International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

Fusion Methods for Forecasting Personality Employing Machine Learning

[1]Dr. Karishma Desai, [2]Girish Kalele, [3]Rakesh Kumar Yadav, [4]Bhuvana Jayabalan, [5]Manish Kumar Goyal

Submitted: 29/12/2023 **Revised**: 05/02/2024 **Accepted**: 13/02/2024

Abstract: Personality prediction has gained substantial attention in various fields, such as psychology and human-computer interaction. Traditional approaches to personality prediction often rely on self-report questionnaires, which can be time-consuming, subject to biases, and limited by the respondent's self-awareness. In this paper, we proposed a Support Vector Machine Stacker (SVMS) for the categorization and prediction of personalities, leveraging the fusing of data and classification levels. Our model uses Learning to transfer in natural language processing by utilizing two the BERT annulment, one of the top pre-trained language models. By incorporating these powerful language models, our proposed approach demonstrates promising results in personality prediction. In our experiments, we compare the SVMS against other methods to evaluate the performance of the proposed model. The evaluation metrics include accuracy, precision, recall, and F1-score. The results demonstrate that the SVMSachieves competitive performance in predicting personality traits.

Keywords: *Personality, Support Vector Machine Stacker (SVMS), BERT and ULMFiT, Machine learning (ML)*

1. Introduction

Studying personality is essential for comprehending human behaviors and processes of decision-making. It includes a range of behaviors, attitudes, and ways of thinking that help to define a person [1]. Personalized advertising, psychological evaluation, and interaction between humans and computers are just applications where personality characteristic forecasting can have a big impact. Recent developments in ML and Natural Language Processing (NLP) methods present new opportunities for accurately predicting characters from textual information [2]. The goal of the project is to create a fusion predictor to forecast personality characteristics by combining the strengths of NLP and ML. To improve the reliability and accuracy of psychology forecasts, the fusion classifiers integrate numerous categorization models. It can accurately capture many facets of a person's personality and offer a more thorough analysis using several ML methods [3]. A person's thoughts, emotions, and language habits are reflected in textual information such as social media posts, review sites, or written works. NLP techniques are used to extract these elements. Sentiment analysis, word usage trends, syntactic patterns, and semantic representations are some of these aspects. The fusion classification can find associations between specific linguistic patterns and personality characteristics by analyzing and interpreting these literary aspects [4].

The fusion classifier is trained on labeled datasets using ML methods such as Support Vector Machines (SVM), Random Forests, or Neural Networks. The Big Five Personality Traits (Agreeableness, Extraversion, Openness, Conscientiousness, and Neuroticism) and other well-known personality test methodologies were used to derive the personality trait labels for the text samples in the labeled collections [5]. Utilizing group learning methods, a fusion classifier integrates the results from multiple classification models to arrive at its final prediction on an individual's character attributes [6]. By utilizing the benefits of each model, this strategy seeks to minimize the drawbacks and inefficiencies that could result from relying just on one classification.

The findings of this study could improve our comprehension of how NLP and ML can be utilized for predicting personality characteristics from the textual information.

Using information and classifier-level combination, they suggested a Support Vector Machine Stacker (SVMS) for personality prediction and categorization in this study. Two of the most popular pre-trained language models, BERT annulment and TL, are used by our model to process natural language. Our suggested method shows promising results in personality prediction by combining these potent models of language.

The rest of this paper is as follows: part 2 is a literature

[1]*Department of ISME, ATLAS SkillTech University, Mumbai, Maharashtra, India, Email Id- karishma.desai@atlasuniversity.edu.in, Orcid Id- 0009-0009-5171-9588*
[2]*Chitkara University, Rajpura, Punjab, India, girish.kalele.orp@chitkara.edu.in, https://orcid.org/0009-0003-7801-8737*
[3]*Maharishi University of Information Technology, Lucknow, India, Email Id- rkymuit@gmail.com, Orcid Id- 0000-0002-0151-4981*
[4]*Jain (Deemed to be University), Bangalore, Karnataka, India, Email Id- j.bhuvana@jainuniversity.ac.in, Orcid Id- 0000-0002-8372-6311*
[5]*Department of Computer Science & Application, Vivekananda Global University, Jaipur, India, Email Id- manish.kumar.goyal@vgu.ac.in, Orcid Id- 0009-0001-6769-2860*

review; Part 3 contains the proposed method explained; Part 4 includes the results and analysis; and Part 5 Discusses the conclusion.

