ISSN : 1693-3214 print | 2721-0111 online

Bootcamp Seminar and Machine Learning Algorithm Workshop for the Data Science Club

Agus Eko Minarno[1], Lailis Syafa'ah[2], Moch. Chamdani Mustaqim[3]
[1,2,3] Department of Informatics, Universitas Muhammadiyah Malang, Malang, Indonesia

* Corresponding author: aguseko@umm.ac.id

ARTICLE INFO

Article history
Received April 7, 2021
Revised May 3, 2021
Accepted September 30, 2021
Available Online November 27, 2021

Keywords
Workshop,
Data Science,
Python,
Data Science Club.

ABSTRACT

The development of data and information needs in the era of Society 5.0 is very crucial because it determines many business decisions. Data in the past becomes valuable when it becomes a historical fact that can illustrate findings to assess future business directions. Based at the University of Muhammadiyah Malang, the Data Science Club has a total membership of more than 200 people spread across East Java. The problem that often occurs in the Data Science community is Machine Learning algorithms' low literacy, especially for new members. Coupled with the development of the Machine Learning algorithm that is so fast and massive. For that, we need activities that can directly impact the Data Science community by presenting the latest algorithms and programming techniques. This service activity proposes a Machine Learning workshop for Data Science by teaching various computational algorithms to the Indonesian Data Science community, which has spread in Indonesia and the East Java region. This activity presents 12 workshop materials for participants who will be delivered by speakers who have expertise in their fields, both from the University of Muhammadiyah Malang, and present national speakers in collaboration with the Data Science Club of the University of Muhammadiyah Malang.

1. Introduction

In the midst of the rapid development of artificial intelligence (AI) technology today, not many people know that artificial intelligence consists of several branches, one of which is machine learning or machine learning [1]. Machine learning (ML) technology is a branch of AI that is very attractive because machine learning is a machine that can learn like humans. In general, artificial intelligence is divided into seven branches, namely machine learning, natural language processing, expert systems, vision, speech, planning, and robotics. The branching of artificial intelligence is intended to narrow the scope when developing or learning AI because, basically, artificial intelligence has a very broad scope.

Please cite this article as: Eko Minarno, A., Syafa'ah, L & Chamdani Mustaqim, M (2021). Bootcamp Seminar and Machine Learning Algorithm Workshop for the Data Science Club. Jurnal Dedikasi, 18(2), 1-11. doi:https://doi.org/10.22219/dedikasi.v18.i2.16189

