Supervised learning and resampling techniques on DISC personality classification using Twitter information in Bahasa Indonesia

Received 21 March 2021
Revised 28 May 2021
12 August 2021
Accepted 23 August 2021

Ema Utami and Irwan Oyong
Universitas Amikom Yogyakarta, Sleman, Indonesia
Suwanto Raharjo
IST AKPRIND, Yogyakarta, Indonesia, and
Anggit Dwi Hartanto and Sumarni Adi
Universitas Amikom Yogyakarta, Sleman, Indonesia

Abstract

Purpose – Gathering knowledge regarding personality traits has long been the interest of academics and researchers in the fields of psychology and in computer science. Analyzing profile data from personal social media accounts reduces data collection time, as this method does not require users to fill any questionnaires. A pure natural language processing (NLP) approach can give decent results, and its reliability can be improved by combining it with machine learning (as shown by previous studies).

Design/methodology/approach – In this, cleaning the dataset and extracting relevant potential features "as assessed by psychological experts" are essential, as Indonesians tend to mix formal words, non-formal words, slang and abbreviations when writing social media posts. For this article, raw data were derived from a predefined dominance, influence, stability and conscientious (DISC) quiz website, returning 316,967 tweets from 1,244 Twitter accounts "filtered to include only personal and Indonesian-language accounts". Using a combination of NLP techniques and machine learning, the authors aim to develop a better approach and more robust model, especially for the Indonesian language.

Findings – The authors find that employing a SMOTETomek re-sampling technique and hyperparameter tuning boosts the model's performance on formalized datasets by 57% (as measured through the F1-score).

Originality/value – The process of cleaning dataset and extracting relevant potential features assessed by psychological experts from it are essential because Indonesian people tend to mix formal words, non-formal words, slang words and abbreviations when writing tweets. Organic data derived from a predefined DISC quiz website resulting 1244 records of Twitter accounts and 316.967 tweets.

Keywords Supervised learning, Resampling techniques, Profiling analysis, DISC, Twitter information, Bahasa Indonesia

Paper type Research paper

The authors would like to acknowledge the valuable funding provided by KEMENRISTEK DIKTI in Hibah Penelitian Terapan Unggulan Perguruan Tinggi (PTUPT) of the fiscal year 2019–2021, contract uumber: 227/SP2H/LT/DRPM/2019.

Applied Computing and Informatics
Emerald Publishing Limited
e-ISSN: 2210-8327
p-ISSN: 2634-1964
DOI 10.1108/ACI-03-2021-0054

1. Introduction

Personality is what distinguishes individual humans from each other and defines their tendencies in their reactions and actions. Although academics and researchers have long attempted to gather knowledge about personality, it remains an evergreen area of research today. Analysis of individuals' personal social media accounts offers a promising approach, as this method does not require users to complete any questionnaires, thereby reducing necessary time and increasing credibility. Social media usage is increasing every day, and thus a huge amount of textual and visual data is uploaded to the Internet daily [1]. For such tasks, Twitter and Facebook are among the two most popular social media platforms, as they provide accessible application programming interfaces (APIs) that might be used in conjunction with external testing applications for corpus and data collection [2]. The Indonesian language, also known as Bahasa Indonesia, is an Austronesian language that is the official language of Indonesia and one of the official languages of ASEAN; according to 2021 Statista data, it is currently the 11th most spoken language in the world [3]. Twitter has a tremendous number of tweets from users, especially in Indonesia, and this is beneficial for personality profiling; it has been proven that data volume correlates positively with profiling accuracy [4].

