

Ensemble Learning Approach for Identifying Personality Traits based on Individuals' Behavior

Sri Lanka Journal of Social Sciences and Humanities
Volume 3 Issue 1, February 2023: 107-116
ISSN: 2773 692X (Online), 2773 6911 (Print)
Copyright: © 2023 The Author(s)
Published by the Faculty of Social Sciences and
Languages, Sabaragamuwa University of Sri Lanka
Website: <https://www.sab.ac.lk/sljssh>
DOI: <http://doi.org/10.4038/sljssh.v3i1.91>



K.M.G.S Karunarathna¹, M.P.R.I.R. Silva², and R.A.H.M. Rupasingha^{3,*}

^{1,2} Department of Information Technology, Sabaragamuwa University of Sri Lanka, Belihuloya, Sri Lanka.

³ Department of Economics and Statistics, Sabaragamuwa University of Sri Lanka, Belihuloya, Sri Lanka.

Received: 11 November 2022, **Revised:** 27 December 2022, **Accepted:** 16 January 2023.

How to Cite this Article: Karunarathana, K.M.G.S., Silva, M.P.R.I.R., & Rupasingha, R.A.H.M. (2023). Ensemble learning approach for identifying personality traits based on individuals' behavior. *Sri Lanka Journal of Social Sciences and Humanities*, 3(1), 107-116.

Abstract

Understanding human personality is essential for natural and social engagement. Arises a significant connection between users' personalities and their behavior. Our primary goal is to identify and classify individuals' personality traits based on their behaviors. Understanding personality types can help better understand preferences and potential differences between them. This study uses users' answers based on the questionnaire on personality to automatically identify the personality type based on behaviors. After pre-processing the data, we researched many classification techniques for automated recognition, including Naive Bayes, Support Vector Machine (SVM), Multilayer Perception (MLP), Random Forest, Logistic, and Decision Tree using a 10-fold cross-validation method. The second observational study combined all the algorithms using an Ensemble Learning algorithm, where by Vote algorithm. Accuracy, precision, recall, f-measure, and error values have been used to measure the systems' performance. According to the comparison analysis, SVM outperforms (85.8%) the other five personality trait detection algorithms. But after combing the five algorithms which contain the highest accuracy by Ensemble Learning algorithm obtained the highest performance (90.5%) than the SVM algorithm and obtained the highest recall, f-measure, and precision values, and the lowest error rates. It demonstrates that an ensemble learning approach that incorporates multiple distinct algorithms may yield greater accuracy than any one of its individual algorithms separately. Our findings are helpful for understanding how to manage and make a relationship with humans by predicting their personality earlier.

Keywords: Classification, Ensemble Learning, Human Personality, Machine Learning

INTRODUCTION

There are many different types of people around the world. They have different characteristics, behaviors, attitudes, etc. Based on their personality characteristics they also change with each other. There are so many characteristics in personality including honesty, brave, loyal, leadership, etc. People have demonstrated these personality traits in various ways. These personality characteristics can be used for various purposes such as being organized, various kinds of social situations getting decisions, etc. Because personality contains various types of characteristics, should be able to identify the personality type according to the purpose at that time. Therefore, identifying personality types can be an important task.

The well-known phrase "Face is a mirror of the mind" demonstrates that a person's appearance is determined by their inner qualities. Because a person's conduct and appearance describe his or her personality, an examination of appearance and behavior provides insight into personality features (Ilmini & Fernando, 2018). Some people are as transparent as a book.; friends appear to be ability to anticipate their actions by reading their thoughts and sentiments. Others persist a mystery even after years of

acquaintance (Colvin et al., 1993). In the world, there are various people. They differ from each other on many factors such as behavior, attitude, feelings, external appearance, etc. Additionally, we cannot say for certain that individuals have a personality in same way. As a result, it should thoroughly study them in order to at least understand them. Because of this, personality classification is important to study people to some extent and learn their personality traits.

Every day, people connect verbally and non-verbally in a variety of ways, from casual banter to serious dialogues (Park et al., 2020) such as various social settings like kindergarten, school, college, family, work teams, etc. (Talasbek et al., 2020). Therefore, it is important to identify the personality of the people. Personality is a dynamic and organized collection of a person's traits that impact that person's cognition, motivation, and behavior. And defining traits that influence how they will respond in various situations. Here, people's personalities are determined by how they connect with other people (Katiyar et al., 2020).

Human interaction is fundamentally influenced by personality (Iacobelli & Culotta, 2013). Personality can not

* Corresponding author: Tel.: +94 (71) 832 4740; Email: hmrupasingha@gmail.com
<https://orcid.org/0000-0003-3922-4290>

only anticipate and characterize a person's conduct but also includes their thoughts and feelings and influences their motivations, preferences, emotions, and even their physical health (Ahmad et al., 2020). Everyday interactions with others rely on our intrinsic ability to understand their behavior. By analyzing observable behavioral indicators, we can characterize others as being more talkative, assertive, social, or as being more or less susceptible to stress or rage (Lepri et al., 2016). Results from personality tests are frequently used in a variety of industries, including clinical psychology, social psychology, educational studies, Sentiment analysis, opinion mining, tailored healthcare, and personalized services (Wang et al., 2021) (Batrinca et al., 2012). And it is also useful in the areas of hiring new employees, career counseling, relationship counseling, and health counseling (Pratama & Sarno, 2016). As well as personality type has a significant impact on suitability for a certain employment, how well handle daily tasks, and even how satisfied with job as a whole.

When considering a person's personality, important to remember that some people tend to be outgoing and confident, others are oddballs and quiet, and yet others could be neither of these things at all. Some personality traits are favored by everyone, while others are despised since each personality is distinct. It possesses some traits that we either find alluring or repellent. As well as our world is one of heated competition. Having a strong personality makes you stand out to employers. On the other hand, each of the numerous models and theories relating to motivation and personality presents a unique viewpoint. However, a person's personality traits affect more than just how they perceive themselves. Therefore, it generates a problem how these various ways of feeling, thinking, and doing be measured?