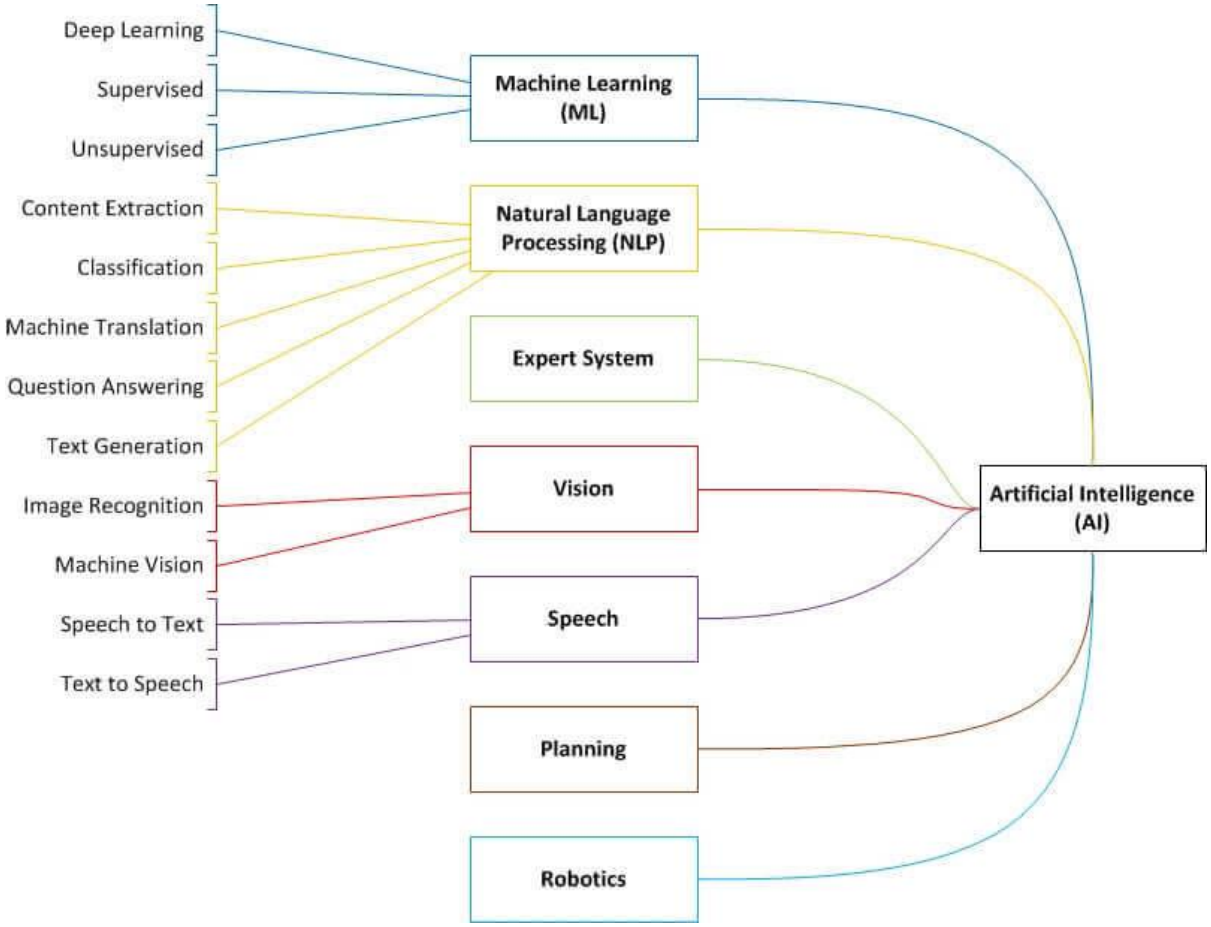

Fig 1. Sub-topic discussion of AI material.

Machine learning (ML) technology is a machine developed to be able to learn on its own without direction from its users. Machine learning is developed based on other disciplines such as statistics, mathematics, and data mining so that machines can learn by analyzing data without needing to be reprogrammed or ordered [2]. In this case, machine learning has the ability to obtain existing data by its own command. ML can also study existing data and the data it obtains so that it can perform certain tasks. The tasks that can be done by ML are very diverse, depending on what they learn. The term machine learning was first put forward by some mathematical scientists such as Adrien Marie Legendre, Thomas Bayes, and Andrey Markov in the 1920s by suggesting the basics of machine learning and its concepts [3]. Since then, many ML have developed. One example of a well-known ML application is Deep Blue, made by IBM in 1996 [4].

The role of machine learning helps people in many fields. Even today, the application of ML can easily be found in everyday life. For example, when using the face unlock feature to open a smartphone device [5], or when browsing the internet or social media, you will often be presented with several advertisements. The advertisements that appear are also the result of ML processing which will provide advertisements according to the person of the user. The concept of Machine Learning is to make machine learning and behave in certain ways after entering certain types of data as input. Machine Learning offers several techniques to solve data processing problems, including supervised learning and unsupervised learning techniques [6].

Supervised Learning uses labeled training data sets and maps input data to desired outputs. Classification and Regression are two common Supervised Learning

Please cite this article as: Eko Minarno, A., Syafa'ah, L & Chamdani Mustaqim, M (2021). Bootcamp Seminar and Machine Learning Algorithm Workshop for the Data Science Club. Jurnal Dedikasi, 18(2), 1-11. doi:https://doi.org/10.22219/dedikasi.v18.i2.16189

ISSN : 1693-3214 print | 2721-0111 online

techniques. Classification assigns a sample of data into one of the discrete, predefined classes; Regression can be viewed as a statistical methodology that is generally used for numerical predictions.

Clustering is a general unsupervised learning technique that takes certain data sets as input and classifies them into a finite number of clusters according to some similarity index [7]. When used on a set of objects, it helps identify some inherent properties present in objects by classifying them into subsets that have meaning in the context of a particular problem. More specifically, an object is represented by a set of features that characterize it. Feature objects are usually described as data points in multidimensional space [8]. So, grouping can be considered as a partitioning of data points based on homogeneity criteria. When the number of clusters, K, is known as a priori knowledge, the groupings are formulated so that objects in the same cluster are more similar in some ways than those in different clusters. This involves minimizing some extrinsic criteria. The K-means algorithm, starting with k arbitrary cluster centers in space, partitions a given set of objects into k subsets based on the distance metric. The cluster center is updated on a recurring basis based on the objective function optimization. This method is one of the most popular clustering techniques widely used because it is very easy to implement and very efficient, with linear time complexity [9][10].

