SIGMUND: Optimization of DISC Methodology and distribution of groups with Machine Learning

Cleiton Silva Ribeiro, Higor Barbosa Ferreira, João Pedro Rodrigues, Ricarte Elias, Victor Santos Boroni, Rafaela Priscila Cruz Moreira*, Luiz Melk de Carvalho, Flávio Henrique Batista de Souza

Centro Universitário de Belo Horizonte UNIBH, Brazil

*Corresponding author

Abstract.

In this work, the SIGMUND software is presented and an optimization process for the behavioral assessment questionnaire used in DISC (Dominance, Influence, Steadiness, Compliance) methodology is described. DISC is a tool that sets out to establish objectives aligned with the professional profile of employees by measuring strengths and weaknesses to achieve better results. This questionnaire is typically expensive because it covers several aspects through a broad variety of questions to assess the professional profile. Therefore, in this work, the Bagged Decision Trees (BDT) algorithm was implemented to reduce the number of questions without losing the quality of the test. The BDT estimates the importance of attributes within a database, returning a score for each attribute. In this algorithm, the higher the score, the greater the importance. After optimizing the questionnaire, it was used to define the profile of candidates and, subsequently, it was created academic groups that allow for better interaction and experience between members. As these groups are created randomly or by free choice, there may be conflicts or even not taking full advantage of the contribution that each one would have if the profile of these members were taken into account. For the creation of these smart groups based on the profile of the candidate, the K-means clustering algorithm was applied to define an ideal number of people in each group to guarantee that there is a balance. As a result, the BDT managed to reduce the number of questions by 52% with an accuracy level of 75.8% and for the division of groups in equal variation, the K-means obtained an accuracy of 97%.

Keywords: Academic Groups; Behavioral Assessment; Data Mining; K-Means Algorithm; Quiz.

1. Introduction

It is notorious the difficulty of teachers and/or advisors of academic projects to create groups in a room with a large number of students, where groups are usually created at random through the call, which can end up generating groups that have conflicts between members or groups that produce with underperformance. To assist teachers with this activity, this application will be based on the renowned methodology DISC (Dominance, Influence, Stability, Compliance) which assists organizations and team administrators in their decision making in order to trace the behavioral profile of people, which in this work will be optimized,

using machine learning techniques (ML) that is an area of artificial intelligence where algorithms are created to teach the machine to perform certain tasks (Utami et al, 2021; Hunt, 2018; Hutter et al, 2019)

Therefore, about the objective proposed in the project, which is the optimization of the DISC questionnaire and the construction of groups based on the profile of each student, a machine learning algorithm called Bagged Decision Trees will be used, which can estimate the importance of attributes in a database, being returned the score of each attribute, the higher the score, the greater the importance of the attribute (Pedregosa et al., 2011). With this, it is expected to obtain the same effectiveness of the DISC methodology, however, in a fast and practical way. The DISC methodology works on four fronts of dominant profile for each person (Utami et al, 2021; Hunt,2018;), in this adaptation, they will be separated according to the nomenclature:

- Communicator: Who are communicative people and usually endorsed with great charisma and power of persuasion and always enthusiastic about projects and novelties, tends to be very optimistic and relates with ease.
- Executor: It is a profile of people endorsed with extreme self-confidence, this type of person is dominant and in extreme cases, can be authoritarian and dictatorial, accepts and gets along well with challenges and difficulties, has an extreme sense of competitiveness and is usually courageous in their postures and in defending their points of view.
- Planner: They are stable and patient people, of constant rhythm and high degree of conservatism, hardly panic and have a small capacity to improvise.
- Analyst: They are detailed and meticulous people, the analyst person is organized, responsible and highly conservative, being skilled in controlling repetitive processes and routines.

Therefore, using the behavioral profile of each student in the class, the application will create the groups, however, for this activity, the clustering algorithm called K-Means will be used, where according to Sinaga & Yang (2020) the algorithm aims to group the data trying to separate samples into n groups of equal variation, minimizing a criterion known as inertia or sum of squares within the cluster.