2. Related works

In study [7], a personality identification system was developed that analyses offline handwritten signature images of a person to predict several aspects of that person's personality. It was observed that many researchers have worked on personality analysis of a person using different methodologies and obtaining promising results with their dataset and standard dataset. The main focus of the proposed work was on feature-level fusion algorithms and random forest classifiers. Study [8] proposed a method to obtain character-based classification without relying on predefined dictionaries or contextual information—research entirely convolutional network architecture that generates symbols streaming any length from the handwritten text. Incoming units are normalized to a canonical interpretation during preparation, eliminating the requirement for pricey recurring symbol misalignment corrections. Using a conventional multilayer network to measure and then recalculate input streams to ecclesiastical participation initially, they describe a novel offline recognition of handwriting technique.The article [9] proposed an optical character recognition (OCR) for identifying and information entry in various industries. Through ML, they improve and simplify existing banking processes, which can be accomplished by autonomously performing inspections. The method used, known as Handwritten Character Recognition, integrates recognition of patterns with ML to produce an optical characters recognizer for numerals and capital letters that can be written with the hand and reproduced. The executed, split images of all characters are sent to the skilled model to get the desired outcomes.Study [10] describes various ML approaches to the standard MBTI personalities dataset, called mbtikaggle, to categorize the text into different personality traits, which was the main focus of the study. The findings indicate that while the accuracy of the XGBoost classifier is exceptional, the scores obtained by every classifier over every characteristic are acceptable.Study [11] applied an unsupervised learning of features method in the paper to offer a first look into group-level personality detection. They pioneer the NRL approach and rigorously compare our model to other well-known research. The findings support using an unsupervised method and a group viewpoint when using psychology assessment.

A study [12] suggested a personality recognition model that classifies data across several labels using multi-text interpretation and psychological variables. They introduce a neural network that combines the linguistic and emotional components of the text with the pre-trained BERT model to create a multi-label character recognition model. To analyze emotions, they use a sentiment dictionary, which adds some capacity to clarify characters. Study [13] a novel semantic-enhanced personality recognition neural network (SEPRNN) model that aims to reduce reliance on feature development, enables the same model to adjust to identifying five distinct personality characteristics without changing the framework itself and makes use of neural techniques and atomically aspects of papers to construct a vectorial word-level illustration for personality trait acceptance. The experimental findings show that the suggested strategy significantly outperforms numerous starting points for multi-labeled personality characteristics. A study [14] was conducted to provide a mathematical method for ML that is less complicated and more effective at predicting personalities. Cronbach's Alpha is used to test the survey created using the HEXACO model to determine the validity of each variable, and factors are considered before the ML algorithm is performed. The outcome shows that the HEXACO model algorithm developed for personality prediction and the HR manager report have a significant major association.

2.1 Transfer learning in NLP

Natural language processing (NLP) has undergone a revolution thanks to TL, which has made it possible to build incredibly accurate models even from sparsely labeled data. Bidirectional Encoder Representations from Transformers BERT and Universal Language Model Fine-tuning ULMFiT are well-liked TL techniques in NLP. Transformer-based models are used in BERT (Bidirectional Encoder Representations from Transformers). In a variety of NLP tasks, such as sentence categorization, named entity recognition, question resolution, and text production, it has excelled remarkably. Contrary to conventional models that only take into account right-to-left or left-to-right contexts, BERT uses a bidirectional method that enables it to gather contextual information from both the left and right contexts of a word. Using a sizable corpus of internet-sourced unlabeled text, BERT is pre-trained. The model gains the ability to anticipate masked words in a sentence and recognize the connections between several sentences in a document during pre-training. BERT may acquire intricate contextual representations of words and sentences. The pre-trained BERT model may be fine-tuned with very few labeled data for diverse NLP applications and provides a solid foundation for language interpretation by capturing rich contextual information. Thanks to its outstanding performance and adaptability, it has become a prominent option for TL in NLP.