Companies have increasingly prioritized the selection of new prospective workers based on their personalities, as they perceive particular attitudes and characters as indicative of good work performance. No companies want to risk potential losses from employees' misconduct and bad behavior [5]. In Indonesia, industrial production measures the output of businesses in the industrial sector (including in manufacturing, mining and utilities). Several models are popularly used for personality assessment in industrial societies, including the Big Five (OCEAN) personality model, the Myers-Briggs Type Indicator (MBTI) and the Keirsey Temperament Sorter. This study, however, uses the dominance, influence, stability and conscientious (DISC) assessment framework, as it explicitly concentrates on behavioral preferences and thus is more applicable, explanatory and comprehensible than the other models mentioned above [6, 7]. First proposed by William Moulton Marston, the DISC model divides individuals' feelings and behaviors into four different dimensions: dominance, influence, stability and conscientious [8]. Kim Yun-Yong *et al.* investigated office workers' DISC behavior style and its effect on organizational commitment, job satisfaction and job performance, finding that persons with "dominant" personalities tend to have better job performance than individuals characterized as "steady" [9]. An investigation by Fariha Tabasum *et al.* found a significant positive relationship between the personality of a salesperson and consumers' perceptions. They also showed that, where a salesperson has an attractive personality, the sales of specific products and services increase [10]. Likewise, Joy Eberechukwu Agodi, Emmanuel Onyedikachi Ahaiwe and Aniekan Eyo Awah found a strong positive relationship between sales performance and personality traits. Successful salespersons, they noted, were empathetic, assertive and ambitious [11].

This study is a continuation of preliminary research conducted by Utami *et al.* that used a pure natural language processing (NLP) approach for profile analysis [12]. Adi *et al.* used three machine learning techniques—stochastic gradient descent (SGD), gradient boosting and stacking—to conduct personality recognition using Indonesian-language Twitter posts, finding that SGD and super learner are better than XGBoost in this case [13]. Machine learning is a growing branch of artificial intelligence that learns from data patterns to make decisions without human intervention. Machines can be trained to cognize and assess individuals' personalities [19]. Similarly, by employing support vector machine and linear regression with an LFM-1b dataset, user demographics might be identified based on music listening information [14]. Another study also used datasets from Twitter, Facebook and YouTube to recognize individuals' personalities using the decision tree and support vector machine approaches [15]. In such cases, as mentioned by Gu *et al.* [16], it is necessary to use

NLP to clean the dataset and extract relevant potential features (as assessed by psychological experts) as Indonesians tend to mix formal words, non-formal words, slang and abbreviations when writing social media posts, with formal words dominating their compositions [17]. We aim to develop a better approach, combining NLP techniques and machine learning to form a more robust model. As Tadesse *et al.* mentioned in their research, using social network features for personality prediction can return better results than using only linguistic features [18].

2. Related studies

Empirical studies of job satisfaction, organizational commitment and job performance have been conducted for many years. For example, a survey conducted in D City between January 28 and May 30, 2010, which collected data from 315 office workers and analyzed it using SPSS/WIN 17.0, found personality has a significant influence on organizational commitment, job satisfaction and job performance [9]. A study conducted by Fariha Tabasum *et al.*, employed random sampling through SPSS software to conduct correlation and reliability analysis of questionnaire data collected from 172 respondents, finding that customer perception and sales are influenced by salespersons' personality traits—particularly their agreeableness. These findings were recommended to help managers develop deeper insights regarding their sales strategies, thereby enabling them to develop optimal approaches [10].

Joy Eberechukwu Agodi, Emmanuel Onyedikachi Ahaiwe and Aniekan Eyo Awah found that a strong and positive relationship exists between empathy, assertiveness and ambitiousness and sales performance. They thus underlined the need to improve the integrity, trust, capability and confidence of salespersons by setting specific targets rather than comparing individual salespersons or comparing individuals against the rest of the team [11].

A study by Utami *et al.* found that a pure NLP approach could be used to predict the personality traits of Twitter users, but needed to be enhanced using a better approach [12]. In four experiments, of 139 users validated by psychological experts, the best accuracy rate (37.41%) was returned using a not stemmed–not weighted keyword vocabulary. Several different machine learning methods have been used by researchers for prediction. For example, one study employed advanced classifiers such as XGBoost and ensemble for prediction, finding that ensemble has high accuracy (82.59%) for real-time Twitter datasets [19]. Significant improvements can be made by achieving a 1.0 ROC AUC score with SGD and super learner in research for personality recognition on Twitter in the Indonesian language [4]. Tommy Tandera *et al.* experimented on traditional machine learning algorithms such as Naive Bayes, SVM, logistic regression, gradient boosting and LDA, using three features (LIWC, SPLICE and SNA). They proved that the SVM algorithm had the highest average accuracy in manually gathered datasets, but the results did not differ much from other algorithms [20].