This research aims to demonstrate the development of a web and mobile application to assist teachers and /or advisors of academic projects in the formation of groups in a practical, efficient and simple way. To accomplish this goal, the Bagging Decision Trees machine learning algorithm will be used to optimize the behavioral assessment questionnaire used in the DISC methodology to define a person's predominant profile. This algorithm is basically a set modeling that uses a "divide to conquer" technique where it involves a predictive model to obtain a better accuracy and stability of the model. Once this is done, with the help of the K-means clustering algorithm, the construction of the groups will be carried out automatically and relatedly. Therefore, K-means will correlate students' responses to determine groups and the Bagged Decision Trees algorithm, after running in a database, will define the questions that are most relevant to defining a person's behavioral profile.

2. Theoretical Concepts

To support the present work on the contextualization, it was researched the following two main concepts. Firstly,

1

2

People's emotions: in the 1920s researcher William Marston wrote a book called The Emotions Of Normal People and in this book he describes 4 main profiles in which he divided the population and that is precisely where the acronym: DISC. So, D means Dominance profile; Influence I; S of Stability and C of Caution (Chen et al., 2016). There are some variations and adaptations of the study by researcher Marston, for this, this project was highlighted a mental mapping tool that has some differences and the main one is the nomenclature of the 4 profiles. Soon we have: Executor, Communicator, Planner and Analyst. It is important to note that all have characteristics of the 4 profiles, so there is no one person being just a planner or just executor for example, but rather, determining the profile with the highest predominance, which in most cases is one or two type. The DISC methodology brings a lot of assertiveness to the process of understanding a particular person. Using the method with a 25-question questionnaire to determine a person's behavioral profile, DISC brings with it a series of information on trends in the behavior of that candidate you don't already know (Chen et al., 2016).

The second concept was based in Machine Learning classification's algorithms. One of the topics covered by machine learning is classification algorithms that involve assigning classes to objects in a dataset. Three main approaches can be considered for classification: supervised, semi-supervised and unsupervised classification (Rodriguez et al., 2019). In this research, where applied. (Pedregosa et al., 2011; Sinaga & Yang, 2020; Hutter et al, 2019):

- **K-Means:** K-means performs three steps. The first step chooses the initial centroids, the most basic method being to choose $\boldsymbol{k}$ samples from the $\boldsymbol{X}$ data set. After initialization, K-means consists of looping between the two other steps. The first step assigns each sample to the nearest centroid. The second step creates centroids by taking the average value of all samples assigned to each previous centroid. The difference between the old and new centroids is calculated and the algorithm repeats these last two steps until this value is less than a limit.

- **Ensemble Method:** The ensemble method is a machine learning technique that combines the result of multiple models in search of producing a better predictive model. There are several pre-fixed algorithms of ensemble classifiers, such as: bagging, boosting, bayesian averaging, among others.

- **Random Forest:** Random Forests is a classifier that evolves from decision trees. To sort a new instance, each decision tree provides a classification for input data. The random forest collects the ratings and chooses the most voted prediction as a result. The

input of each tree is given to sample the original dataset. In addition, a subset of features are randomly selected from among the optional features to increase the tree on each node. Each tree is grown without pruning. Essentially, random forest allows many weak or weakly correlation classifiers to form a strong classifier.

So, based on those concepts, the research was performed.

3. Methodology

The research methods were based in frameworks (structures and algorithms) selection, development process and experiments definitions.