The Data Science Club is a community that houses students who have an interest in learning about science data at the University of Muhammadiyah Malang. As a community, of course, there is a work program or activity carried out by the community [11]. In carrying out its activities, there are problems faced by the community. The problem that occurs in the Data Science Club community is understanding Machine Learning algorithms which are partially unstructured so that it is confusing to understand Machine Learning algorithms comprehensively. The second problem is the lack of resources, both tutors and source code, which provide hands-on experience from concept to machine learning algorithm implementation.

2. Methods

2.1 Lecture and Question and Answer Methods

This method is carried out in a way where the trainer explains the concept of the material to be discussed to the training participants. Suppose some still don't understand, a question-and-answer session is held, so participants can ask questions that are not understood regarding the material. The trainer will answer by explaining it to the participants.

2.2 Demonstration Methods and Examples (practice)

A demonstration shows and plans how a job or how something is done. This method involves breaking down and demonstrating through examples. This method is very effective because it is easier to show participants how to do a task and because it is combined with learning aids such as pictures, text material, discussions, or practicing the material taught.

The entire arrangement of activities designed is arranged by mutual agreement between the tutors, participants, and the committee as the event organizer. The mechanism and realization of the Bootcamp Seminar and Machine Learning Algorithm Workshop for the Data Science Club are described in the following points:

Please cite this article as: Eko Minarno, A., Syafa'ah, L & Chamdani Mustaqim, M (2021). Bootcamp Seminar and Machine Learning Algorithm Workshop for the Data Science Club. Jurnal Dedikasi, 18(2), 1-11. doi:https://doi.org/10.22219/dedikasi.v18.i2.16189

a. The number of participants involved was 50 people.
b. Place : Google Meet (Online)
c. Bootcamp Activity Time :

Table 1. Bootcamp Activity Time

No	Timing (minutes)	Activity
1.	0 – 5	Opening
2.	5 – 85	Delivery of materials
3.	85 – 90	Closing

Each meeting is held 3 times a week, namely on days:
- Monday (07:00 PM)
- Friday (07.00 PM)
- Saturday (07:00 PM)

In addition, quiz submission activities are carried out every last week. Meanwhile, the final project is carried out after all the meetings are completed within 2 weeks.

d. Bootcamp Activity Materials and Speakers

Table 2. Bootcamp Activity Materials and Speakers

No	Materials	Speakers
1.	Introduction to Data Science	R. Panji Maharjo Tri (Education Staff at DSI Jatim)
2.	Basic Python 1	Moch. Chamdani Mustaqim (Head of Data Science Club UMM)
3.	Basic Python 2	Rangga Putra Kurnia Wiratama (Informatics Laboratory Assistant - UMM)
4.	Excel Data Processing	Muhammad Yusril Hasanuddin (Informatics Laboratory Assistant - UMM)
5.	Introduction to Pandas	Muhammad Hussein (Community Members of DSC-UMM)
6.	Sharing Session	Ahmad Hamdani (Lead Tools Programmer at Agate International)
7.	OOP in Python	Moch. Shandy Tsalasa Putra (Community Members of DSC-UMM)
8.	Preprocessing Data, Variable Reduction and Selection	Yussyafrida C. R (Education Staff at DSI Jatim)
9.	Variable Selection and Reduction	Yussyafrida C. R (Education Staff at DSI Jatim)
10.	Basic Statistics	Rangga Putra Kurnia Wiratama (Informatics Laboratory Assistant - UMM)
11.	Data Visualization	Moch. Daffa Shafwan Chairullah

Please cite this article as: Eko Minarno, A., Syafa'ah, L & Chamdani Mustaqim, M (2021). Bootcamp Seminar and Machine Learning Algorithm Workshop for the Data Science Club. Jurnal Dedikasi, 18(2), 1-11. doi:https://doi.org/10.22219/dedikasi.v18.i2.16189

