Tec Empresarial
P-ISSN: 1659-2395; E-ISSN: 1659-3359

OPTIMIZING AUTOMATED PERSONALITY INSIGHT: A HYBRID MACHINE LEARNING APPROACH

Pravin Amin
Sr. Manager, Veritas Technologies LLC, Roseville, MN, USA

Abstract: This scholarly article explores enhancing automated personality prediction using combined machine learning methodologies. Utilizing the Myers-Briggs Type Indicator (MBTI) dataset sourced from Kaggle is central to this investigation, specifically focusing on refining the accuracy of predicting the Judging-Perceiving (J/P) dichotomy. Past research has underscored notable hurdles in forecasting this aspect, frequently resulting from challenges such as data leakage and model inefficiencies. In order to confront these barriers, a range of machine learning algorithms, specifically Logistic Regression, Random Forest, K-Nearest Neighbors (KNN), and XGBoost are examined in this analysis. The findings indicate that ensemble models, especially XGBoost, perform superior to traditional single-model approaches in terms of precision and reliability.

Additionally, advanced data preprocessing methods like TF-IDF vectorization and SMOTE are integrated to address data imbalance issues and further elevate model efficacy. By amalgamating these diverse modelling techniques, this research establishes a sturdy framework for automated personality forecasting, supplying valuable insights for fields such as psychology, marketing, and human resources. By bridging critical gaps in existing methodologies and proposing innovative remedies for more precise and dependable personality prediction, this paper makes a significant contribution to the academic domain.

Keywords: Automated Personality Prediction, Machine Learning, MBTI, Ensemble Modeling, Text Classification

1. Introduction

1.1 Background:

Personality prediction is pivotal across multiple sectors, including education, employment, interpersonal relationships, health, and overall well-being [1]. Diverse methodologies, like facial recognition software, sentiment mining, and AI algorithms, have been applied to project personality traits accurately [2][3][4]. These predictions benefit personnel selection, job performance assessment, and enhancing interactions between humans and computers. The capacity to forecast personality traits is a valuable tool for individuals preparing for job interviews or public speaking. However, it also aids recruiters and hiring managers in making informed choices during selection. In general, the precise prediction of personality traits using diverse methods provides opportunities to enhance decision-making processes and mitigate subjective biases in assessing individuals for various objectives.

1.2 Problem Statement:

Current personality prediction models encounter various limitations. Primarily, they heavily depend on pre-trained language models, overlooking sentiment information within psycholinguistic features [5]. Secondly, specific models encounter challenges with training duration and the capacity to grasp the true essence of words, resulting in a contextual loss [6]. Moreover, conventional measurement mechanisms for personality traits often need to be more practical, impeding their integration into predictive models. Moreover, a lack of consensus exists regarding psycholinguistic features, resulting in insufficient sentiment analysis [7]. Lastly, the selection of covariates in personality-outcome relationships can significantly influence the robustness of the models, underscoring the necessity for meticulous consideration in model development [8].

1.3 Goals:

The principal objective of amalgamating diverse modelling strategies for personality prediction is to enhance the precision and efficiency of Automatic Personality Prediction (APP) systems [9]. By integrating a variety of modalities like audio, text, and video data through a multimodal fusion technique, a more comprehensive understanding of human personality traits can be achieved, surpassing the capabilities of each modality individually [10]. Additionally, the goal is to enhance predictive accuracy beyond the conventional Five-Factor Model (FFM) by leveraging supervised machine learning techniques for dimensionality reduction and constructing a more anticipatory personality summary, known as the "Predictive Five" (PF) [11]. Furthermore, the amalgamation of traditional psycholinguistic features with language model embeddings in deep learning models aims to accomplish cutting-edge performance in personality prediction while offering insights into language features' influence on personality models [12].

2. Literature Review:

2.1 Previous Work:

Prior research has extensively delved into personality prediction by utilizing machine learning methodologies. These investigations have emphasized a variety of factors such as the comprehensibility of predictive models [13], determining the optimal level within the personality hierarchy for the creation of predictive algorithms [14], incorporating social network analysis in conjunction with passive sensing technologies to model personality states [15], and the advancement of deep learning systems for automated personality prediction utilizing multimodal characteristics [16]. These scholarly endeavours have illustrated the efficacy of machine learning in forecasting personality traits, improving the models' precision, comprehensibility, and dependability, and exhibiting potential applications in disciplines such as psychology, psychiatry, and digital forensics. Nigam et al. (2012) demonstrated the effectiveness of using a Semi-Supervised Support Vector Machine (SVM) in a bug classification model, achieving high classification accuracy with a combination of labeled and unlabeled data [17]. This approach highlights the potential of semi-supervised techniques in handling complex classification tasks with limited labeled data [18]. The number of training documents is important in formation of word sets used to determine the class of a new document. The greater number of word sets from training documents reduces the possibility of failure to classify a new document [19].