ULMFiT (Universal Language Model Fine-tuning): This transfer learning (TL) strategy uses language modeling. It focuses on enhancing a language model already trained for certain tasks while utilizing the broader linguistic knowledge acquired during pre-training. A useful and efficient method for transferring information from a pre-trained language model to a target task is provided by ULMFiT. The ULMFiT improves performance even with little labeled data by capturing general language patterns and semantics by pre-training on a large corpus. By facilitating (TL), BERT and ULMFiT have considerably improved the field of NLP. These models have paved the way for more precise and reliable NLP applications, making it simpler to create effective language models for various tasks without needing enormous volumes of labeled data.

3. Proposed Method

Transferable techniques for learning backed by LMs who have already received training have been adopted by ULMFiT and BERT. The mechanisms used by each model vary, but they all rely on word incorporation and adjusting methods. These models contribute significantly to NLP tasks. But figuring out which approach to use for a specific NLP assignment is difficult. Detailed information will determine the exact result. Fusion techniques can be applied at various levels, including the stories of data, features, classifiers, and decisions. Fusion techniques could boost classification performance in general and increase forecast precision.

To improve the overall efficacy of character estimation, our suggested approach combinesclassifier-level and data-level fusion methodologies. The pre-trained ULMFiT, BERT, were accepted. They utilized the Essays and My Character databases to improve the proposed model. The goal is to make the suggested temperament prediction framework more effective and universally applicable, regardless of the information's sources. Figure 1 depicts the high-level structure of the suggested model for predicting temperament.

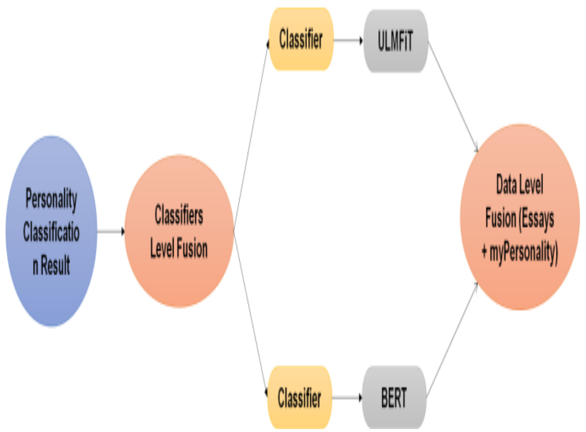

Fig 1: Overview of proposed model.

3.1 Dataset

Both datasets that were pooled to create the corpus for this study are frequently used in personality forecasting studies; more information about the datasets is provided in Table 1 [16].

Table1: Dataset Description

Title	Essays	my personality
sources	Students	Facebook
Records	2,468 essays	9,918 status
Words	1,900,000	143,600

Our study was due to a portion of the original My Personality dataset. Both datasets have been marked with the character qualities of the authors: Table 2 presents the pair of the binary label (yes/no) values for "EXT, NEU, AGR, CON, and OPN."

Table 2: Major Personality Characteristics

Five Qualities	Summary
AGR	Is the person unreliable, convoluted, meager, and boastful, or are they straightforward, generous, and modest?
NEU	Is the person secure and confident or sensitive and uneasy?
CON	Is the person orderly and tidy or disorganised and careless?
OPEN	Is the person innovative and inquisitive rather than conservative and dogmatic?
EXT	Is the person friendly, chatty, active, or quiet and solitary?

The overall distribution of both datasets based on the type of personality attribute is shown in Tables 3 and 4. All of the characteristics in the Essays dataset have equilibrium, as shown in Table 3. However, NEU and OPN are the only qualities that lack peace, as shown in Table 4.

Table 3: Essays Dataset Distribution

Value	Yes	No
EXT	1278	1192
NEU	1236	1235
AGR	1312	1158
CON	1255	1215
OPN	1218	1198

Table 4: Distribution of Personality Dataset

Value	Yes	No
EXT	4210	5709
NEU	3718	6202
AGR	5269	4648
CON	4558	5362
OPEN	7372	2549

3.2 Support Vector Machine Stacker (SVMS)

In this paper, we integrate Support vector machine with stacking classifier for predicting personality.