3. Materials and methods

This study was conducted with the guidance of two psychological experts as well as previous successful works on related issues. Both hold masters' degrees in industrial psychology and are involved mostly in maintaining psychological standards while developing the instrument and defining the features. The most challenging part of the study was the textual feature preparation, which we handled by using NLP techniques (with an Indonesian-language NLP toolkit, which is relatively limited compared to English-language ones) to clean and pre-process the data. An overview of this study's methodologies is presented in Figure 1 and described in Section 3 below.

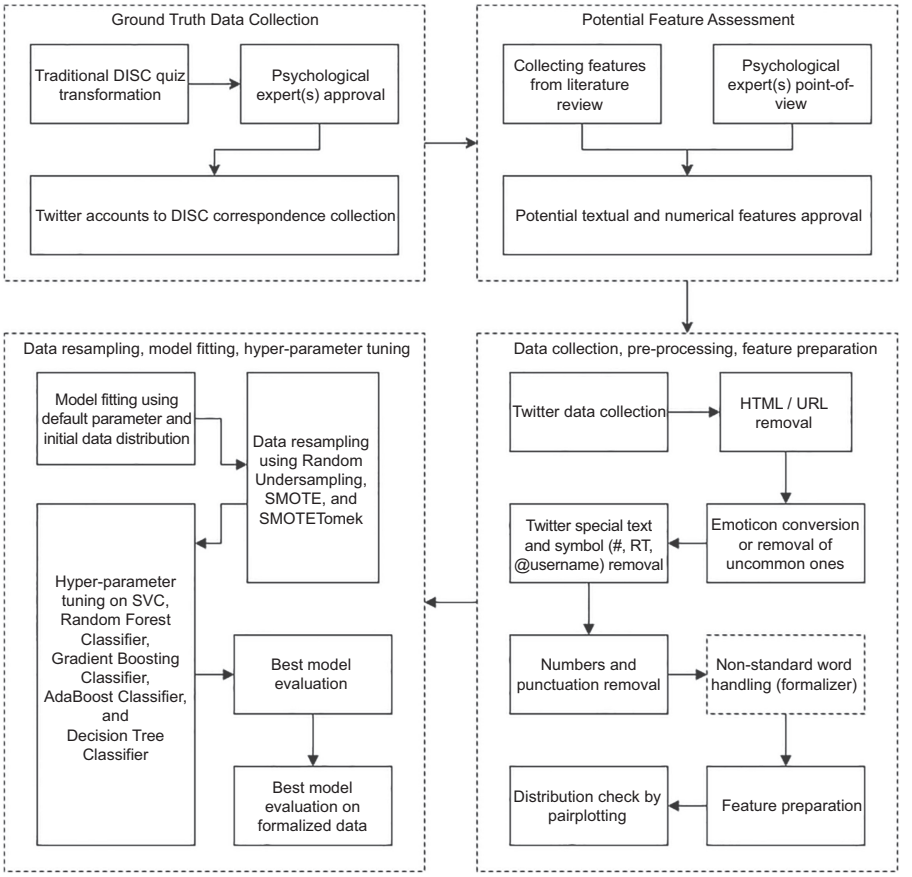

Figure 1.
Overview of the methodologies

3.1 Ground truth data collection

An Internet-based DISC instrument, developed based on industrial psychology experts' direction, was used to collect ground truth data. This remains the golden standard for data collection, as it is the closest to what experts usually do during interviews. Explicit requests for visitors' Twitter accounts and disclaimers on the academic usage of collected data were given. This quiz was based on the assessment instruments used by industrial psychology experts during live interviews in their regular business process, using a different format while still maintaining the instrument's assessment ability and credibility [8, 21]. The aforementioned DISC instrument is included as supplementary material and accessible using the provided link (see Appendix).

3.2 Feature assessment

During the data collection process, experts were also consulted regarding the potential features collected from the literature reviews and the state-of-the-art classification techniques. Eight features were identified as potentially useful for DISC profiling analysis [16]. Some were approved by experts, who also added other potential features; ultimately, nine features were identified (see Table 1).