		(Informatics Laboratory Assistant - UMM)
12.	Classifications Algorithm (KNN - DT - SVM)	Muhammad Hussein (Community Members of DSC-UMM)
13.	Basis Data	R. Panji Maharjo Tri (Education Staff at DSI Jatim)
14.	XGBoost and Random Forest	Moch. Shandy Tsalasa Putra (Community Members of DSC-UMM)
15.	Model Evaluation and Cross Validation	Amri Muhaimin (RnD Staff at DSI Jatim)
16.	Data Clustering	Yufis Azhar, M. Kom (Informatics Study Program Lecturer - UMM)
17.	Time Series	Amri Muhaimin (RnD Staff at DSI Jatim)
18.	ARIMA	Amri Muhaimin (RnD Staff at DSI Jatim)

e. Facility
   1. Learning Modules
   2. Learning Recordings
   3. Certificate
   4. Final Project Reward

2.3 Schedule

Table 3. Bootcamp agenda schedule

No	Activity Name	Month(s)											
		1	2	3	4	5	6	7	8	9	10	11	12
1.	Preparation												
2.	Communication with partners												
3.	Module's creation												
4.	Implementation												
5.	Evaluation of activities												
6.	Preparation of activity reports												
7.	Publication of activities												

3. Results and Discussion

3.1 Inhibiting Factors

a. Some participants came late, so the event's rundown changed several times according to the predetermined final time.
b. Slow internet connection at the time of delivery of the material, so it is necessary to hold repeated material delivery.

Please cite this article as: Eko Minarno, A., Syafa'ah, L & Chamdani Mustaqim, M (2021). Bootcamp Seminar and Machine Learning Algorithm Workshop for the Data Science Club. Jurnal Dedikasi, 18(2), 1-11. doi:https://doi.org/10.22219/dedikasi.v18.i2.16189

c. The level of discussion of the material makes participants a little overwhelmed to understand.
d. The speaker was deemed too fast in delivering the material.
e. The delivery of the material is still not clear.
f. Some of the materials presented did not match the expectations of the participants.

3.2 Supporting Factors

a. Modules and Slides are easy for the participants to understand.
b. All participants who attend get the facilities provided, except for the door prize, which is only distributed to one participant based on the final project assessment.
c. The moderator has been able to arrange the course of the event to continue effectively and conducive.
d. Resetting the changed schedule is done quickly so that the presenters and participants can immediately adjust (conditionally).

3.3 The Level of Success of Events

Table 4. The final score of the bootcamp participants

Name	Attendance (Att)	Weekly Task (WT)	Att Total Score	WT Total Score	Final Score	Status
Pringgo Arif Himantoro	3		4.5	0	4.5	Drop out
Mochammad Khairun Nur Hidayat	1		1.5	0	1.5	Drop out
Wahyu Budi Utomo	5		7.5	0	7.5	Drop out
Yudanty Dwi Setya Wardhani	8	2	12	28	40	Drop out
Farli Nahrul Javier	13	2	19.5	28	47.5	Drop out
Muhammad Zein Ihza Fahrozi	15	3	22.5	42	64.5	Pass with Note
Dwi Wdiainto	17	3	25.5	42	67.5	Pass with Note
Muhammad Naufal Al Ghifari	18	3	27	42	69	Pass with Note
Devanis Dwi Sutrisno	19	3	28.5	42	70.5	Pass with Note
Fachry Fathurahman	19	3	28.5	42	70.5	Pass with Note
M Fauzi Rais	19	3	28.5	42	70.5	Pass with Note
Bella Dwi Mardiana	14	4	21	56	77	Pass
Aulia Ligar Salma Hanani	17	4	25.5	56	81.5	Pass

Please cite this article as: Eko Minarno, A., Syafa'ah, L & Chamdani Mustaqim, M (2021). Bootcamp Seminar and Machine Learning Algorithm Workshop for the Data Science Club. Jurnal Dedikasi, 18(2), 1-11. doi:https://doi.org/10.22219/dedikasi.v18.i2.16189

