Supervised Machine Learning Algorithm to Classify Gender in Google Colaboratory

Praveen Gujjar J. [1] *and Naveen Kumar V.* [2]

Abstract

Supervised machine learning uses the label dataset to classify data. In this paper supervised machine learning algorithm Random Forest Classifier and Support Vector Machine learning algorithm are used to classify gender. Google Colaboratory is a Cloud based service which is also known as Colab. Google Colab is based on Jupyter Notebook where machine learning and deep learning concepts can be implemented. Google Colab provides free access to GPU which is very much required to disseminate Random Forest and Support Vector Machine Learning concepts. This paper made an attempt to classify gender based on the person's height, weight, and shoe size using Random Forest and Support Vector Machine. The result shows that Random Forest Classifier accuracy is relatively better when compared with Support Vector Machine learning algorithm.

Keywords : Gender classification, Google Colab, Jupyter Notebook, Random forest classifier, Support vector classifier

I. INTRODUCTION

A Supervised machine learning model uses the label data for classification and prediction. Random forest classifier is also called an ensemble algorithm, which means that Random Forest Classifier uses one or more algorithms for classifying objects. Support Vector Machine mainly focuses on creating decision boundary or best line segregate n-dimensional space into classes. The shoe size measurement technique varies from nation to nation. Shoe size is measured in inches in India, US uses different measurement techniques as compared with UK. The gender classification may help in intent analysis followed by sentiment analysis. To implement any deep learning and machine learning concept, researchers may require higher end infrastructure to carry out their work. Some of the examples of deep learning and machine learning applications are natural language recognition, sentiment analysis, e-commerce suggestions/recommendations and social network recommendations etc. [1]. Many deep learning and machine learning applications rely on heavy computations on massive datasets. GPU is an acronym for Graphical Processing Units which may help in parallel task processing [2]. Most of the applications in deep learning and machine learning use NVIDIA GPUs [3] [4]. Google Colab will help in providing the necessary infrastructure to carry out the task for free by default (Google colab with 12 GB RAM and 32 GB of Disk space). If any researcher wants to use more, Google Cloud provides necessary infrastructure in a pay-by-hour manner to use the hardware with a fully configured GPU for deep learning applications [5] [6]. The paper is organized into four parts. Section I is the introduction ; Section II is the background and related works ; Section III covers MobileNetV2 and xlm ; Section IV presents the results and discussion, and Section V presents the conclusion.

Manuscript Received : July 12, 2021 ; Revised : September 6, 2021 ; Accepted : September 9, 2021. Date of Publication : October 5, 2021.

[1] P. Gujjar J., *Associate Professor*, Email : dr.praveengujjar@cms.ac.in ; ORCID iD : https://orcid.org/0000-0003-0240-7827
[2] N. Kumar V., *Assistant Professor,* Email : Naveenkumar_v@cms.ac.in ; ORCID iD : https://orcid.org/0000-0002-2618-9978
[1,2] CMS Business School, Jain (Deemed-to-be University), 17, Seshadri Road, Gandhi Nagar, Bengaluru - 560 009, Karnataka.

DOI : https://dx.doi.org/10.17010/ijcs/2021/v6/i5/166516

II. BACKGROUND AND RELATED WORKS

In this paper two supervised learning algorithms random forest classifier and support vector classifier learning have been used. Random forest classifier is also called as ensemble algorithm, which means random forest classifier uses one or more algorithms for classifying objects. Random Forest Classifier uses the decision tree ; these trees are created from subset of training set. Aggregating all the possible decision tree final class of the test object has been chosen. On the other hand, Support Vector Machine is a classification model and it is based on supervised machine learning. Support vector machine uses two group classification problems. Support vector machine mainly focuses on creating decision boundary or best line segregate n-dimensional space into classes. This decision boundary or best line is called as hyperplane. Amazon and Microsoft both will provide high performance computing hardware and the necessary infrastructure to carry out deep learning and machine learning applications. Relatively, Google Colab is cost effective [7] [8]. Expósito et al. [9] showed that Amazon EC2 has performance bottlenecks in application scalability, especially with high performance computing infrastructure.A convolutional neural network is used for image classification and prediction [10]. Transfer learning is nothing but a reuse of the trained model for classification and prediction [11]. Sentiment analysis and opinion mining can be done using the Textblob library using the tool Google Colab [12] [13].