2.2 Gaps in Research:

The following are the primary deficiencies it aims to rectify in this research:

1. **Integration of Models**: A significant gap in current research is the isolated utilization of diverse machine learning models without delving into their potential synergies. This dissertation investigates how amalgamating various modelling strategies can augment the efficacy of personality prediction systems. By incorporating models such as LSTM, Transformers, BERT, CNN, and RNNs, the study explores whether a hybrid approach could result in superior accuracy and dependability compared to single-model methodologies.
2. **Enhancement of Model Performance:** There is often a tendency to focus on employing a singular algorithm for personality prediction without thorough comparison or optimization across different models. This study not only contrasts individual models but also delves into optimizing their amalgamation for enhanced performance, considering elements like hyperparameter optimization and adjustments in model architecture.
3. **Processing of Textual Data:** The dissertation concentrates on refining how textual data is processed and understood in the context of personality prediction. Numerous studies may need to fully exploit the intricacies of natural language processing techniques, which are crucial for comprehending the nuances of human language. The research endeavours to enhance these techniques within the framework of integrated models to refine the predictive capabilities of the systems.
4. **Utilization of Advanced Machine Learning Techniques**: While certain studies may rely on fundamental machine learning techniques for personality prediction, this dissertation pushes boundaries by employing more advanced methodologies such as deep learning and ensemble techniques. This could address the gap in leveraging state-of-the-art technology to elevate the sophistication and precision of predictive analytics in personality assessment.
5. **Assessment Metrics:** The dissertation strives to utilize and formulate comprehensive evaluation metrics to gauge the efficacy of integrated models. The efficient evaluation of model performance, particularly in a comparative context across varied strategies, aids in identifying optimal practices and methodologies for personality prediction.

By rectifying these deficiencies, the dissertation makes a significant contribution to the progression of the field of personality prediction. It proposes novel techniques that have the potential to enhance the accuracy and reliability of predictions (Nomura et al,2021), thus benefiting applications in fields like psychology, marketing, human resources, and beyond [20].

2.3 Theoretical Framework:

The Cross-Industry Standard Process for Data Mining (CRISP-DM) is a widely recognized framework utilized for managing ethical concerns in data science projects [21]. Having its roots in the latter part of the 1990s, CRISP-DM remains a prominent standard in data mining and projects related to knowledge discovery [22]. It aims to tackle the obstacles encountered in the field of social sciences by offering a standardized approach for examining extensive amounts of unstructured data, thereby augmenting the productivity and quality of research endeavours[23]. Moreover, an extension of CRISP has been made to create a sophisticated software known as CRISP, which assists in the detection, fusion, and categorization of molecular data sourced from biological samples utilizing GC × GC–TOFMS technology [24]. This framework accentuates the significance of longitudinal studies about dementia care giving, with a focus on the evolving requirements of caregivers and individuals who have dementia as they navigate through the course of the illness [25]. This framework is extensively utilized in various domains, such as industrial machinery monitoring, cryptocurrency fraud detection, and ethics management in data science projects [26]. It presents a methodical approach to overseeing data science projects and assisting in developing, assessing, and implementing models [27]. Moreover, enhancements have been made to CRISP-DM to confront emerging issues like biases related to fairness in machine learning applications, resulting in the introduction of Fair CRISP-DM. This framework assists professionals in identifying and alleviating biases and contributes to the body of knowledge on machine learning advancement and equity. The rationale behind using CRISP-DM stems from its flexibility across various fields, structured methodology ensuring the triumph of projects, and continual adaptation to address current challenges in data science projects.

In this research, CRISP-DM guides the integration of various machine learning models, such as LSTM, CNN, BERT, and XGBoost, to explore their combined effectiveness in predicting personality traits. Each phase of CRISP-DM helps manage different aspects of this integration, ranging from initial data analysis and model selection based on theoretical and empirical evidence to the evaluation of combined models against established benchmarks [21].

The application of CRISP-DM in this research justifies its use based on the framework's structured, iterative, and comprehensive nature [28]. It highlights its suitability for managing complex, innovative projects that aim to push the boundaries of current knowledge in personality prediction.