Support vector machines (SVMs) are supervised learning models that explore data and discover data samples needed for categorization. They also have associated learning methodologies. An SVM training method creates a framework that classifies current data sets into one stage or the other using a collection of labeled training data about one or both classes, creating a non-probabilistic discrete linear classification algorithm from the other. A Support Vector Machine (SVM) model represents samples as points across space, drawn so that the data of the various classes are divided by as wide a gap as allowed. Then, fresh specimens are projected into a type made based on that same space and projected into it. Model for Predictingpersonalityusing SVM. SVMs may effectively conduct asymmetric categorization in addition to linear classification by translating input data into feature spaces with high dimensions using the kernel method.

Creating test and training data sets is a step in the categorization process. One desired value and multiple characteristics are present in every sample in the learning set. The goal of SVM is to build a model that forecasts the expected values of the test data using the training information.

Assume a training set where and, The following efficiency must be resolved using support vector machines.

$$min_{x,a,\varepsilon} \frac{1}{2} x^S x + D \sum_{j=1}^{m} \quad \xi_j \tag{1}$$

subject to

$$z_j\left(x^S \emptyset \left(w_j\right) + a\right) \geq 1 - \xi_j, \xi_j \geq 0 \tag{2}$$

Using the function, sets of training xi are converted into an additional dimension area in this instance. In this higher dimension space, SVM finds a linear separation hyperplane with the largest margins. The error item is produced when C > 0.But a new character has been introduced by researchers. The following four fundamental personalities are:

Linear:

$$B\left(w_j, w_i\right) = w_j^S w_i \tag{3}$$

Polynomial:

$$B\left(w_j, w_i\right) = (\gamma w_j^S w_i + q)^c, \gamma > 0 \tag{4}$$

Gaussian (Radial basis function (RBF))

$$B\left(w_j, w_i\right) = exp\ exp\ \left(-\gamma\ \frac{||w_j - w_i||^2}{2\sigma^2}\right), \gamma > 0 \tag{5}$$

Sigmoid:

$$B\left(w_j, w_i\right) = tang\ \left(\gamma w_j^S w_i + q\right) \tag{6}$$

Here,γ, q, and care personality parameters.

The Gaussian kernel has been used in the current study.

3.3 Stacking classifier

The arrangement classifier can be a helpful fusion approach in Natural Language Processing (NLP) when Support Vector Machines (SVMs) are added to ensemble systems. The Stacking Classifier makes the final forecast by combining the results of various basic classification algorithms, including SVMs.Several SVM models are trained on the NLP information as part of the Stacking Classifier with SVM in the setting of NLP. Alternative subsets of the data or alternative hyperparameter combinations can be used to train each SVM model. In the arrangement structure, these SVM models are regarded as the fundamental classifiers.

Following SVM, the development of models and forecasts are fed into a meta-classifier. The meta-classifier can combine the SVM models' predictions and provide a final prediction based on the collected data. The SVM models can work together and utilize their unique strengths in the stacking technique. The NLP data may have distinct patterns or focuses in each SVM model, and the meta-classifier can learn to combine their results properly. The assembly classification can increase the overall predictive accuracy and robustness in NLP applications by merging numerous SVM models and using the knowledge gained from their projections.

4. Result

To achieve the best efficiency for the two classification algorithms, we tested with several fine-tuning methods using ULMFiT and BERT. These strategies performed similarly and provided the most favorable outcomes for the two algorithms. In the ULMFiT model for BERT, we kept the classifier and included a classification algorithm for feature categorization. Each test involved training five separate networks with the same architecture on all five personality variables.

Accuracy is a metric for how well a model forecasts the results or labels of a specific dataset. It is computed by dividing the total number of forecasts made by the number of accurate predictions, commonly expressed as a percentage. Figure 1 depicts the accuracy outcome. The value obtained for our suggested approach is superior to current methods (SVMS). This demonstrates that our recommended strategy successfully identifies a personality.