Rahmi Nurazizah	17	4	25.5	56	81.5	Pass
Riski Setiawan	19	4	28.5	56	84.5	Pass
Moch. Daffa Shafwan Chairullah	12	5	18	70	88	Pass
Dinda Arinawati Wiyono	14	5	21	70	91	Pass
Putri Sari Asih	14	5	21	70	91	Pass
Siti Alfiyatun Ni'Mah	14	5	21	70	91	Pass
Muhammad Rifal Alfarizy	15	5	22.5	70	92.5	Pass
Wana Salma Labibah	15	5	22.5	70	92.5	Pass
Indah Rezki Ananda	17	5	25.5	70	95.5	Pass with Appreciation
Muhammad Fadhlan	17	5	25.5	70	95.5	Pass with Appreciation
Ulfah Nur Oktaviana	17	5	25.5	70	95.5	Pass with Appreciation
Annisa Fitria Nurjannah	18	5	27	70	97	Pass with Appreciation
Mega Dewi Giridrawardani	18	5	27	70	97	Pass with Appreciation
Mico Winaryo Dwi Putranto	18	5	27	70	97	Pass with Appreciation
Agus Hendriyawan	19	5	28.5	70	98.5	Pass with Appreciation
Faras Haidar Pratama	19	5	28.5	70	98.5	Pass with Appreciation
Jalu Nusantoro	19	5	28.5	70	98.5	Pass with Appreciation
M.Randy Anugerah	19	5	28.5	70	98.5	Pass with Appreciation
Ricky Hendrawan	19	5	28.5	70	98.5	Pass with Appreciation
Andi Shafira Dyah Kurniasari	20	5	30	70	100	Pass with Appreciation
Adhigana Priyatama	20	5	30	70	100	Pass with Appreciation
Alfin Yusriansyah	20	5	30	70	100	Pass with Appreciation
Dicky Prabowo Octianto	20	5	30	70	100	Pass with Appreciation
Mochammad Hazmi Cokro Mandiri	20	5	30	70	100	Pass with Appreciation

Please cite this article as: Eko Minarno, A., Syafa'ah, L & Chamdani Mustaqim, M (2021). Bootcamp Seminar and Machine Learning Algorithm Workshop for the Data Science Club. Jurnal Dedikasi, 18(2), 1-11. doi:https://doi.org/10.22219/dedikasi.v18.i2.16189

Moh. Badris Sholeh Rahmatullah	20	5	30	70	100	Pass with Appreciation
Muhammad Nuril Huda	20	5	30	70	100	Pass with Appreciation
Wien Nurul Dewani	20	5	30	70	100	Pass with Appreciation

Based on the data of Table 4, Fig. 2 shows the high success rate of the event based on the number of participants who passed the title (either with or without notes or appreciation).

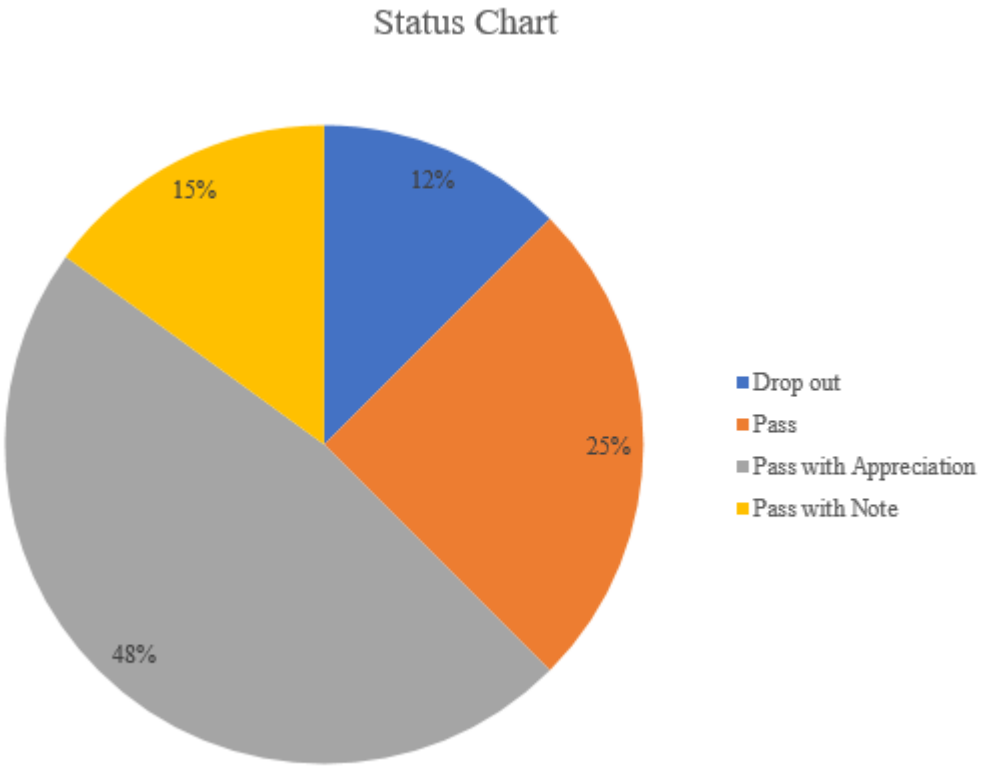

Fig. 2. Status Chart

3.4 The Results of The Participant's Final Project

3.4.1. Classification of human brain tumor types

Fig. 3 and Fig. 4 shows the classification results of the types of brain tumors in humans and compares the accuracy of the models used in the classification.