No	Features	Dominance	Influence	Stability	Compliance
1	Self- mentioning	Aku	Saya	Kami	Self-name addressing
2	Post characteristics	All about his/her own opinion or desire	Something funny or influencing	Giving useful information to others	Mostly about giving critics and arguing
3	Post frequencies	1-2 times a day	2-5 times a week	1-2 times a week	Less than 1 time a week
4	Tendency while questioning other person's post	On what	On who	On how	On why
5	How they react to mentions	Not so responding	Respond and reply expressively	Replying clearly	Replying carefully
6	Hashtag usage (#)	Only using 1 or a few hashtags picturing themselves	Using many variance of hashtags	Not using many hashtags	Tend not to use hashtags
7	Openness of location	Tends to inform any locations	Only informing favorite places	Creating his/her own location, ex: "Inside my house"	Not informing location
8	Post length	Broad and long but no continuity	Broad and long with continuity	Only the core and substantiate parts	Short and compact
9	Emoticon usage	Using serious, concentrating and angry emoticons	Using smiling and laughing emoticons	Using emoticons containing feelings, ex: Heart, moon	Tend not to use emoticons

Table 1.
Studied potential features by the experts

A dump of the DISC website provided the records of 3,132 people who accessed and answered the quiz. Using the Twitter username entered by the participants, we gathered corresponding posts from the previous three months. Unfortunately, not all of the records contained a Twitter username, and not all of the recorded Twitter accounts were valid for data collection, either because they were non-existent or protected. We further filtered accounts to remove bots, non-personal accounts and non-Indonesian speaking accounts, thereby reducing noise. Ultimately, data were collected from 1,244 Twitter accounts, producing a corpus of 316.967 tweets which are all analyzed.

3.3 Data preprocessing and feature extraction

Because Twitter users often write in non-standard forms of Indonesian, it was inevitable that collected tweets were unready for immediate classification. As such, data cleaning and preprocessing were first necessary. For this, we employed InaNLP (the Indonesian language NLP toolkit) to formalize the non-standard form of language including handling abbreviations or shortened form of a word to its original form [22] and see its impact on model performance. After the data were cleaned, we extracted features from the dataset based on the information collected in raw textual form.

3.4 Distribution check and pair plot

The collected quiz results contained the following distribution of classes: $'S'$: 499,$'C'$: 359,$'I'$: 229,$'D'$: 157. Where classes are unevenly distributed, the model usually performs poorly when used to classify a more general condition. To overcome this issue, oversampling (duplicating samples from the minority class) or undersampling (deleting samples from the majority class)

techniques are often used to adjust the class distribution of a data. The pair plot distribution of some initial features is shown in Figure 2.

4. Results and discussion

4.1 Model fitting using default parameter and initial data distribution

Perfect balance rarely occurs in class distribution, and thus immediate usage in model-building will not produce accurate and robust model. Based on the best results identified by previous studies, we handpicked several base classifier models to be used. As expected, initial data distribution performed poorly, as seen from the SVC (support vector classifier) example in Table 2 and Figure 3.