Because there are so many variables involved, personality classification is become very important and complex (Keh & Cheng, 2019). But we can describe other people, attempt to predict their behavior, and modify our own behavior further

by using the form of intelligence (Lepri et al., 2016). In our study demonstrates how several categorization algorithms are used to automatically identify personality traits in various data.

Following Figure 1 shows the mapping between research questions and objectives.

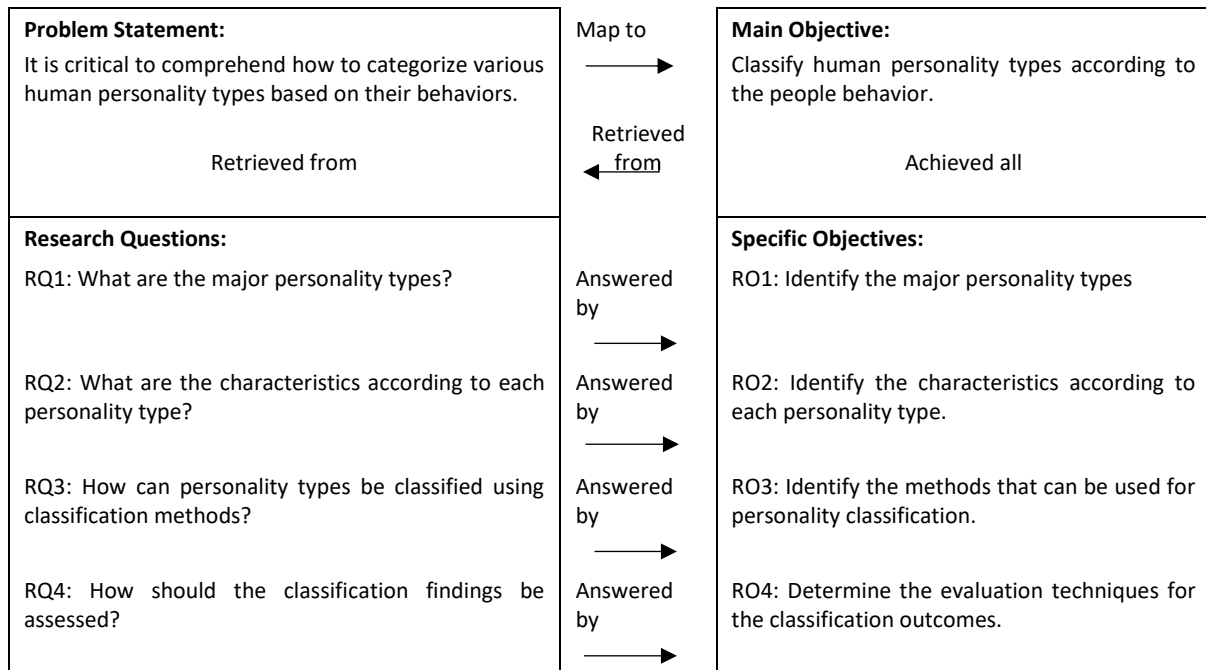
The proposed approach will help to recognize one's personality type and to understand their preferences and how or why they could differ. Understanding personality types can improve our ability to lead, persuade, cooperate, negotiate, and handle stress. As well as personality classification is playing a huge role when we want to identify our strengths and weakness of own self. And also knowing your personality type might be helpful in the job in many different ways. It may help with communication, leadership style, conflict resolution, knowing how other people think, mentoring others, enhancing sales skills, etc.

In contrast to single traits, personality types that consider how qualities interact with one another paint a more complete picture of a person's character. This study examines the personality traits of each person and thereby gives an idea of what kind of personality they have.

This paper's primary contribution is a performance analysis of various categorization using six classification algorithms and combine five classification algorithms among them with highest accuracy, using ensemble learning algorithm. Additionally, we go over how to create a system that accurately predicts personality based on responses to a questionnaire that individuals fill out. Measurement of the system's performance using precision, recall and f-measure and lowest error rate additional contribution made by this research.

The following is a description of the document's format: The similar work is described in Chapter 2, the suggested research approach is described in chapter 3, the findings and discussion of the proposed research are represented in chapter 4, and the research work is concluded and future directions for research are defined in chapter 5.

Figure 1 Mapping between research questions and objectives



LITERATURE REVIEW

Related Work

According to previous researches, personality is a combination of an individual's characteristics and behavior in dealing with various situations (Pratama & Sarno, 2016). One of the interesting traits that might be explored for adaptation is personality. A person's personality may be defined as a set of specifications that demand a pattern on the person's actions; this tendency remains constant throughout time and locations (Souri et al., 2018).

The researchers of (William, 2020) in machine learning, classification refers to the process of labeling given input transmitted data into predetermined groups. There are many popular categorization algorithms based on Bayes, trees, functions, or rules. This article examines classification techniques for identifying and forecasting personality using the HEXACO Model dataset, and it is supported by implementation performance. The process of identifying expressive, distinctive, and fascinating relationships, trends, and patterns in massive amounts of data utilizing pattern recognition technology as well as computational and analytical tools is known as data mining. Data mining encompasses data prediction and interpretation in addition to data gathering and processing. When people attempt to evaluate or create connections between various parts, errors frequently arise, making it difficult to develop answers to specific problems. Machine learning may be useful in solving these difficulties by improving the effectiveness of processes and computer designs.

The paper (Joo & Hwang, 2019) explains the involvement in the PAN 2019 shared task on author profiling, determining if the author of a tweet is a bot or a human, and identifying the author's gender in the case of a human for English and Spanish datasets. They study the complementarities of stylometry approaches and content-based methods in this research, proposing several ways for developing flexible features.