Please cite this article as: Eko Minarno, A., Syafa'ah, L & Chamdani Mustaqim, M (2021). Bootcamp Seminar and Machine Learning Algorithm Workshop for the Data Science Club. Jurnal Dedikasi, 18(2), 1-11. doi:https://doi.org/10.22219/dedikasi.v18.i2.16189

Probability Table

Prediction of : **Proposed_model** | Predition Time : **0.1804** second

#	Label	Probability
0	Glioma	1.0
1	Meningioma	0.0
2	Pituitary	0.0

Prediction of : **Blance_trial_v1** | Predition Time : **0.1638** second

#	Label	Probability
0	Glioma	0.0048
1	Meningioma	0.9952
2	Pituitary	0.0

Fig. 3. Probability table result of human brain tumor classification.

Result

Based on the highest predictive probability of all labels We know that :
The **PROPOSED_MODEL** predicts that the input image is **100 % Glioma**
The **BLANCE_TRIAL_V1** predicts that the input image is **99 % Meningioma**

Blance_trial_v1 is the fastest model with 0.1638 second of prediction time.
Proposed_model is the slowest model with 0.1804 second of prediction time.

Blance_trial_v1 Is **1.10 * FASTER** Then The Slowest Model.


Input Image

Fig. 4. Classification result of human brain tumor types

3.4.2. Visualization of data analysis results from various topics used in the final project.

Fig. 5 shows a visualization of data analysis results based on the topics used in the final project.

Please cite this article as: Eko Minarno, A., Syafa'ah, L & Chamdani Mustaqim, M (2021). Bootcamp Seminar and Machine Learning Algorithm Workshop for the Data Science Club. Jurnal Dedikasi, 18(2), 1-11. doi:https://doi.org/10.22219/dedikasi.v18.i2.16189

Fig. 5. Visualization of data analysis results from various topics used in the final project.

3.4.3. Web application for heart attack prediction using the KNN algorithm.

Fig. 6 shows a web application to predict a person's likelihood of having a heart attack based on the attribute values of the features provided by the web, which are then predicted using the KNN algorithm.

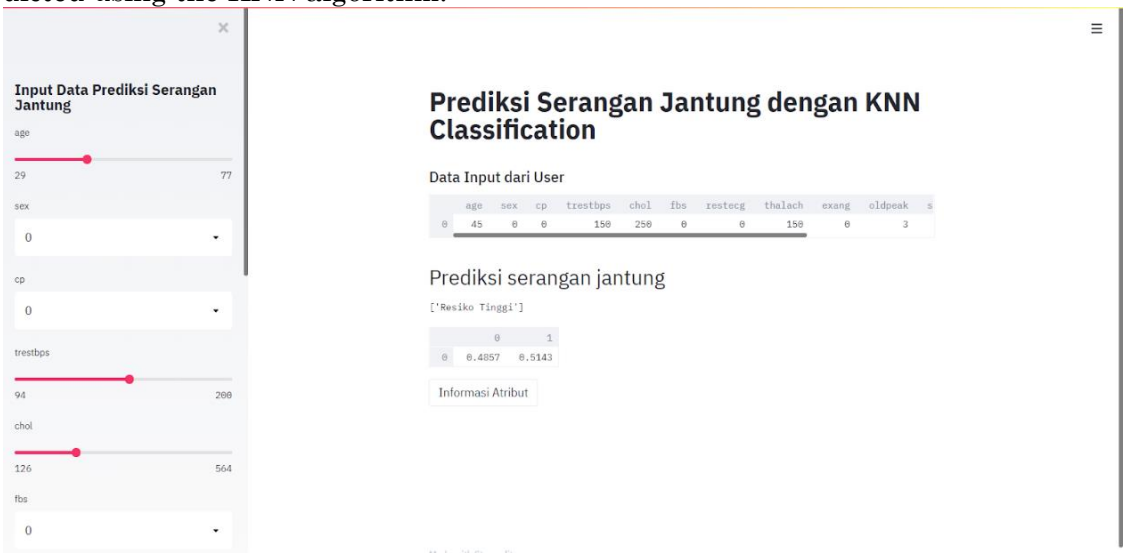

Fig. 6. Web application for heart attack prediction using the KNN algorithm.