3

3.1 Frameworks Selection

The frameworks and development tools were selected (Siahaan & Sianipar, 2019; Singh & Bhadani, 2020):

- **Web Application:**. For the frontend, the programming languages HTML5.2, CSS and Javascript. For backend, the tools used are PHP 7 and Python 3, responsible for the operation of the site and communication with database. For the Database Management System (DBMS) postgre SQL 12.
- **Mobile Application:** Flutter, Google's free, open source framework used to create high-performance, high-fidelity apps for IOS, Android, from a single code base.
- **Infrastructure:** to support the platform, the Infrastructure as a Service (IaaS) mode of the Amazon Web Services (AWS) platform is used. Two specific types of services have been determined: Elastic Compute Cloud (EC2) that offers a resizable computing capability and the Amazon Relational Database Service (RDS) that facilitates the configuration, operation, and scalability of relational databases. In this application we used a preconfigured image of a Linux Apache PHP and PostgreSQL (LAPP) web server developed and made available by Bitnami in ubuntu 16.04 version 7.3.18-0.
- **Algorithms and Structure:** for the definition of the questionnaire that is present in the application, the behavioral evaluation methodology was used based on the DISC methodology, which by default has 25 questions. The applicability of the DISC methodology with this questionnaire had a reformulation, with redefinitions of their questions. The amount of questions was reduced, with the help of an algorithm that defines the questions that are most relevant to determine the behavioral profile of a person. For this, it was used the Python language, statistical regression and machine learning with the Scikit Learn library that allows avoiding excessive adjustments and reducing variations. In data mining, the K-means algorithm was used. Thinking about test automation was used Selenium, a popular open source tool for automated testing for web applications, this tool is used for the execution of scripts and test cases, allowing in this case the execution of tests in different browsers for the purpose of quality and optimization of the application.

3.2 Development Process

Among the numerous educational institutions, group work follows a series of factors that end up influencing the performance of academic work and projects, thus inevitably obtaining minimal or intermediate performance, because members may not be focused on the same objectives and guidelines for some reason, such as lack of affinity with members or even lack of knowledge in the designated part of the work. It refers by working group or work project, a set of different people who are part of a project developing tasks in search of a single goal previously defined.

With this, we can aim at the need for a platform that facilitates the organization and elaboration of groups in the academic environment and for this was used methodologies, libraries and algorithms for the elaboration of the Sigmund platform that, in addition to facilitating the elaboration, makes the performance ideal bringing groups with the possible performance to improve the performance dividing the members by area of activity. So the following steps were performed:

- **Data Collection:** In order to perform these optimization studies and reach an expressive result for our research, a collection of questionnaires to evaluate the DISC behavioral profile was performed, we used a python algorithm that performed access to the "http://tribus.dx.am/" website, answering the questionnaires randomly, thus generating a base with a total of 59,219 records.
- **Database balancing:** In the first phase of optimization, the base was balanced in order to match the registration number for each profile classification ("Analyst - Blue" – 17714 records, "Communicator - Green"-13801 records, "Executor - Yellow"-12198 records, "Planner - Red" 15506 records), figuring as a unbalanced database. The data obtained above were collected as a result obtained from the data collection using the DISC questionnaire method adjusted with the platform parameters with only the 12 most equivalent questions for determining a behavioral profile. The data were adjusted according to the smallest amount obtained from a profile, that is, we took the four profile and matched all with the lowest results of the profiles, which in this example is 12,198 students per profile. After balancing, a total of 48,792 records were available, distributed equally, 12,198 records for each behavioral profile.
- **Optimization:** The optimization of the application was divided into two modules, aiming at the first module to optimize the DISC evaluation questionnaire and in the second module aims to optimize the ordering of groups based on the questionnaires answered.
- **Optimization of the evaluation questionnaire:** to define the questions that will be equivalent to trace the behavioral profile of a person we use the Extra Tree Classifier class that belongs to the ensemble methods, which aims to combine the predictions of several base estimators built with a given learning algorithm in order to improve the generalization / robustness of a single estimator, that is, using the Extra Tree Classifier class it will be possible to classify the columns (variables) of the collected base and thereby select the columns that obtained the result above the calculated average, and this result is considered satisfactory to consider that question (variable) as a candidate to be implemented in the optimized DISC questionnaire model.