III. PROPOSED METHOD

For gender classification training data used consists of height, weight, and shoe size of people. The proposed classification method includes the following algorithm :

(1) Random forest classifier

(2) Support vector machine

Random forest classifier is also called as ensemble algorithm, which means that random forest classifier uses one or more algorithms for classifying the objects. Random forest classifier uses the decision tree ; these trees are created from subset of training set. Aggregating all the possible decision tree final class of the test object has been chosen. On the other hand, support vector machine is a classification model and it is based on supervised machine learning. Support vector machine uses two group classification problems. Support vector machine mainly focuses on creating decision boundary or best line segregate n-dimensional space into classes. This decision boundary or best line is called as hyperplane. The shoe size measurement technique varies from nation to nation in India. Shoe size is measured in inches, US uses different measurement technique as compared with UK shoe size. The gender classification may help in intent analysis followed by sentiment analysis. The shoe size of women is given in Table I and that of men is given in Table II.

TABLE I.

SHOE SIZE FOR WOMEN

US SIZE	EUROPEAN SIZE	UK SIZE	FOOT LENGTH (Approximate)
7.5	38	5.5	~9.375 in / 23.8 cm
7	37-38	5	~9.25 in / 23.5 cm
6.5	37	4.5	~9 in / 23 cm
6	36-37	4	~8.75 in / 22.5 cm

TABLE II.

SHOE SIZE OF MEN

US SIZE	EUROPEAN SIZE	UK SIZE	FOOT LENGTH (Approximate)
8.5	41-42	7.5	~10.125 in / 25.7 cm
7	40	6	~9.6 in / 24.4 cm
8	41	7	~9.9 in / 25.2 cm
7.5	40-41	6.5	~9.75 in / 24.8 cm

IV. RESULTS AND DISCUSSION

In this paper, the researcher has used random forest classifier and support vector machine to classify the gender on the basis of height, weight and shoe size. Random forest classifier and support vector machines are implemented in Google Colab. Random forest classifier is also called *ensemble algorithm*, which means random forest classifier uses one or more algorithms for classifying objects. Random forest classifier uses the decision tree ; these trees are created from subset of training set. Aggregating all the possible decision tree

final class of the test object has been chosen. On the other hand support vector machine is a classification model and it is based on supervised machine learning. Support vector machine uses two group classification problems. Support vector machine mainly focuses on creating decision boundary or best line segregate n-dimensional space into classes. This decision boundary or best line is called as hyperplane. In this paper, random forest classifier and support vector machine method is used by calling the proper library in Google Colab. The result has been discussed next. The data set has been assigned to list x and corresponding label has been assigned to list y as follows :

The x variable h

```
X = [[177, 70, 43], [181, 80, 44], [154, 54, 37], [160, 60, 38],
    [166, 65, 40], [175, 64, 39], [190, 90, 47]
    [159, 55, 37], [171, 75, 42], [177, 70, 40]]
    [181, 85, 43], [168, 75, 41], [168, 77, 41]]
```

The list y has been assigned with the corresponding label set.

```
Y = ['male', 'male', 'female', 'female', 'male', 'female',
    'male','female','female', 'male', 'female', 'female', 'male']
```

Using the Random Forest Classifier method available in sklearn ensemble library prediction has been made for the following test data :

```
test = [[154, 75, 38], [181,65,40], [190, 70, 43]]
```

The pseudocode for the random forest classifier is shown below :

```
Randomforestclassifier = RandomForestClassifier()
Randomforestclassifer.fit(X,Y)
result = randomforestclassiifer.predict(test)
print(result)
```

The pseudocode for the support vector machine is shown below :

SVC method has been invoked by using library

```
sklearn.svm
supportvectorclassifier = SVC()
supportvectorclassifier.fit(X,Y)
result = supportvectorclassifier.predict(test_data)
print(result)
```

The result of random forest classifier shows ['female' 'female' 'male']. The result of the support vector machine shows ['female' 'male' 'female']. The researcher states that relatively random forest classifier is better than support vector machine for gender classification problem having the data set height, weight, and shoe size.

V. CONCLUSION

In this paper, the researcher presented two supervised learning algorithms : random forest classifier and support vector machine learning. Random forest classifier is also called ensemble algorithm, which means that random forest classifier uses one or more algorithms for classifying objects. Support vector machine uses two group classification problems. Support vector machine mainly focuses on creating decision boundary or best line segregate n-dimensional space into classes. This paper has attempted to demonstrate the use of Google Colab for implementing random forest classifier and support vector machine. The overall conclusions of this study is that random forest classifier relatively gives better accuracy as compared to support vector machine. For future research direction, researchers can employ the GPU to process deep learning application using Google Colab. In the future scope, use of supervised learning and unsupervised learning can be used to address the challenges faced in intent analysis, particularly for Emojis.

AUTHORS' CONTRIBUTION

Dr. Praveen Gujjar J. and Professor Naveen Kumar V. conceived the idea to study transfer learning. The implantation of transfer learning was carried out jointly by them using Google Colab. Dr. Praveen Gujjar J. wrote the manuscript in consultation with Prof. Naveen Kumar.