4. Conclusion

The entire series of activities for the Bootcamp Seminar and Machine Learning Algorithm Workshop for the Data Science Club can run well, and all obstacles can be

Please cite this article as: Eko Minarno, A., Syafa'ah, L & Chamdani Mustaqim, M (2021). Bootcamp Seminar and Machine Learning Algorithm Workshop for the Data Science Club. Jurnal Dedikasi, 18(2), 1-11. doi:https://doi.org/10.22219/dedikasi.v18.i2.16189

handled optimally. The results of the final project of the Bootcamp participants were also very satisfying, even though some participants were excluded because they did not follow the series of events properly. In the future, the results of the evaluation of activities and criticism of suggestions from all parties are used as correction material to improve this activity so that it can run even better. This program is entirely dedicated to members of the DSC community, and all material components and the final project can be managed by the community and participants involved in it. In the future, if the covid-19 pandemic has ended, this kind of Bootcamp can be done face-to-face and can be done regularly so that communication and knowledge transfer is even more intense.

Acknowledgements

Thank you to all the participants in this Bootcamp for the enthusiasm. And also, thanks to all committee of this Bootcamp that helped this event to be successful. And thank you also for the DSC and DSI that helped to carry out this activity.

References

[1] "KECERDASAN BUATAN - Test Repository." http://eprints.unm.ac.id/4532/ (accessed Mar. 26, 2021).

[2] "Mathematics for Machine Learning - Marc Peter Deisenroth, A. Aldo Faisal, Cheng Soon Ong - Google Books." https://books.google.co.id/books?hl=en&lr=&id=pFjPDwAAQBAJ&oi=fnd&pg=PR9&dq=mathematics+machine+learning&ots=VLgr8HJcHa&sig=81iJlWhGVkd08gfbFrLAk-CZSbQ&redir_esc=y#v=onepage&q=mathematics machine learning&f=false (accessed Mar. 26, 2021).

[3] M. Haenlein and A. Kaplan, "A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence," *Calif. Manage. Rev.*, vol. 61, no. 4, pp. 5–14, Aug. 2019, doi: 10.1177/0008125619864925.

[4] M. Campbell, A. J. Hoane, and F. H. Hsu, "Deep Blue," *Artif. Intell.*, vol. 134, no. 1–2, pp. 57–83, Jan. 2002, doi: 10.1016/S0004-3702(01)00129-1.

[5] S. Waghmare, M. Chatterjee, and S. L. Varma, "Authentication System for Android Smartphones," *Int. J. Comput. Appl.*, vol. 87, no. 5, pp. 19–24, Feb. 2014, doi: 10.5120/15204-3602.

[6] B. C. Love, "Comparing supervised and unsupervised category learning," *Psychon. Bull. Rev.*, vol. 9, no. 4, pp. 829–835, Dec. 2002, doi: 10.3758/BF03196342.

[7] A. Ahmadyfard and H. Modares, "Combining PSO and k-means to enhance data clustering," *2008 Int. Symp. Telecommun. IST 2008*, pp. 688–691, 2008, doi: 10.1109/ISTEL.2008.4651388.

[8] C. D. Megawati, E. M. Yuniarno, and S. M. S. Nugroho, "Clustering of Female Avatar Face Features Consumers Choice using KMeans and SOM Algorithm," 2019, doi: 10.1109/ISITIA.2019.8937279.

[9] C. Y. Chen and F. Ye, "Particle swarm optimization algorithm and its application to clustering analysis," in *2012 Proceedings of 17th Conference on Electrical Power Distribution, EPDC 2012*, 2012, pp. 789–794, doi: 10.1109/icnsc.2004.1297047.

[10] C. Y. Chen and F. Ye, "Particle swarm optimization algorithm and its application to clustering analysis," 2004.

[11] S. A. Abstak, "POTENSI DAN KEKUATAN MODAL SOSIAL DALAM SUATU KOMUNITAS," 2013. Accessed: Mar. 26, 2021. [Online]. Available: https://journal.unhas.ac.id/index.php/socius/article/view/381.