Those were the references for experiments.

3.3 Experiment's Definitions

In possession of the results obtained through the method mentioned above, we obtained 14 columns with scores of the calculated mean. With this, the accuracy of the model was tested with the 14 columns above the calculated average and obtained the value of 78.7% of hit. An accuracy test was also performed with the 12 main columns returned by the model and 75.8% accuracy was obtained in this model. Therefore, in order to provide a better user experience, especially for students in the classroom and for having reached a level of accuracy of 75.8%, the author decided to use only the 12 main columns, which are the questions that will be listed for students to answer to determine their predominant behavioral profile.

After defining the questions will be used in the Sigmund questionnaire, the construction stage of the classification model begins, which will be responsible for dividing and ordering the groups intelligently. For this, the K-means algorithm that integrates the clustering module of the scikit learn library will be used, which aims to group the data trying to separate samples into n groups of equal variation, minimizing a criterion known as inertia or sum of squares within the cluster. This algorithm requires specifying the number of clusters to be mounted, for this reason, was implementing a function by the author that will receive the number of students and find the best number of groups to be formed, that is, the function will return to the model the specific number of clusters that should be created.

To assess the accuracy of the model, 3 stages of tests were performed. In the first stage, the model was executed 30 times in a row, and each execution was performed arbitrarily in a database containing random data of number of students and profiles. Thus, 96.7% accuracy was obtained with the logic that the algorithm implements in the code, which in turn will define the number of ideal groups for a given student base. Following the same logic used in the previous step, the model was performed again 100 times and obtained 97% accuracy. In the last stage of the test the model was executed for another 200 times and reached the accuracy of 94.5%. Therefore, based on the positive results in the applied tests, the author concluded that the algorithm is responding as expected.

4. Results

A Given the proposed objective of optimizing the DISC questionnaire (25 questions) and to carry out the construction of groups of academic projects intelligently, data collection was performed in order to create a database for testing, where it was possible to perform the construction and adjustments of the algorithms that perform the selection of the main variables used to define the behavioral profile of a person. This basis was also applied in the tests carried out in the construction of the logic used to define the number of groups that a given base would have to have to achieve satisfactory levels of application. All this to be able to provide students and teachers with a tool that can assist them in the academic day-to-day, with the support on two platforms, web and mobile, which will have their results presented in this topic.

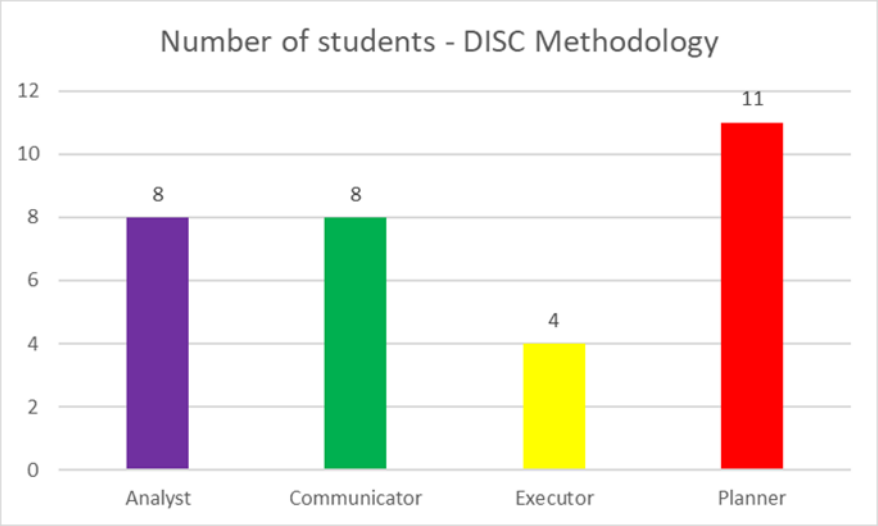
4

4.1 Optimization of the evaluation questionnaire

In this step, a random sample was randomly taken on a balanced basis, that is, containing the same number of records for each of the behavioral profiles. This sample contains 31 records that had their behavioral profiles classified by disc methodology, having 8 records with analyst profile, 11 records with planner profile, 8 records with communicator profile and 4 records with executor profile, as illustrated in Figure 1. The graph shows the general classification of the base, in which it has 8 analysts, 8 communicators, 4 executors and 11 planners.