3. Methodology

3.1 Dataset Description

The analysis dataset used in this research concerns the MBTI, where personality traits are assessed through the Myers-Briggs Type Indicator (MBTI). The primary emphasis is on forecasting the Judging-Perceiving (J/P) dichotomy [29]. Past research has illuminated notable obstacles when effectively predicting this particular facet, predominantly attributed to information seepage within the Personality Forum Cafe dataset, leading to excessively positive outcomes [26]. Drawing parallels with the findings of Nigam et al. (2017), this study carefully selects algorithms that balance processing efficiency and accuracy in predicting personality traits, ensuring that the chosen model can handle large datasets effectively [21]

3.2 Data Collection and Preprocessing

We sourced the MBTI Personality Types 500 Dataset from Kaggle to ensure robust and accurate predictions. This dataset includes approximately 106,000 preprocessed posts categorized into 16 personality types. Preprocessing steps involved:

- Removing punctuation, stop words, and URLs
- Applying lemmatization to standardize text
- Splitting the dataset into a 70:30 ratio for training and testing, respectively

3.3 Exploratory Data Analysis

Includes dataset overview, distribution of personality types, post-length distribution, word frequency analysis, and word clouds.

3.4 Feature Engineering

It involves adding new columns with individual personality traits and preprocessing steps like lemmatization and removing MBTI terms from posts.It involves adding new columns with individual personality traits and preprocessing steps like lemmatization and removing MBTI terms from posts**.**

3.5 Data Splitting

The dataset is segmented into training (70%) and testing (30%) sets for analysis and model building.

3.6 Model Selection and Evaluation

We compared five machine learning algorithms to identify the most effective model for predicting the J/P dichotomy. The models evaluated include:

1. Logistic Regression
2. Random Forest
3. K-Nearest Neighbors (KNN)
4. XGBoost

3.7 Feature Extraction

Feature extraction was conducted using TF-IDF vectorization to convert textual data into numerical format suitable for machine learning models. This process ensured that the most relevant textual features were utilized in model training.