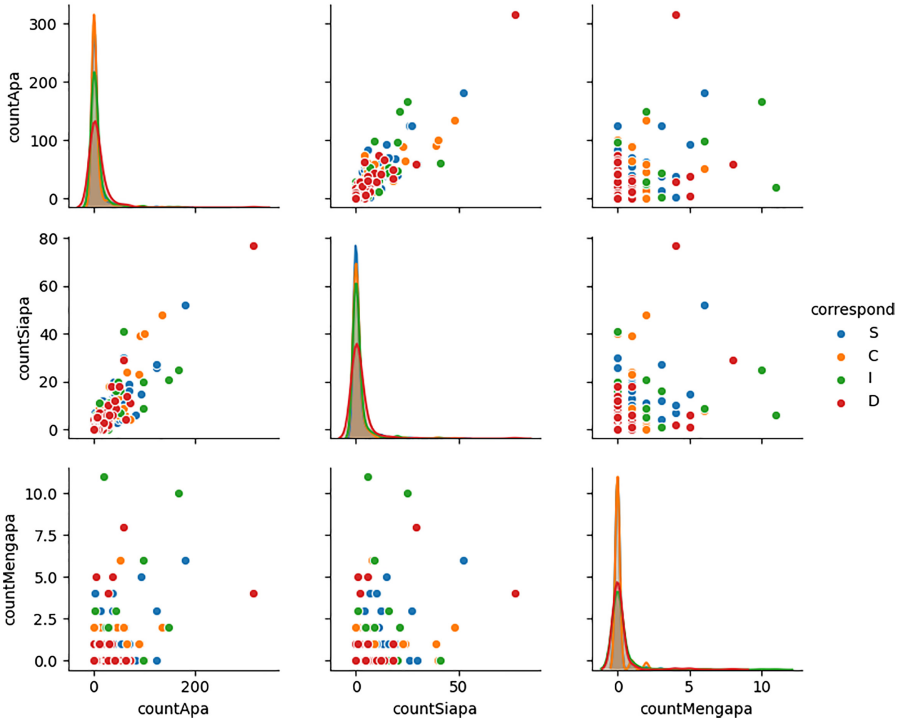

Figure 2.
Pair plot distribution of some initial features

	Precision	Recall	F1-score	Support
Dominance	0.00	0.00	0.00	111
Influence	0.00	0.00	0.00	166
Stability	0.39	0.99	0.56	340
Conscientious	0.30	0.01	0.02	253
Accuracy			0.39	870
Macro average	0.17	0.25	0.15	870
Weighted average	0.24	0.39	0.23	870

Table 2.
Evaluation table of SVC default parameter and initial data

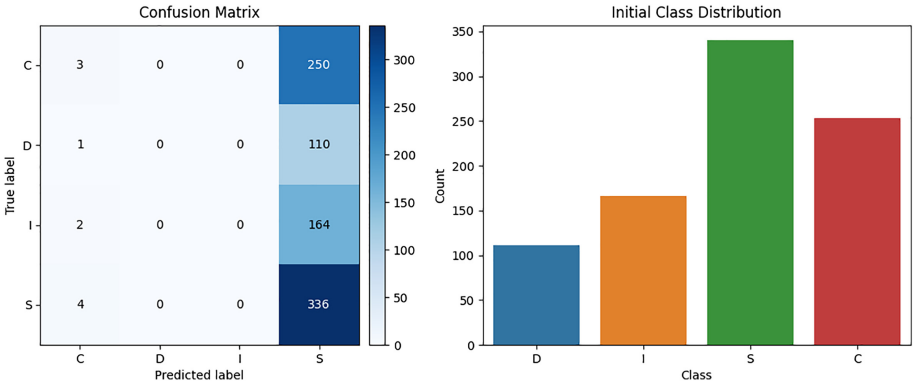

Figure 3.
SVC using default parameter confusion matrix and initial class distribution

4.2 Model hyper-parameter tuning on resampled data

Using default parameters and initial data distribution gave us benchmarks for improvement, and this was realized by using hyperparameter tuning and resampling on the data. Several resampling methods have been found to perform well with imbalanced datasets. Based on related studies, we handpicked several of the best performing methods: random under sampling, SMOTE and SMOTETomek [23–27]. To automate the process of finding the best hyper-parameters tuning and resampling technique, a grid search approach was used. In the grid search approach, hyperparameter tuning is performed in order to determine the optimal values for a given model; here, it is implemented using GridSearchCV from scikit-learn [28]. A comparison of hyperparameter tuning and data resampling performance, as shown from the GridSearchCV results, is provided in Table 3.

The random undersampling technique involves randomly selecting examples from the majority class (in this case, the S and C classes) to delete from the training dataset, thereby

Estimator	Minimum score	Mean score	Maximum score	STD score
Random undersampling GridSearchCV				
SVC	0.191	0.251	0.362	0.064
Decision tree classifier	0.163	0.265	0.357	0.065
Ada boost classifier	0.243	0.277	0.347	0.037
Random forest classifier	0.206	0.271	0.346	0.048
Gradient boosting classifier	0.181	0.245	0.331	0.049
SMOTE GridSearchCV				
SVC	0.484	0.531	0.581	0.034
Decision tree classifier	0.301	0.342	0.382	0.029
Ada boost classifier	0.492	0.540	0.574	0.030
Random forest classifier	0.342	0.377	0.445	0.035
Gradient boosting classifier	0.270	0.320	0.385	0.038
SMOTETomek GridSearchCV				
SVC	0.520	0.555	0.597	0.025
Decision tree classifier	0.270	0.320	0.385	0.038
Ada boost classifier	0.342	0.377	0.445	0.035
Random forest classifier	0.513	0.554	0.593	0.029
Gradient boosting classifier	0.416	0.477	0.567	0.049