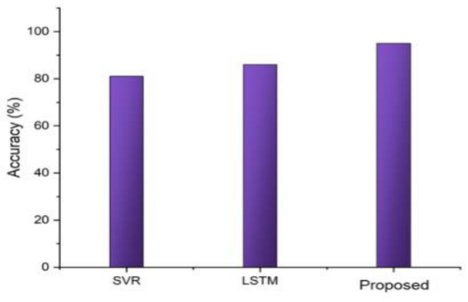

Fig 1: Result of accuracy

A statistical parameter called precision assesses the categorization or prediction model's accuracy. Out of all occurrences anticipated to be favorable (true positives plus false positives), it calculates the percentage of correctly predicted positive instances (true positives). The value obtained using our suggested approach is superior to that current methods . This demonstrates that our proposed strategy successfully identifies a personality.

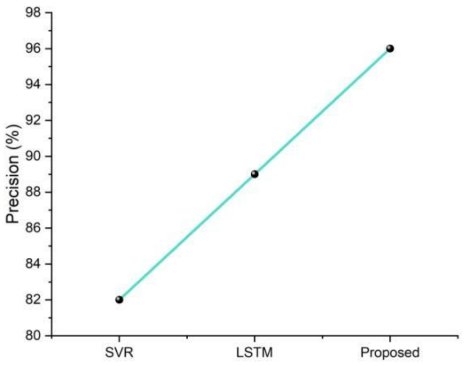

Fig 2: Result of precision

Recall, also called sensitivity or true positive rate, quantifies the percentage of positive occurrences the model properly accepted. The ratio of true positives to the total of true and misleading negatives is computed. Compared to the existing approaches (SVR,LSTM) the value produced using our proposed method (SVMS) is higher. This shows that our proposed way is effective in locating a personality.

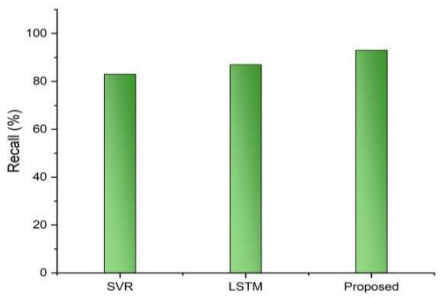

Fig 3: Result of recall

The F1-score is sometimes referred to as the F-score or F-measure. F1-score is defined as the harmonic mean of precision. When compared to the existing approaches, our proposed method is better. This shows that our proposed way is effective in locating a personality.

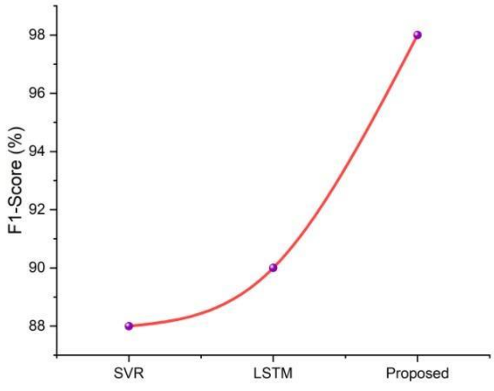

Fig 4: Result of f1-score

5. Conclusion

To forecast and categorize personalities, this research developed a Support Vector Machine Stacker (SVMS) technique.[17] Conventional character prediction methods that rely on self-report surveys take a lot of time, are subject to prejudice, and are only as good as the respondents themselves. The suggested way makes use of TL in processing natural languages and two pre-trained models of language, BERT, and ULMFiT, to get over these limitations. The stacking classifier with the SVM technique showed promising outcomes in predicting personality by combining these potent models of language. Compared to previous approaches, the model performed competitively in accuracy, precision, recall, and F1 score. This suggests that the SVMS method works well in identifying personality traits. The proposed strategy creates possibilities for more study and applications in fields like psychology and interaction between humans and computers. These findings show that TL and fusion approaches are mature for extracting personality traits from literature. In terms of future work, we intend to test the application of the findings in this research to additional expressive categories and subjective variables such as sentiment, emotion, and mood in order to develop a flexible model applied in the social network domain.

References

[1] Mehta, Y., Majumder, N., Gelbukh, A. and Cambria, E., 2020. Recent trends in deep learning-based personality detection. Artificial Intelligence Review, 53, pp.2313-2339.