In (Karunarathna et al., 2022b), six algorithms are individually used to determine the personality types of humans. As well as in (Karunarathna et al., 2022a), the researchers have applied clustering algorithms to identify the personality behaviors.

According to this study (Ahmad et al., 2020), the daily use of digital devices with Internet access, such as tablets and smartphones, has increased dramatically in recent years, impacting Internet and social media network usage. When people use social networks, they disclose personal information that is broadcast to other users, providing organizations with useful information. Characterizing users based on their social media behavior is a new topic of study in Natural Language Processing (NLP), and this research will present an overview of how personality can be determined through online data.

The primary idea behind an ensemble methodology is to combine a series of models to fulfill the task, resulting in a better composite global model with more accurate and judgments than a single model. There are numerous methods for training an ensemble model to get a desired result. Diversity: The higher performance of ensemble models is mostly due to the employment of multiple inductive biases. Predictive performance: The individual inducer's predictive performance should be as high as feasible and at least as good as a random model (Rokch, 2019).

In (Boyd & Pennebaker, 2017) suggest a supplementary model that is based on a large data solution: word analysis. Language-based personality tests were utilized to capture/model lower-level personality processes that are more closely related to important objective organizational results. Furthermore, the growing availability of language data, as well as developments in statistical methodologies and technical capability, are rapidly opening up new possibilities.

In the study (Fallah & Khotanlou, 2015), a personality recognition system is used to automatically extract character attributes from persons handwritten writings. Their suggested method generates feature vectors by combining independent data with context information such as top margin value, word extraction, character sizes, line space, word space, word tilts, horizontal to vertical character ratio, and lie tilts. For categorization, the MLP neural network is applied.

The study (Stachl et al., 2020) looks at how well people's personality traits (measured at broad domain and narrow facet levels) may be predicted by six types of behavior: In a large sample, communication and social behavior, music consumption, app usage, mobility, overall phone activity, and day and night time activity were studied. The cross-validated results show which Big Five personality traits are predicted and which unique behaviors are associated with which characteristics, with communication and social behavior being the most predicting overall. The study (Stachl et al., 2019) demonstrates that personality trait levels can be predicted using data collected from off-the-shelf smartphones. The data set used in this study was made up of three independent datasets obtained in three separate investigations as part of the phone study research project at Ludwig Maximilians-Universität München. They present how variables may be used to predict self-assessments of the big five personality characteristics at the component and facet level using a machine learning technique (random forest, elastic net). The research (Kassarnig et al., 2018) is based on data collected from smartphones used as primary phones by the students for two years. They can directly measure the explanatory power of individual and societal factors since multi-channel data from a single population is accessible. The researchers discover that the most revealing measures of performance are based on social links, and that network indicators outperform individual traits in terms of model performance. Among individual traits, class attendance is the most important predictor in their findings show that considerable homophily and/or peer effects exist among university students.

Research Gap

Most of the existing research is considered based on the social network, usage of the internet, language, and handwriting patterns to identify personality. But in our study, we considered human personality behaviors in humans' day-to-day lives.

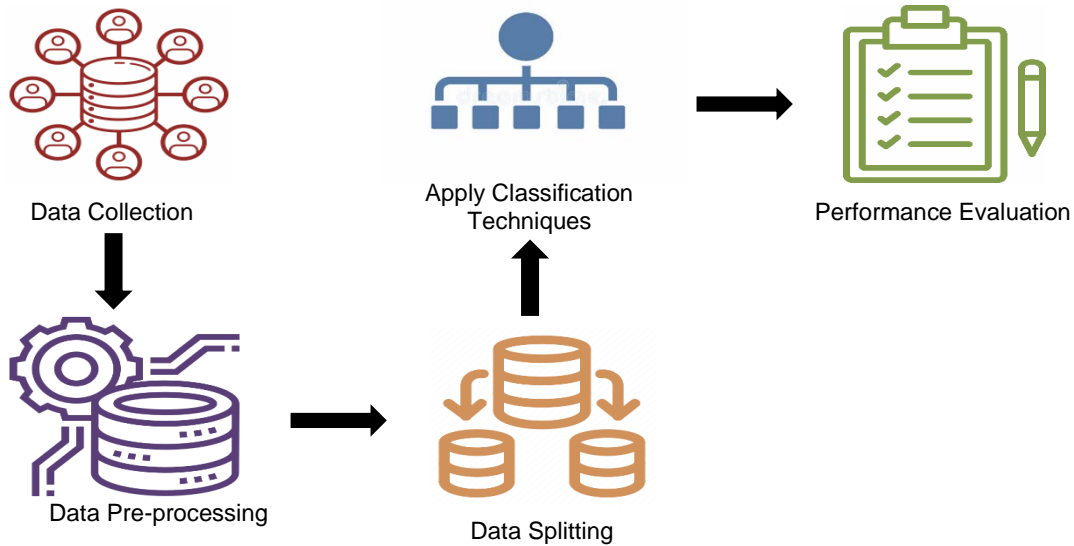
Furthermore, the data set we used was obtained through a questionnaire. In other recent research, the information has been obtained by observing the posts, links, and text patterns in social media networks. The information obtained from the observer may vary depending on the information seeker's state of mind, knowledge, and other factors. It may affect the information received. But through the questionnaire given to the respondent, the information can be obtained directly from the person. Therefore, the results

obtained from our study are more accurate than other study results.

Previous research used one or two classification algorithms in their approach. We increased it to six individual algorithms and also combined five algorithms using an ensemble learning algorithm to improve the results.

METHODOLOGY

The suggested automatic personality classification system applies specific data mining techniques to categorize the Figure 2: Proposed approach



personality type according to human’s behavior. With the guidance of the suggested model, user personality can be determined using answers given by the people. This system evaluates user characteristics and behaviors, and then creates its own patterns that assist anticipate the user’s personality and categorize all users into distinct groups.