Table 3.
The comparison on how the hyper-parameter tuning performs along the data resampling result table sorted by maximum F1-Macro score

reducing each class to the same number. As no new information is introduced, any underlying issues with absolute rarity are not addressed [27].

Meanwhile, the synthetic minority oversampling technique (SMOTE) oversamples the data by introducing new, non-replicated data to the minority classes (in this case, the I and D classes) from the five nearest neighbors [27]. The SMOTE preprocessing algorithm is considered the de facto standard in the framework of learning from imbalanced data [29].

SMOTETomek is a good way to avoid the disadvantages of the SMOTE and Tomek Link techniques. The SMOTETomek technique is applied using the library from imbalanced_learn and includes a SMOTE function for oversampling as well as a Tomek Link function for undersampling [29]. The algorithm flow of the SMOTETomek method is to combine SMOTE and Tomek Link to form a pipeline [25].

4.3 Best model evaluation

After identifying the best performing scenario using SVC and SMOTETomek resampling technique, we did a five-fold cross-validation using the best hyperparameter values. Figure 4 shows how the class distribution transformed after resampling. As seen in Table 4, a

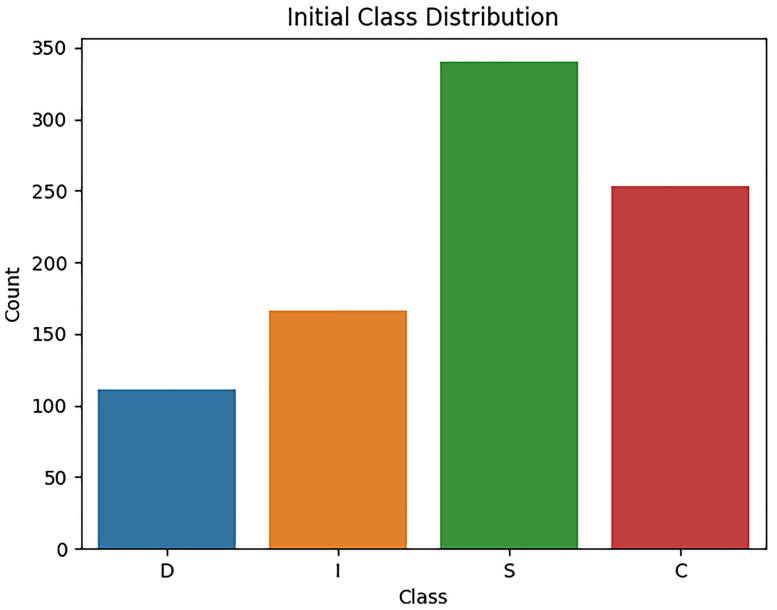

Figure 4.
SMOTETomek resampled class distribution

Table 4.
Evaluation table of SVC using SMOTETomek resampling and hyper-parameter tuned confusion matrix

	Precision	Recall	F1-score	Support
Dominance	0.72	0.68	0.70	310
Influence	0.66	0.56	0.61	293
Stability	0.34	0.45	0.39	265
Conscientious	0.54	0.50	0.52	280
Accuracy			0.55	1148
Macro average	0.57	0.55	0.55	1148
Weighted average	0.58	0.55	0.56	1148

considerable improvement was realized over the default parameter and initial data distribution benchmark.

4.4 Best model evaluation on formalized text data

We employ hyperparameter tuning on a formalized text dataset to see how the results differ. It is shown that the formalized dataset performed slightly better, with a 0.009 difference in the F1-score. Details are presented in Table 5.