[2] Gornale, S.S., Kumar, S. and Hiremath, P.S., 2021.

Handwritten signature Biometric data analysis for personality prediction system using machine learning techniques. Transactions on Machine Learning and Artificial Intelligence, 9(5), pp.1-22.

[3] López-Santillán, R., González, L.C., Montes-y-Gómez, M. and López-Monroy, A.P., 2023. When attention is not enough to unveil a text's author profile: Enhancing a transformer with a wide branch. Neural Computing and Applications, pp.1-20.

[4] Amin, J., Sharif, M., Gul, N., Raza, M., Anjum, M.A., Nisar, M.W. and Bukhari, S.A.C., 2020. Brain tumor detection by using stacked autoencoders in deep learning. Journal of medical systems, 44, pp.1-12.

[5] Pandeya, Y.R. and Lee, J., 2021. Deep learning-based late fusion of multimodal information for emotion classification of music video. Multimedia Tools and Applications, 80, pp.2887-2905.

[6] Nguyen, H.H., Ngo, V.M., Le, T.T.P. and Van Nguyen, P., 2023. Do investors' personalities predict market winners? Experimental setting and machine learning analysis. Heliyon, 9(4).

[7] Gornale, S.S., Kumar, S. and Hiremath, P.S., 2021. Handwritten signature Biometric data analysis for personality prediction system using machine learning techniques. Transactions on Machine Learning and Artificial Intelligence, 9(5), pp.1-22.

[8] Ptucha, R., Such, F.P., Pillai, S., Brockler, F., Singh, V. and Hutkowski, P., 2019. Intelligent character recognition using fully convolutional neural networks. Pattern recognition, 88, pp.604-613.

[9] Srivastava, S., Priyadarshini, J., Gopal, S., Gupta, S. and Dayal, H.S., 2019. Optical character recognition on bank cheques using 2D convolution neural network. In Applications of Artificial Intelligence Techniques in Engineering: SIGMA 2018, Volume 2 (pp. 589-596). Springer Singapore.

[10] Panhwar, M.A., Memon, K.A., Abro, A., Zhongliang, D., Khuhro, S.A. and Memon, S., 2019, July. Signboard detection and text recognition using artificial neural networks. In 2019 IEEE 9th international conference on electronics information and emergency communication (ICEIEC) (pp. 16-19). IEEE.

[11] Mehta, Y., Majumder, N., Gelbukh, A. and Cambria, E., 2020. Recent trends in deep learning-based personality detection. Artificial Intelligence Review, 53, pp.2313-2339.

[12] Khan, A.S., Hussain, A., Asghar, M.Z., Saddozai, F.K., Arif, A. and Khalid, H.A., 2020. Personality classification from online text using machine learning approach. International journal of advanced computer science and applications, 11(3).

[13] Panhwar, M.A., Memon, K.A., Abro, A., Zhongliang, D., Khuhro, S.A. and Memon, S., 2019, July. Signboard detection and text recognition using artificial neural networks. In 2019 IEEE 9th international conference on electronics information and emergency communication (ICEIEC) (pp. 16-19). IEEE.

[14] Wu, X., Luo, C., Zhang, Q., Zhou, J., Yang, H. and Li, Y., 2019. Text Detection and Recognition for Natural Scene Images Using Deep Convolutional Neural Networks. Computers, Materials & Continua, 61(1)

[15] Abdullah, S.M.S.A., Ameen, S.Y.A., Sadeeq, M.A. and Zeebaree, S., 2021. Multimodal emotion recognition using deep learning. Journal of Applied Science and Technology Trends, 2(02), pp.52-58.

[16] Soundar Rajan, S. & Venkateswari, P. (2021). Construction of Artificial Minds with a Probability Learning Model. Technoarete Transactions on Industrial Robotics and Automation Systems (TTIRAS), 1(1), 11-14. https://doi.org/10.36647/TTIRAS/01.01.A003

[17] Jayanthi, R,. & Sunethra, B. (2022). Review on Quantum Computers in Machine Learning. Technoarete Transactions on Advances in Computer Applications (TTACA), 1 (1), 20-24. doi: doi.org/10.36647/TTACA/01.01.A005