CONFLICT OF INTEREST

The authors certify that they have no affiliations with or

involvement in any organization or entity with any financial interest, or non-financial interest in the subject matter, or materials discussed in this manuscript.

FUNDING ACKNOWLEDGEMENT

The authors received no financial support for the research, authorship and/or for the publication of this article.

REFERENCES

[1] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature,* vol. 521*,* no., pp. 436 – 444, 2015. [Online]. Available: https://doi.org/10.1038/nature14539

[2] A. Brodtkorb, C. Dyken, T. R. Hagen, J. M. Hjelmervik, and O. O. Storaasli, "State-of-the-art in heterogeneous computing," *Scientific Programming,* vol. 18, no. 1, pp. 1–33, 2010. https://doi.org/10.3233/SPR-2009-0296

[3] NVIDIA Corporation, *"Tesla V100 performance guide: Deep learning and HPC applications,"* NVIDIA Corporation Whitepaper, 2016.

[4] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. H. Katz, A. Konwinski, G. Lee, D. A. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, "Above the clouds: A Berkeley view of Cloud computing," Tech. Rep. UCB/EECS-2009–28, EECS Dept., University of California, Berkeley, Tech. Rep., 2009. [Online]. Available: https://www2.eecs.berkeley.edu/Pubs/TechRpts/2009/EECS-2009-28.pdf

[5] NVIDIA Corporation, "Introduction to NVIDIA GPU cloud," NVIDIA Corporation Application Note, 2018.

[6] Google, "Colaboratory: Frequently asked questions," 2018. [Online]. Available: https://research.google.com/colaboratory/faq.html. Accessed on: June 6, 2018.

[7] G. Juve, E. Deelman, K. Vahi, G. Mehta, B. Berriman, B. P. Berman, and P. Maechling, "Scientific workflow applications on Amazon EC2," in *2009 5th IEEE Int. Conf. on E-Sci., Workshops, IEEE, 2009,* pp. 59–66. https://doi.org/10.1109/ESCIW.2009.5408002

[8] K. R. Jackson, L. Ramakrishnan, K. Muriki, S. Canon, S. Cholia, J. Shalf, H. J. Wasserman, and N. J. Wright, "Performance analysis of high-performance computing applications on the Amazon web services cloud," in *2010 IEEE Second Int. Conf. on Cloud Computing Technol. and Sci.,* pp. 159–168. [Online]. Available: 10.1109/CloudCom.2010.69

[9] R. R. Expósito, G. L. Taboada, S. Ramos, J. Touriño, and R. Doallo, "Performance analysis of HPC applications in the cloud," *Future Generation Comput. Syst.,* vol. *29,* no. 1, pp. 218–229, 2013.

[10] N. Sharma, V. Jain, and A. Mishra, "An analysis of convolutional neural networks for image classification", *Procedia Comput. Sci.,* vol. *132,* pp. 377–384, 2018. [Online]. Available: https://doi.org/10.1016/j.procs.2018.05.198

[11] M. Shaha and M. Pawar, "Transfer learning for image classification," In *2018 Second Int. Conf. on Electronics, Communication, and Aerospace Technol. (ICECA), IEEE,* pp. 656–660, 2018.

[12] J. P. Gujjar and R. P. Kumar H., "Sentimental analysis for running text in email conversation," *Int. J. of Comput. Sci. and Eng.,* vol. 9, no. 4, pp. 67 – 68, 2020.

[13] T. Manjunatha and P. Gujjar, "Performance analysis of Indian Inform. Technol. companies using DuPont Model," *IUP J. of Manage. Res.,* vol. 17, no. 4, pp. 7–14, 2018.

About the Authors

Dr. Praveen Gujjar J. is a bright academician with qualifications such as B.E. (Information Science), M.B.A. in Technology Management, M.Tech in Computer Science and Engineering, and Ph.D. in Business Administration. Dr. Praveen won second prize in Doctoral competition organized by AIMS at Symbiosis Institute of Management Studies, Pune in January 2019.

Prof. Naveen Kumar V. worked in the corporate sector for eight and a half years. He worked as a Process Associate with HCL, as Market Analyst with Reuters for two years and three months, and as Senior Analyst and Team Manager with Deutsche Bank for four years and two years respectively.