The proposed research is useful for forecasting peoples' personalities. Following Figure 2 shows the proposed approach for the current study.

DATA COLLECTION

This was done using the secondary data set, which was received from the Kaggle website (Kaggle, 2022). The proposed methodology uses the responses from the

questionnaire that the participants provided. There are 62 questions in the survey. Those are regarded as attributes. The top 34 attributes are chosen after the attributes are ranked using the information gain ranking algorithm in Waikato Environment for Knowledge Analysis (WEKA) data mining tool. Following Table 1 shows the 34 attributes.

Table 1 Attributes of personality

Attributes of Personality		
You frequently make new friends	You appreciate witnessing heated debates	You are frequently the one who makes contact with acquaintances and starts events in your social circle
You are still troubled by faults you made in the past	Your moods can vary dramatically	It's easy to feel like you want to cry when you see other people crying
Even under intense pressure, you typically maintain your calm	You frequently find yourself taking action at the very last minute	You are more in charge of your emotions than the other way around
You would rather complete one assignment entirely before beginning another	You typically choose to be with people rather being by yourself	Your own work style is more similar to random bursts of energy than deliberate, regular efforts
It is easy for you to approach someone you find fascinating and start a discussion	It's simple to relate to someone whose experiences diverge greatly from your own	If your job required you to work alone most of the time, you would adore it
You tend to follow your intellect more often than your heart	You do not frequently question your decisions	You are drawn more to noisy, energetic locations than to quiet, private ones
You hardly ever give thought to how you come across to new acquaintances	You like visiting art galleries	You frequently feel overburdened
You enjoy reading and watching books and movies that force you to conceive of your own conclusion	You enjoy making a daily to-do list	You spend a lot of your free time researching a variety of arbitrary subjects that catch your attention

You have so many interests that it can be challenging to decide what to explore next	You try to avoid making calls	Frequently, a backup plan for a backup plan is made
In group settings, you resist taking on leadership responsibilities	You find things that are regarded controversial to be incredibly intriguing	You actually rarely make an effort to introduce yourself at social gatherings; instead, you choose to converse with people you already know
You believe that if people relied more on reason and less on emotion, the world would be a better place You could start to question your general knowledge and talents after making even a tiny error	You find deadlines difficult	You have a sensitive nature

In the original data set, there were 16 personality types considering their behaviors. In our study, depending on their personality traits, the personality types were condensed to 5 types.

- *The Supervisor* - Extraverted, sensing, thinking, and judging are its initials. This personality type is characterized by a person who feels energized by being near people (extroverted), who focuses on facts and information rather than concepts and ideas (sensing), who uses logic and reason to make decisions, and who values planning and organization over dynamic and unplanned (judging) (Owens, 2012).
- *The Commander* - Extraverted, intuitive, thinking, and judging are its initials. This kind of personality is characterized by someone who is energized by social interaction (extroverted), prefers thoughts and conceptions to facts and information (intuitive), basing judgments on reason and logic (thinking), and likes to be planned and arranged as opposed to fluid and spontaneous (Judging) (Owens, 2012)
- *The Inspector* - Its explains introverted, sensing, thinking, and judging. This personality type depicts someone who finds energy in solitude

(Introverted), likes facts and details to ideas and hunches (Sensing), based their judgments on rationality and logic (Thinking), and values planning and organization above being impulsive and adaptable (Judging) (Owens, 2012)

- *The Doer* - Extraverted, sensing, thinking, and perceiving is what it stands for. This personality type is characterized by extraversion (being energized by other people), sensing (concentrating on information and specifics rather than concepts and ideas), thinking (making choices based on reason and logic), and preference for spontaneity and flexibility over planned and organized behavior (Perceiving) (Owens, 2012).
- *The Idealist* - Introversion, intuition, feeling, and perception make up this type. This is introverted, intuitive, feeling-based, and focused on ideas and concepts rather than details and facts. Additionally, they are more likely to be unplanned and flexible than organized and rigid (Perceiving)(Owens, 2012).

Following Figure 3 shows the summary of the five personality types.



Figure 3 personality types

Pre-processing

Before putting the data into model training, through data pre-processing, the initial dataset is changed. into a useful and typical dataset to improve data quality and prevent dirty data. The pre-processing procedure entails ranking the attributes. The dataset consists of responses from 59999 individuals to the questionnaire. The questionnaire included 62 questions as attributes. As well as in original data set

focused on 16 target variables as personality types. From these 62 original data sets the top 34 attributes are chosen after the attributes are ranked using the information gain ranking algorithm. From the 16 personality types selected 5 personality types after studying those personality types. The 5 major and common personality types were selected here. Based on those 5 personality types, 200 data were randomly

selected for each personality type. Finally, a total of 1000 datasets were applied to the model.

Classification

Using the WEKA data mining tool, the classification process is applied to the pre-processed data set. The prediction model is constructed in order to identify personality traits. The algorithms such as Random Forest, Naive Bayes, Decision Tree (J48), Logistic, SVM, and MLP were all applied to the data set. The 5-fold and 10-fold cross-validation is used in order to validate of the classification model.

- *Random Forest* - Using their average for classification and majority vote for regression, it constructs decision trees from various data. The Random Forest technique's capacity to handle data sets with both continuous variables, as in regression, and categorical variables, as in classification, is one of its most crucial aspects. (Sruthi, 2021).
- *Naïve Bayes* - It is a classification method built on the Bayes Theorem and predictor independence. The premise behind a Naive Bayes classifier is that the presence of one feature in a class has no influence on the presence of any further features. The Naive Bayes model is a very useful tool for very large data sets and is simple to construct. Naive Bayes is well renowned for doing better than even the most sophisticated classification approaches, in addition to being simple. (Saini, 2021b).
- *Decision Tree* – A decision tree, where the internal node of the tree represents attributes and A class label relates to the leaf node.(Sharma, 2021).
- *Logistic* - Can be applied to model a class or event's probability. When the result is binary or dichotomous and the data can be separated linearly, it is used. (Saini, 2021).
- *SVM* - Each data point is represented as a point in n-dimensional space by the SVM method, and the value of each feature corresponds to the value of a certain coordinate. Simply explained, Support

vectors are each observation's coordinates. (Ray, 2017).