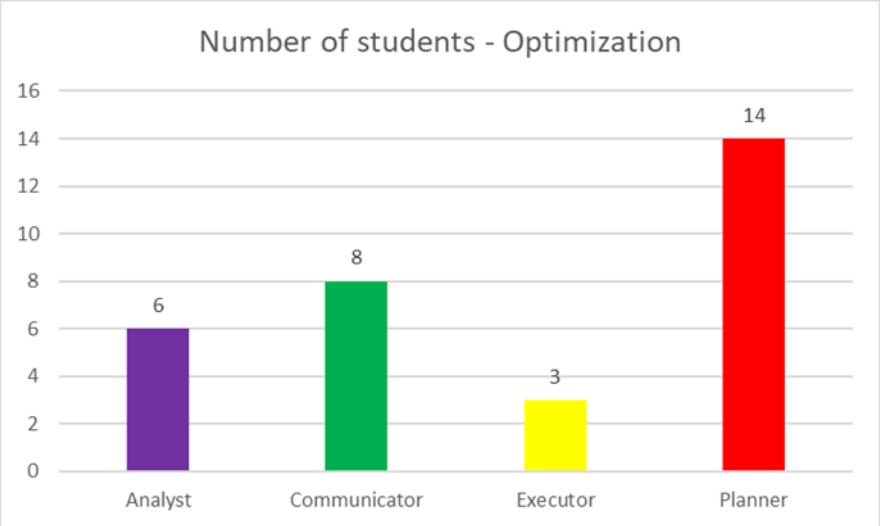
Thus, by applying the optimization proposed by the platform to perform the analysis of the answers and trace the profile of the student using the Extra Tree Classifier class that belongs to the ensemble methods of the scikit learn library, we obtained 77.4% of correct answers in comparison with the classification of the DISC methodology, getting the following results as shown in the Figure 2. Figure 2 shows the overall classification of the base, using the model proposed by the application, where students will answer the 12 questions that the Extra Tree Classifier class returned as the most important.

Figure 7: Base classified by the following DISC methodology



Source: Author, 2020

Figure 2: Optimized classification result



Source: Author, 2020

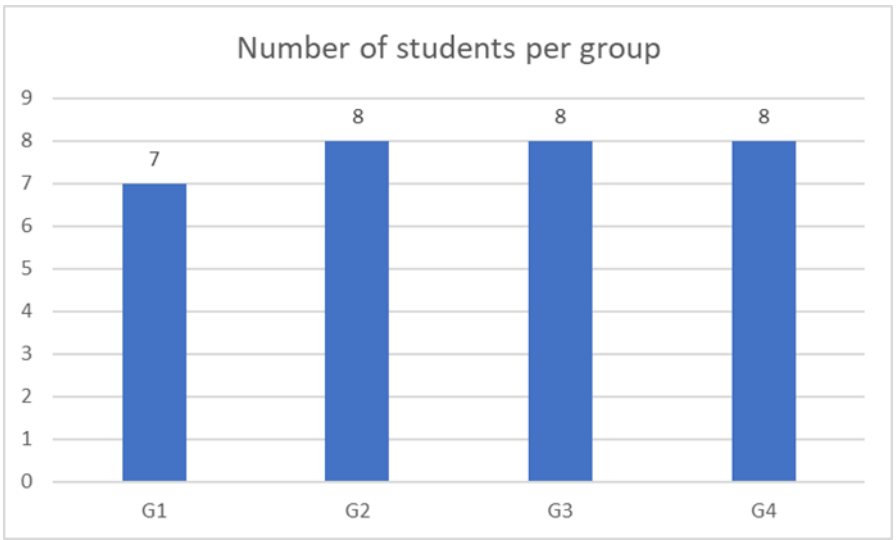
Thus, the results were: 6 analysts, 8 communicators, 3 executor and 14 planners.