ARTHSHASTRA INDIAN JOURNAL OF ECONOMICS & RESEARCH

ISSN 2278 - 1811, IC Value = 78.95 , Indexed in Indian Citation Index (ICI)

SUBSCRIPTION FORM

Subscription Charges

Period	Rate	Discount	Amount Payable
One Year	₹ 1600/-	Nil	₹ **1600/-**
Two Years	₹ 3200/-	₹ 200/-	₹ **3000/-**
Three Years	₹ 4800/-	₹ 300/-	₹ **4500/-**

Subscription Details

Amount

- ❑ ₹ **1600/-**
- ❑ ₹ **3000/-**
- ❑ ₹ **4500/-**

Subscription Period: _____ **to** _____

Payment Details

NEFT/RTGS/M.O./Demand Draft/Cheque No: _____ **dated** _____

in favor of **INDIAN JOURNAL OF ECONOMICS & RESEARCH, payable at New Delhi.**

(Outstation cheques are not accepted. Only payable at par cheques are accepted)

Subscriber No. (Renewal) _____

Delivery Details

Name : _____

Address : _____

_____ **Pin** _____

Email : _____

SEND YOUR SUBSCRIPTION TO :

Meenakshi Gilani
Subscription Manager
Arthshastra Indian Journal of Economics & Research
Y-21, Hauz Khas
New Delhi-110016
Telephone: 011-40586303, (Whatsapp : 9667428899)
Url : http://www.indianjournalofeconomicsandresearch.com
Email: sub@indianjournalofmarketing.com

Account Details for Online Transfer :
Indian Journal of Marketing
Current A/c: 65270940713
Bank Name: State Bank of India, Asian Games Village, New Delhi, India
Branch Code: 07407 - Asian Games Village, New Delhi
IFSC Code: SBIN0007407
MICR Code : 110002007

AMC INDIAN JOURNAL OF CIVIL ENGINEERING

ISSN 2581 - 8171, Indexed in Google Scholar, IC Value = 75.67

SUBSCRIPTION FORM

SUBSCRIPTION CHARGES : **AMC INDIAN JOURNAL OF CIVIL ENGINEERING**

Period	Rate	Discount	AMOUNT PAYABLE
One Year (2 issues)	₹ 1200/-	Nil	₹ 1200/-
Two Years (4 issues)	₹ 2400/-	₹ 100/-	₹ 2300/-
Three Years (6 issues)	₹ 3600/-	₹ 200/-	₹ 3400/-

Amount ☐ ₹ 1200/- ☐ ₹ 2300/- ☐ ₹ 3400/-

SUBSCRIPTION RATES - MINI COMBO OFFER FOR 2 JOURNALS

SUBSCRIBE TO OUR TWO JOURNALS **- INDIAN JOURNAL OF COMPUTER SCIENCE AND AMC INDIAN JOURNAL OF CIVIL ENGINEERING**

Period	Rate	Discount	AMOUNT PAYABLE
One Year (8 issues)	₹ 3000/-	₹ 100/-	₹ 2900/-
Two Years (16 issues)	₹ 6000/-	₹ 200/-	₹ 5800/-
Three Years (24 issues)	₹ 9000/-	₹ 500/-	₹ 8500/-

Amount ☐ ₹ 2900/- ☐ ₹ 5800/- ☐ ₹ 8500/-

SUBSCRIPTION PERIOD: _____ **to** _____

PAYMENT DETAILS

NEFT/RTGS/MO/Demand Draft/Cheque No: _____ **dated** _____

in favor of **INDIAN JOURNAL OF CIVIL ENGINEERING, payable at New Delhi.**

(Outstation cheques are not accepted. Only payable at par cheques are accepted)

SUBSCRIBER NO. (RENEWAL) _____

WRITE YOUR DELIVERY DETAILS/ CORRESPONDENCE ADDRESS (COMPLETE WITH PIN CODE)

Name : _____

Address : _____

_____

_____

_____

_____ **Pin** _____

Email : _____ Mobile No _____

SEND YOUR SUBSCRIPTION TO :
Meenakshi Gilani
Subscription Manager
AMC Indian Journal of Civil Engineering
Y-21, Hauz Khas, New Delhi-110016
Telephone : 011- 40586303, (Whatsapp : 9667428899)
Url : http://indianjournalofcivilengineering.com
Email : sub@indianjournalofmarketing.com

ACCOUNT DETAILS FOR ONLINE TRANSFER :
Indian Journal of Marketing
Current A/c: 65270940713
Bank Name: State Bank of India, Asian Games Village, New Delhi, India
Branch Code: 07407 - Asian Games Village, New Delhi
IFSC Code: SBIN0007407
MICR Code : 110002007

PLEASE NOTE : If you are making the payment by NEFT/RTGS, it is MANDATORY to email us the : NEFT UTR Number / RTGS Number, Date of Transfer, Amount Transferred, Journals you wish to subscribe, Subscription Period Preference, and Your Full Address for Sending the journals (with pincode)