- *MLP* - MLP needs several parameters to accommodate multidimensional input since it aims to recall patterns in sequential data. (Pillai, 2020).
- *Ensemble Learning* - By combining the predictions from different models, it is a comprehensive meta-machine learning technique that seeks to improve predictive performance. (Brownlee, 2020)

RESULTS & DISCUSSION

The Kaggle secondary dataset (Kaggle, 2022), which contains data from 1000 people, provided the experimental data for this work. The five personality types are utilized to determine the personality. Utilizing the WEKA data mining tool, the system's performance had been assessed. It is proposed to employ accuracy, precision, recall, and f-measure as assessment metrics and the lowest error rates are calculated using MAE (Mean Absolute Error) and RMSE (Root Mean Square Error).

The processed dataset was split into 90% training and 10% test sets using 10-fold cross validation method. In addition, we obtained comparison between 5-fold cross validation with the 10- fold cross validation and compared five combination rules in ensemble learning algorithm. The test platform featured Microsoft Windows 10 on a computer with an Intel Core i5-8250U processor clocked at 1.60GHz and 4GB of RAM.

Confusion Matrix

The other evaluations were based on the confusion matrix results. The confusion matrix values for the seven machine learning techniques are shown in Table 2.

Here, *FP* is the quantity of actual negative cases that the classifier divides into positive ones and *FN* is the quantity of actual positive instances that the classifier divides into negative ones. *TP* is the number of actual positive instances in a batch of data that the classifier has identified as positive cases, *TN* is the total number of actual negative cases that the classifier has classified as negative cases, and *FP* is the quantity of actual negative cases that the classifier divides into positive ones (sensors).

Table 2 Confusion matrix values

Classification Algorithms	Confusion Matrix			
	TP	FP	FN	TN
Ensemble Learning	33	40	82	845
SVM	81	87	119	713
Naïve Bayes	84	90	134	692
Random Forest	38	34	148	780
Decision Tree	185	225	15	575
Logistic	62	109	186	643
MLP	90	98	108	704

Accuracy of Classification Algorithms

The accuracy of the classification algorithms on the identifying personality traits are observed and discussed in

this section. Equation (1) illustrates the calculation of classification accuracy

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

Figure 4 lists outcomes for accuracy of the each and every classification algorithm.

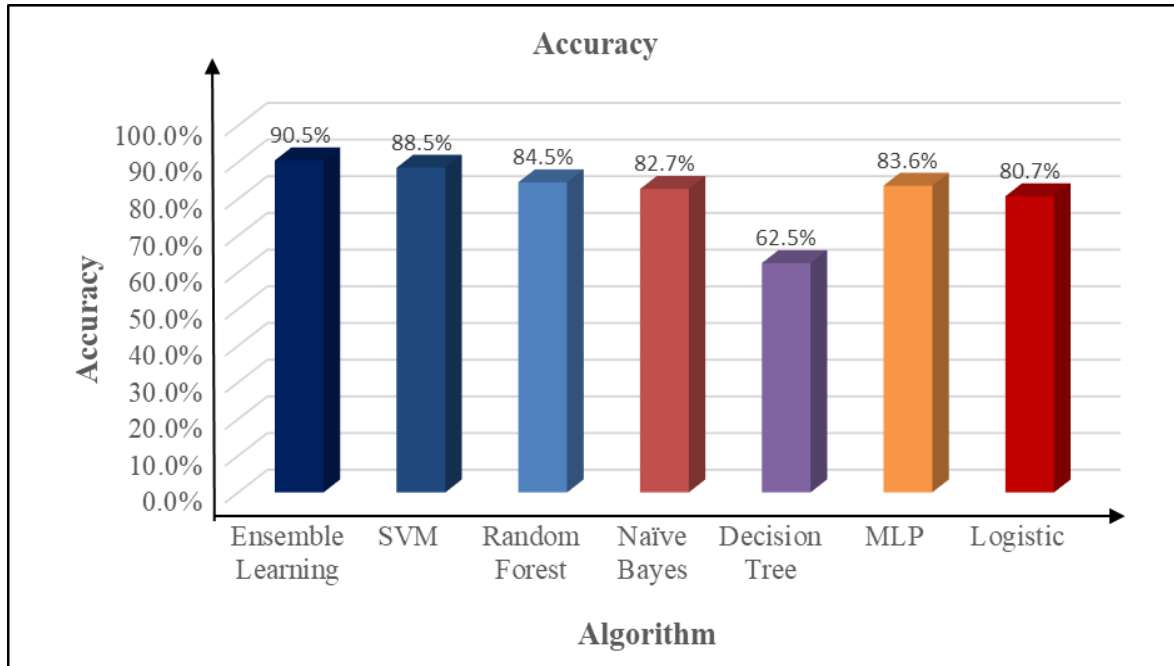


Figure 4 Accuracy of the algorithms

Precision, Recall & F-measure Values

The F1 score was also used to evaluate how accurate the classification model was. It took into account the classification model's precision and recall, This may be considered a harmonic average of the model's recall rate and accuracy.

The following equations (2), (3), and (4) detail the computations for the metrics of precision, recall, and F-measure for each of the seven methods.

$$Precision = \frac{P_s}{P_x} \tag{2}$$

$$Recall = \frac{P_s}{P_{stx}} \tag{3}$$

$$F_{measure} = \frac{2 * Precision * Recall}{Precision + Recall} \tag{4}$$

Here, P_{stx} is the total number of relevant members that comprise the specified-cluster in the corpus, P_x is the total number of members that comprise a specific cluster, and P_s is the total number of relevant members that comprise a particular cluster.