4.2 Group ordering optimization

At the end of the optimization module, the grouping model was constructed capable of evenly distributing students in a certain number of groups. For this, the K-means algorithm was used, which aims to group the data trying to separate samples into n groups of equal variation, minimizing a criterion known as inertia or sum of squares within the cluster.

To present the result, a sample containing 31 records was randomly selected, constructed as follows, 12 responses from the optimized questionnaire and behavioral classification. As soon as, after the execution of the algorithm. Figure 3 shows the resulting groups on the X axis and the number of students per corresponding group on the Y axis.

Figure 3: Result of dividing groups



Source: Author, 2020

4.3 Web Platform

The results obtained in the web platform are related to the gains in the User Experience, a more intuitive and responsive application, to Figures 4 to 9 demonstrate these aspects. Figure 10 shows the home screen of the web application, where is seen of the purpose of the platform, as well as the definition of some topics such as:

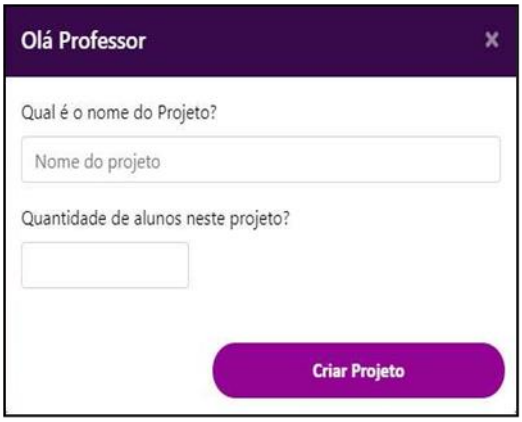
- Group Formation for Academic Projects: In this topic you can create your project so that Sigmund distributes students in groups logically according to the profile outlined in the behavioral questionnaire!
- Join Academic Projects: Here you can access a project to then answer the behavioral questionnaire, where Sigmund will use this questionnaire to put you in a group of people who best match you.
- Behavioral Questionnaire: It's fast! By answering these questions you will soon know what your predominant behavioral profile is! Let's go, shall we?

In Figure 5 the teacher can quickly create a project of his class. With only two information (Project Name and Number of students in the class), the teacher can open a project and already release it to his class through a unique identification, which can be accessed through a QRCODE for mobile devices, or via web via project ID. So that as soon as possible, the platform can capture the data of the students and perform the separation of the groups of this project. Figure 6-a, on the student side, we have the function of linking to a project, where it is necessary for the user to enter name, email, project key.

Figure 4: Home Screen

Source: Author, 2020

Figure 5: Create Project



Source: Author, 2020

Some important aspects of this screen: If access is via mobile device, it will read the QRCODE made available by the teacher and as soon as the key field of the project will already be filled, another important point is that the email field is a unique authenticator, which make the student can register only one email per project. Figure 6-b shows the method used by the platform to be able to perform the 'calculation' and classify a student, all through a questionnaire of 12 questions, which at the end of the same, the user will be able to know their behavioral profile based on the DISC methodology, however, in an optimized way.

Figure 6: Screen Link in Project

a) Link in Project *b) Optimized Questionnaire*

Source: Author, 2020

Finally, the screen returns the groups to the teacher, as indicated in Figure 7, such presentation reveals all groups properly separated automatically by the platform, showing the student's name, his behavioral profile and which group he belongs to.