5. Ethics and privacy

In this study, users' privacy and general ethics were serious concerns, especially during the process of collecting and analyzing data from social media accounts. Boyd and Crawford (2012) write "...it is problematic for researchers to justify their actions as ethical simply because the data are accessible... The process of evaluating the research ethics cannot be ignored simply because the data are seemingly public" [30]. Although open discussions using Twitter differ from protected or private posts on Facebook, our main concern was to not cross the line between public and private posts. We always protected anonymity by replacing usernames with specific codes. Usernames are never published or used in the screening, classification and any manual assessment processes. As our concern deals with human resource development, we are aware that job applicants' social media usernames are commonly collected today and thus informed consent for profiling analysis was necessary for the process. It is also worth mentioning that we included disclaimers regarding the educational purposes of data collection and usage on the DISC test website.

6. Conclusion and future works

In this study, we tried to explore the possibility of conducting DISC analysis using social media posts written in Bahasa Indonesia, the mother tongue of Indonesia. Data collection tried to comply with the golden standard of profiling analysis conducted conventionally by experts using a DISC analysis instrument, which was used to collect textual and numerical information that is considered connected to a person's personality. A combination of NLP and statistical approach, complemented by a machine learning algorithm, has been proven effective in improving performance evaluation. We balanced the dataset using several resampling techniques, with SMOTETomek returning the best performance. Hyperparameter-tuned support vector classifier outperformed several supervised and ensemble learning algorithms, with an F1-score of 56.43%. This affirms that automatic personality classification from social media information in Bahasa Indonesia is feasible to be done and needs more in-depth further research. According to human resource experts, observation analysis of other social media platforms (such as Facebook, Instagram and LinkedIn) is also commonly practiced during the employee selection process. A combination

	Precision	Recall	F1-score	Support
Dominance	0.74	0.65	0.69	312
Influence	0.67	0.60	0.64	291
Stability	0.37	0.60	0.64	291
Conscientious	0.56	0.46	0.51	285
Accuracy			0.56	1160
Macro average	0.58	0.56	0.57	1160
Weighted average	0.59	0.56	0.57	1160

Table 5. Evaluation table of SVC using SMOTETomek resampling and hyper-parameter tuned confusion matrix on formalized text data

of textual and visual information derived from those platforms might be able to provide more comprehensive and better classification results. Text mining of new prospective workers' resumes may also give comparable insight, and we aim to achieve such insight soon in future works.

References

1. Kunte AV, Panicker S. Using textual data for personality prediction: a machine learning approach. 2019 4th International Conference on Information Systems and Computer Networks (ISCON). IEEE: 2019: 529-33.
2. Piedboeuf F, Langlais P, Bourg L. Personality extraction through linkedin. Canadian Conference on Artificial Intelligence; Springer. 2019. 55-67.
3. Szmigiera M. Most spoken languages in the world. 2021. [Online]. Available from: https://www.statista.com/statistics/266808/the-most-spoken-languages-worldwide/ (accessed 20 May 2021).
4. Adi S, Hartanto AD, Utami E, Raharjo S, Oyong I. The effect of data acquisition techniques in profiling analysis based on twitter. 2019 International Conference on Information and Communications Technology (ICOIACT); IEEE: 2019, 868-71.
5. Hartanto AD, Utami E, Adi S, Hudnanto HS. Job seeker profile classification of twitter data using the naïve bayes classifier algorithm based on the disc method. 2019 4th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE); IEEE: 2019, 533-36.
6. Jia J, Zhang P, Zhang R. A comparative study of three personality assessment models in software engineering field. 2015 6th IEEE International Conference on Software Engineering and Service Science (ICSESS); IEEE: 2015, 7-10.
7. Reynierse JH, Ackerman D, Fink AA, Harker JB. The effects of personality and management role on perceived values in business settings. Int J Value-Based Manag. 2000; 13(1): 1-13.
8. Marston WM, Emotions of normal people. Oxfordshire: Routledge; 2013, 158.
9. Kim YY, Baek YH, Park KH, Yoo JH, Jang ES. The effects of disc behavior styles of office workers on job satisfaction, organizational commitment and job performance. Korean J Occup Health Nurs. 2012; 21(2): 98-107.
10. Tabasum F, Ibrahim M, Rabbani M, Asif M. Impact of salesmen personality on customer perception and sales. Global J Manag Business Res. 2015; 14(8): 63-8.
11. Agodi JE, Ahaiwe E, Awah A. Salesman's personality trait and its effect on sales performance: study of fast moving consumer goods (fmcg) in abia state, Nigeria. J Econ Sustain Dev. 2017; 8(24): 81-8.
12. Utami E, Hartanto AD, Adi S, Oyong I, Raharjo S. Profiling analysis of disc personality traits based on twitter posts in Bahasa Indonesia. J King Saud Univ Com Info Sci. 2019; in press.
13. Adi GYN, Tandio MH, Ong V, Suhartono D. Optimization for automatic personality recognition on twitter in bahasa Indonesia. Proc Comp Sci. 2018; 135: 473-80.
14. Krismayer T, Schedl M, Knees P, Rabiser R. Predicting user demographics from music listening information. Multimed Tool Appl. 2019; 78(3): 2897-920.
15. Farnadi G, Sitaraman G, Sushmita S, Celli F, Kosinski M, Stillwell D, Davalos S, Moens MF, De Cock M. Computational personality recognition in social media. User Model User-Adap Int. 2016; 26(2–3): pp. 109-42.
16. Gu H, Wang J, Wang Z, Zhuang B, Su F. Modeling of user portrait through social media. 2018 IEEE International Conference on Multimedia and Expo (ICME); IEEE: 2018. 1-6.
17. Utami E, Hartanto AD, Adi S, Putra RBS, Raharjo S. Formal and non-formal Indonesian word usage frequency in twitter profile using non-formal affix rule. 2019 1st International Conference on Cybernetics and Intelligent System (ICORIS); IEEE: 2019. 1, 173-76.