Following Table 3 shows the evaluation results for seven classification algorithms.

Table 3. Evaluation results for precision, recall & f-measure

Classification Algorithm	Precision	Recall	F-measure
Ensemble Learning	0.905	0.905	0.905
Naïve Bayes	0.830	0.827	0.827
SVM	0.885	0.885	0.885
Random Forest	0.846	0.845	0.845
Decision Tree (J48)	0.626	0.625	0.625
Logistic	0.807	0.807	0.807
MLP	0.836	0.836	0.836

MAE & RMSE Error Values

We determined the MAE and RMSE for each of the seven algorithms, as follows in (5) and (6).

$$MAE = \frac{1}{T} \sum_{x=1}^T |p_{vx} - M_{vx}| \tag{5}$$

$$RMSE = \sqrt{\frac{1}{T} \sum_{x=1}^T (p_{vx} - M_{vx})^2} \tag{6}$$

Following Figure 5 and 6 shows the values for lowest error rates for seven classification algorithms.

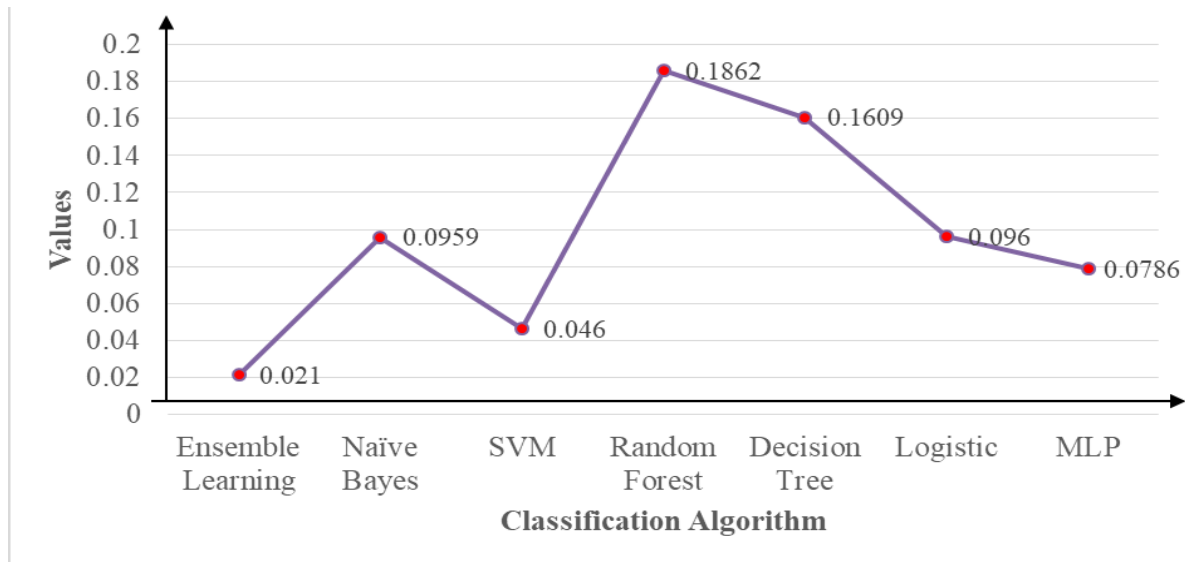


Figure 5 MAE values

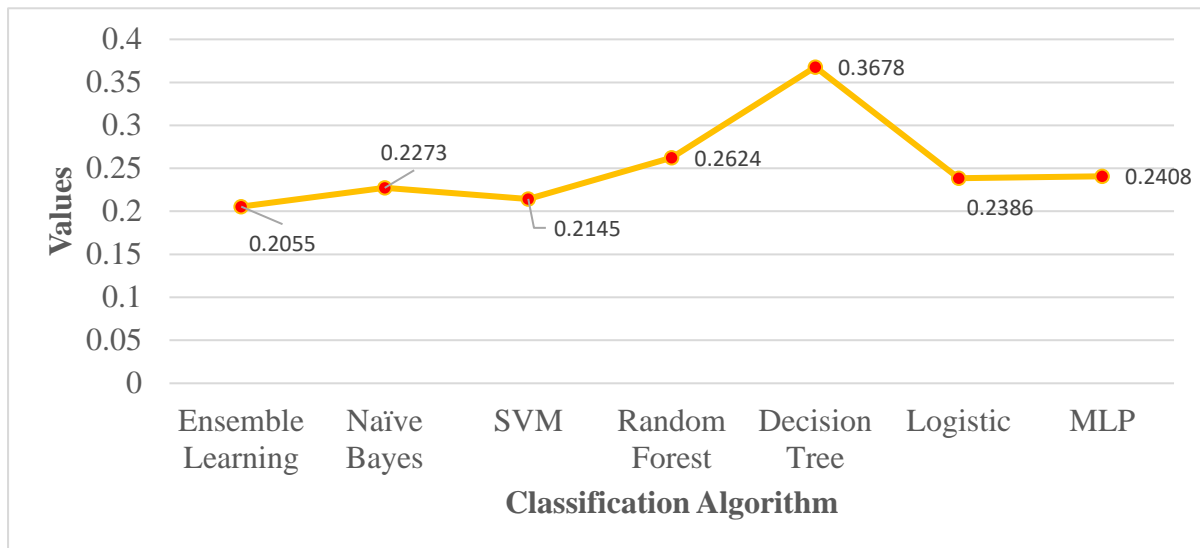


Figure 6 RMSE values

5-fold Cross Validation vs 10-fold Cross Validation

Both 5-fold cross validation and 10-fold cross validation were employed in our research experiment. Table 4 and 5

display the accuracy, precision, recall and f-measure results for both cross validations.