Figure 7: Group Presentation Screen

Source: Author, 2020

4.4 Mobile Platform

The results related to the mobile platform served as modernization, convenience and approach to the public, as mobile services currently help a lot in productivity and can be accessible at any time through a smartphone. Figure 8 is the home screen of the mobile application, where the user can easily start the services listed as needed. If the user wants to participate in a project previously created by his teacher, he/she must enter this option to join a

project, after clicking on the "Scan QR Code" button will activate the reading of his camera, and the user must point to the code made available by his teacher. The function of the screen described in Figure 9-a, is identical to that of Figure 6-a, which basically the user must enter the data requested to be part of a project in which their teacher will perform.

Figure 8: Home Mobile

Source: Author, 2020

As figure 6-b exemplifies Figure 9-b, has the function of capturing students' answers to trace and present their behavioral profile on the screen.

Figure 9: Screen Link in Mobile Project

a) Start Quiz *b) Mobile Quiz*
Source: Author, 2020

5. Conclusion

According the planning of creating a platform to assist teachers and advisors of academic projects in the formation of groups in the classroom, and this division or separation of students will be carried out efficiently, because with the help of the application, which will have specific algorithms to perform this distribution, this task will be carried out in a fully automatic way,

bringing dynamism and practice to the teacher. All this in a very simple and intuitive way. The results of the objectives were achieved in the deliveries of a web and mobile platform that bring ease and convenience to users, in addition to operational tests that were performed for adjustments and validation of the platform. It is remarkable that the application has great potential and there are great possibilities for growth and value aggregation with new features. Thinking about it, some cases that were raised as possible improvements to the author's platform:

- Teacher or team trainer: The system provides functionality to logging and obtain access to the groups he is managing. Reordering groups through filters. Generic class charts with the percentages of each profile in the class.
- Student: Make your environment with more interaction with the teacher, where it will be possible to check schedule of delivery of activities deadlines among other notifications that the teacher will have the possibility to communicate with his student, besides having a history of all his projects delivered, having the possibility to continue a project with a new teacher in the next semester.

Finally, it can be concluded that when using the platform, the user will have with him a tool that will bring agility and automation in the process of separation and organization of academic groups, making this task simple and fast and erudite, thus bringing efficiency to teachers in the classroom.

References

Chen, T. Y., Tsai, M. C., and Chen, Y. M. (2016). A user's personality prediction approach by mining network interaction behaviors on Facebook. Online Information Review.

Hunt, S. J. (2018). Communication and change: A qualitative study of law enforcement team disc personality traits (Doctoral dissertation, Grand Canyon University).

Hutter, F., Kotthoff, L. and Vanschoren, J. (2019). Automated machine learning: methods, systems, challenges (p. 219). Springer Nature.

Pedregosa, F. et al. (2011). Scikit-learn: Machine learning in Python. The Journal of Machine Learning research, 12, 2825-2830.

Rodriguez, M. Z., Comin, C. H., Casanova, D., Bruno, O. M., Amancio, D. R., Costa, L. D. F., and Rodrigues, F. A. (2019). Clustering algorithms: A comparative approach. PloS one, 14(1), e0210236.

Sinaga, K. P. and Yang, M. S. (2020). Unsupervised K-means clustering algorithm. IEEE Access, 8, 80716-80727.

Siahaan, V., & Sianipar, R. H. (2019). OpenCV-Python with PostgreSQL for Absolute Beginners: A Hands-On, Practical Database-Driven Applications. Sparta Publishing.

Singh, A., & Bhadani, R. (2020). Mobile Deep Learning with TensorFlow Lite, ML Kit and Flutter: Build scalable real-world projects to implement end-to-end neural networks on Android and iOS. Packt Publishing Ltd.

Utami, E., Oyong, I., Raharjo, S., Hartanto, A. D. and Adi, S. (2021). Supervised learning and resampling techniques on DISC personality classification using Twitter information in Bahasa Indonesia. Applied Computing and Informatics.