18. Tadesse MM, Lin H, Xu B, Yang L.. Personality predictions based on user behavior on the facebook social media platform. IEEE Access. 2018; 6: 61 959.
19. Kunte A, Panicker S. Personality prediction of social network users using ensemble and xgboost. Progress in computing, analytics and networking; Springer: 2020. 133-40.
20. Tandera T, Suhartono D, Wongso R, Prasetio YL *et al.* Personality prediction system from facebook users. Proc Comp Sci. 2017; 116: 604-11.
21. Owen JE, Mahatmya D, Carter R. Dominance, influence, steadiness, and conscientiousness (DISC) assessment tool; Springer International Publishing: 2017, 1-4.
22. Purwarianti A, Andhika A, Wicaksono AF, Afif I, Ferdian F. Inanlp: Indonesia natural language processing toolkit, case study: complaint tweet classification. 2016 International Conference On Advanced Informatics: Concepts, Theory And Application (ICAICTA); IEEE: 2016, 1-5.
23. Liu B, Tsoumakas G. Dealing with class imbalance in classifier chains via random undersampling. Knowl Base Syst. 2020; 192: 105292.
24. Maldonado S, López J, Vairetti C. An alternative smote oversampling strategy for high-dimensional datasets. Appl Soft Comput. 2019; 76: 380-89.
25. Wang Z, Wu C, Zheng K, Niu X, Wang X. Smotetomek-based resampling for personality recognition. IEEE Access. 2019; 7: 129678-89.
26. Fernández A, Garcia S, Herrera F, Chawla NV. Smote for learning from imbalanced data: progress and challenges, marking the 15-year anniversary. J Artif Int Res. 2018; 61: 863-905.
27. He H, Ma Y, Imbalanced learning: foundations, algorithms, and applications. Hoboken: John Wiley & Sons; 2013.
28. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V *et al.* Scikit-learn: machine learning in python. J Machine Learn Res. 2011; 12: 2825-30.
29. Elhassan T and Aljurf M. Classification of imbalance data using tomek link (*t*-link) combined with random under-sampling (rus) as a data reduction method. Global J Technol Optim. 2016; S1(2).
30. Boyd D, Crawford K. Critical questions for big data: provocations for a cultural, technological, and scholarly phenomenon. Inf Commun Soc. 2012; 15(5): 662-79.

Appendix

Supplementary data

Supplementary data to this article can be found online at https://github.com/irwanOyong/aci-disc-supplementary-material

Corresponding author
Suwanto Raharjo can be contacted at: wa2n@akprind.ac.id