Table 4 Accuracy for 5-fold vs 10-fold cross validation

Fold	Ensemble Learning Algorithm		Accuracy
	Test Data	Training Data	
10	10%	90%	90.5 %
5	20%	80%	85.8 %

Table 5 Values for precision, recall & f-measure for 5-fold vs 10-fold cross validation

Fold	Ensemble Learning Algorithm		Precision	Recall	F-measure
	Test Data	Training Data			
10	10%	90%	0.905	0.905	0.905
5	20%	80%	0.858	0.858	0.858

Combination Rules of the Five Machine Learning Algorithms

The five algorithms with highest accuracy are used for the ensemble learning method. Namely. SVM, Random Forest,

Naive Bayes, MLP and Logistic classification algorithms are used. The accuracy results for each combination rule used in the Ensemble Learning algorithm are shown in Table 6.

Table 6 Accuracy of combination rules

Combination Rules	Accuracy
Average of Probabilities	85.8%
Product of Probabilities	67.2%
Majority Voting	90.5%
Minimum Probability	67.2%
Maximum Probability	81.7%

DISCUSSION

We show and evaluate the classification algorithm findings on the issue of classifying personality traits. 10-fold cross-validations were used throughout testing. From the individual algorithms, SVM had the highest accuracy of the six examined algorithms in the cross-validation testing, with an average accuracy of 88.5% as shown in Figure 4. In the second experiment, after combining five classification algorithms with highest accuracy, using the ensemble learning method whereby Vote algorithm performed the highest accuracy of 90.5% than SVM.

Table 3 represented the performance metrics of the seven classification algorithms. This confirmed that the ensemble learning algorithm is the best classification algorithm, because, the precision (0.905), recall (0.905), and f-measure (0.905) values of the ensemble learning obtained the highest values. The lowest error rate (MAE – 0.021 & RMSE- 0.2055) was also obtained by the ensemble learning algorithm. Figures 5 and 6 illustrated it. Another experiment of this research is compared 10-fold cross-validation with 5-fold cross-validation. Table 4 and 5 shows the accuracy and performance matrix for cross-validation methods. From that 10-fold cross-validation showed the highest accuracy and best performance matrix.

The special feature of the Ensemble Learning algorithm is combination rules. In this study, we experimented with five combination rules in Table 6. From that majority voting combination rule performed best at 90.5%. When a model uses majority voting, it forecasts (votes), with the final forecast being the one it garnered a majority of the votes for each test case. (Demir, 2016).

CONCLUSION

This article is based on the information of personality behaviors and characteristics of the human. Our main objective is to determine and categorize people's personality traits in many sectors based on their behaviors. The main conclusion of this study is Ensemble Learning showed the best accuracy with the highest values for precision, recall, and f-measure using 10-fold cross-validation. According to the evaluation results, the Ensemble Learning algorithm is the best way for predicting human behavior. It also demonstrates that combining such algorithms and generating predictive findings for future forecasts is more effective than using individual algorithms. After collecting information about a person, we can apply the collected data to this model. Then we can get an insight into which personality type he/she belongs to. That way, they will be able to identify their personality traits and choose fields related to them. They can create the environment they need for their career path as well as their life in society. Also, after studying another person, we can find out which personality type that person belongs to.

We intend to explore this approach with deep learning techniques by increasing the size of the data set. And also, expected to enhance the personality data types and personality attributes for more evaluations.

REFERENCES

- Ahmad, H., Asghar, M. Z., Khan, A. S., & Habib, A. (2020). A systematic literature review of personality trait classification from textual content. *Open Computer Science*, 10(1), 175–193. <https://doi.org/10.1515/comp-2020-0188>
- Batrinca, L. M., Lepri, B., Mana, N., & Pianesi, F. (2012). Multimodal recognition of personality traits in human-computer collaborative tasks. *ICMI'12 - Proceedings of the ACM International Conference on Multimodal Interaction*, 39–46. <https://doi.org/10.1145/2388676.2388687>
- Boyd, R. L., & Pennebaker, J. W. (2017). Language-based personality: a new approach to personality in a digital world. *Current Opinion in Behavioral Sciences*, 18, 63–68. <https://doi.org/10.1016/j.cobeha.2017.07.017>
- Brownlee, J., [Jason Brownlee]. (2020, October 23). A Gentle Introduction to Ensemble Learning. *Machinelearningmastery*. Retrieved November 15, 2022, from <https://machinelearningmastery.com/what-is-ensemble-learning/>
- Demir, N., [Necati Demir]. (2016). Ensemble Methods: Elegant Techniques to Produce Improved Machine Learning Results. *Kdnuggets*. Retrieved November 17, 2022, from <https://www.kdnuggets.com/2016/02/ensemble-methods-techniques-produce-improved-machine-learning.html>
- Fallah, B., & Khotanlou, H. (2015). Detecting features of human personality based on handwriting using learning algorithms. *Advances in Computer Science: An International Journal*, 4(6), 31–37.
- Iacobelli, F., & Culotta, A. (2013). Too neurotic, not too friendly: Structured personality classification on textual data. *AAAI Workshop - Technical Report*, WS-13-01, 19–22. <https://doi.org/10.1609/icwsm.v7i2.14472>
- Ilmini, W. M. K. S., & Fernando, T. G. I. (2018). Computational personality traits assessment: A review. *2017 IEEE International Conference on Industrial and Information Systems, ICIIS 2017 - Proceedings*, 2018-January, 1–6. <https://doi.org/10.1109/ICIINF5.2017.8300416>
- Joo, Y., & Hwang, I. (2019). Author Profiling on Social Media: An Ensemble Learning Model using Various Features. *September*, 9–12.
- Mehta, A., [Anshul Mehta], Guzman, J., [Jose Guzman], & Oumaima, B. hammed, [Ben hammed Oumaima]. (2022, February). Personality classification Data: 16 Personalities. *Kaggle*. Retrieved September 5, 2022, from <https://www.kaggle.com/datasets/anshulmehtakaggl/60k-responses-of-16-personalities-test-mbt>
- Karunarathna, K. M. G. S., Silva, M. P. R. I. R., & Rupasingha, R. A. H. M. (2022). Clustering Human Personality Based on Persons' Behaviour. In *3rd International Women in Engineering Symposium (WIESYMP)*. WIESYMP.
- Karunarathna, K. M. G. S., Silva, M. P. R. I. R., & Rupasingha, R. A. H. M. (2023). Human Personality Classification Using Supervised Machine Learning Algorithms. In *1st International Conference on Social Sciences and Languages*, Sabaragamuwa University of Sri Lanka (pp. 102–109). ICSSL.
- Kassarnig, V., Mones, E., Bjerre-Nielsen, A., Sapiezynski, P., Lassen, D. D., & Lehmann, S. (2018). Academic performance and behavioral patterns. *EPJ Data Science*, 7(1), 1–16. <https://doi.org/10.1140/epjds/s13688-018-0138-8>
- Katiyar, S., Walia, H., & Kumar, S. (2020). Personality Classification System using Data Mining. *ICRITO 2020 - IEEE 8th International*

- Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions), 1020–1023. <https://doi.org/10.1109/ICRITO48877.2020.9197803>
- Keh, S. S., & Cheng, I. C. (2019, July). Myers-Briggs Personality Classification and Personality-Specific Language Generation Using Pre-trained Language Models. ArXiv
- Lepri, B., Staiano, J., Shmueli, E., Pianesi, F., & Pentland, A. (2016). The role of personality in shaping social networks and mediating behavioral change. *User Modeling and User-Adapted Interaction*, 26(2–3), 143–175. <https://doi.org/10.1007/s11257-016-9173-y>
- Owens, M. (2021). MYERS & BRIGGS' 16 PERSONALITY TYPES . Truity. <https://www.truity.com/page/16-personality-types-myers-briggs>
- Park, J., Lee, S., Brotherton, K., Um, D., & Park, J. (2020). Identification of speech characteristics to distinguish human personality of introversive and extroversive male groups. *International Journal of Environmental Research and Public Health*, 17(6). <https://doi.org/10.3390/ijerph17062125>
- Pillai, J. (n.d.). A Simple Overview of Multilayer Perceptron (MLP). Analyticsvidhya. Retrieved November 22, 2022, from, <https://www.analyticsvidhya.com/blog/2020/12/mlp-multilayer-perceptronsimple-overview/>.
- Pratama, B. Y., & Sarno, R. (2016). Personality classification based on Twitter text using Naive Bayes, KNN and SVM. *Proceedings of 2015 International Conference on Data and Software Engineering, ICODSE 2015*, 170–174. <https://doi.org/10.1109/ICODSE.2015.7436992>
- Sruthi, E. R. (2021, June 17). Understanding Random Forest. Analyticsvidhya. <https://www.analyticsvidhya.com/blog/2021/06/understanding-random-forest/>
- Ray S (2017), "Understanding Support Vector Machine (SVM) Algorithm from Examples (Along with Code)", 2017, available at <https://www.analyticsvidhya.com/blog/2017/09/understaing-support-vector-machine-example-code/> Accessed on November 23, 2021.
- Lior Rokach. (2019). *Ensemble learning: pattern classification using ensemble methods*. World Scientific Publishing.
- Saini, A. (2021, August 3). Conceptual Understanding of Logistic Regression for Data Science Beginners. Analyticsvidhya. <https://www.analyticsvidhya.com/blog/2021/08/conceptual-understanding-of-logistic-regression-for-data-science-beginners/>
- Saini, A. (2021). Naive Bayes Algorithm: A Complete Guide for Data Science Enthusiasts Analyticsvidhya. <https://www.analyticsvidhya.com/blog/2021/09/naivebayes-algorithm-a-complete-guide-for-data-science-enthusiasts/>
- Sharma A (2021), "Machine Learning 101: Decision Tree Algorithm for Classification", available at <https://www.analyticsvidhya.com/blog/2021/02/machine-learning-101-decision-tree-algorithm-for-classification/> (Accessed: 20 November 2021)
- Souri, A., Hosseinpour, S., & Rahmani, A. M. (2018). Personality classification based on profiles of social networks' users and the five-factor model of personality. *Human-Centric Computing and Information Sciences*, 8(1). <https://doi.org/10.1186/s13673-018-0147-4>
- Stachl, C., Au, Q., Schoedel, R., Buschek, D., Völkel, S. T., Oldemeier, M., Ullmann, T., Hussmann, H., & Bischl, B. (2019). Behavioral Patterns in Smartphone Usage Predict Big Five Personality Traits. 1–24.
- Stachl, C., Au, Q., Schoedel, R., Gosling, S. D., Harari, G. M., Buschek, D., Völkel, S. T., Schuwerk, T., Oldemeier, M., Ullmann, T., Hussmann, H., Bischl, B., & Bühner, M. (2020). Predicting personality from patterns of behavior collected with smartphones. *Proceedings of the National Academy of Sciences of the United States of America*, 117(30), 17680–17687. <https://doi.org/10.1073/pnas.1920484117>
- Talabek, A., Serek, A., Zhaparov, M., Moo-Yoo, S., Kim, Y. K., & Jeong, G. H. (2020). Personality classification experiment by applying k-means clustering. *International Journal of Emerging Technologies in Learning*, 15(16), 162–177. <https://doi.org/10.3991/ijet.v15i16.15049>
- Wang, X., Sui, Y., Zheng, K., Shi, Y., & Cao, S. (2021). Personality classification of social users based on feature fusion. *Sensors*, 21(20). <https://doi.org/10.3390/s21206758>
- William, P. (2020). Evaluating Efficacy of Classification Algorithms on Personality Prediction Dataset. 19(4), 3362–3375.