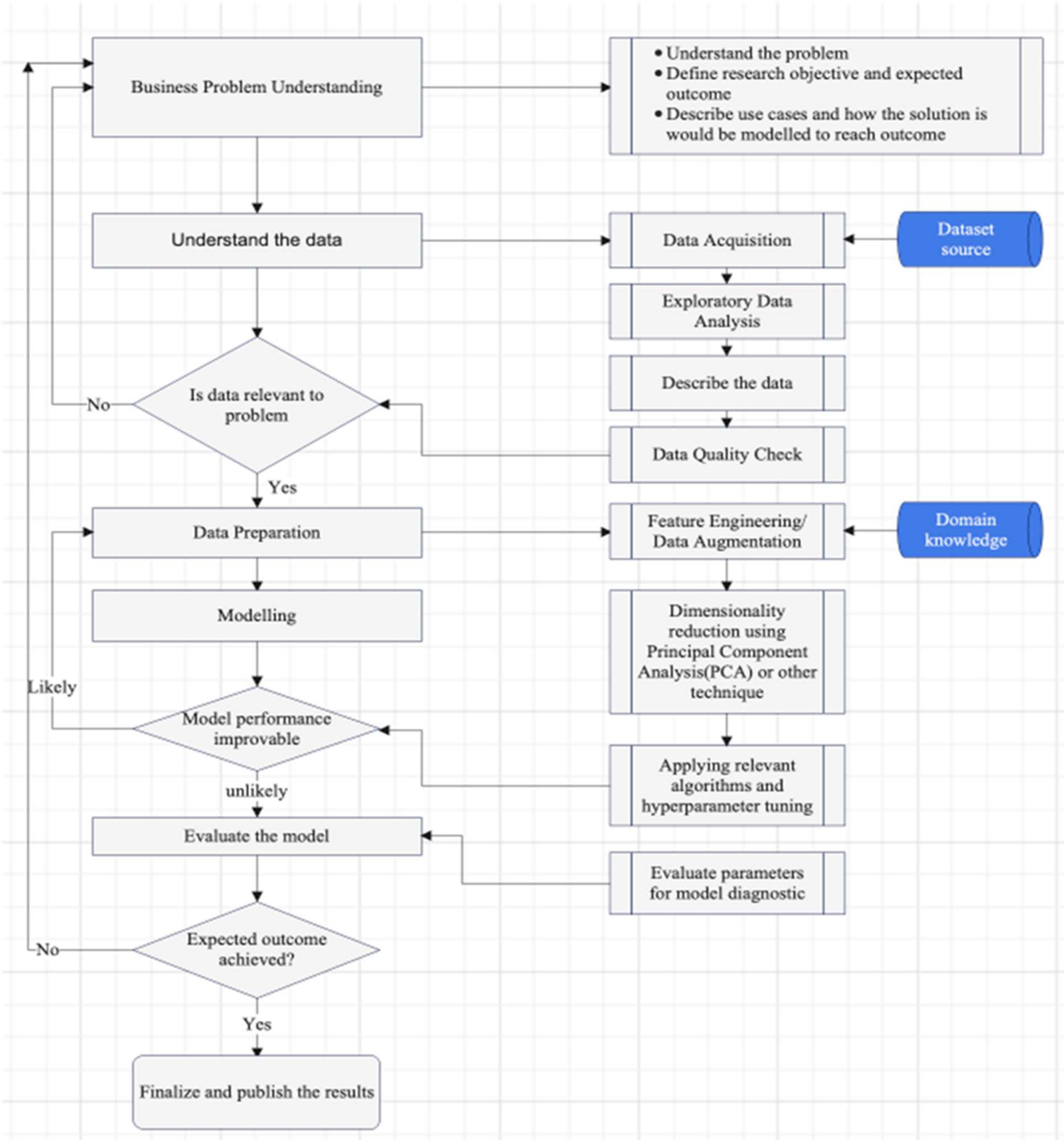

Figure 1. A diagrammatic representation of methodology followed.

3.8 Handling Data Imbalance

To address the class imbalance in the dataset, we employed techniques such as the Synthetic Minority Over-sampling Technique (SMOTE), which generates synthetic samples for minority classes to balance the dataset and improve model performance.

3.9 Model Training and Hyperparameter Tuning

Each model underwent extensive training and hyperparameter tuning to optimize performance. The training process involved:

- Splitting the data into training (70%) and testing (30%) sets

- Utilizing grid search and cross-validation for hyperparameter tuning
- Evaluating models based on accuracy, precision, recall, F1-score, and ROC-AUC

The assessment and comparison of models were grounded on their respective performance indicators. The XGBoost model emerged as the most efficacious in forecasting the J/P dichotomy, showcasing heightened levels of accuracy and dependability.

4. Result and Evaluation

4.1 Model Evaluation

Models will be evaluated based on various parameters as below:

Confusion Matrix: A machine learning classification problem with two or more possible classes to the output is a good candidate for a Confusion Matrix evaluation. This table has four possible permutations of expected and observed values [24].

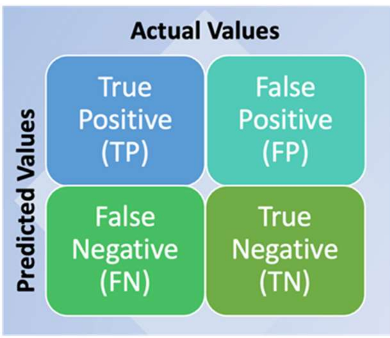

Figure 1: Confusion Matrix

4.2 Evaluation of Results on Train and Test Data

Performance-based analysis is conducted, with the best-performing model chosen based on metrics like Accuracy, Sensitivity, Specificity, Precision, F1-score, and ROC.

4.3 Logistic Regression Classification

Results in table 1 show detailed precision, recall, and f1-scores for the Introvert/Extrovert, Intuition/Sensing, Feeling/Thinking, and Judging/Perceiving classifications.

The logistic regression model performed well on this binary classification task, with the following results:

Precision: 81% for 'Introvert', 64% for 'Extrovert'
Recall: 95% for 'Introvert', 27% for 'Extrovert'
F1-score: 87% for 'Introvert', 38% for 'Extrovert'

4.4 Random Forest Classification

Accuracy and visualization of TPR against FPR for different personality types are provided in Table 2 and 3.

The Random Forest model, an ensemble of decision trees, was built and evaluated with the following configuration:

- n_estimators: 100
- max_depth: [3, 5, 7, 10, 12, 15]
- max_features: [0.05, 0.10, 0.15, 0.2]
- criterion: ["gini", "entropy"]

The model showed strong performance across different personality types: Introvert Train-Test Accuracy

Train Accuracy: Ranging from 95% to 99% across various types. Test Accuracy: 85% to 98%.

4.5 K-Nearest Neighbor Technique

Results for K=1 and K=2 are discussed, indicating overfitting scenarios in table no 4 and 5.
K=1
Accuracy: 44% on test data, indicating overfitting due to significantly higher train accuracy. K=2
Accuracy: 44% on test data, similar to K=1, with improvements needed to avoid overfitting.

4.6 XGBoost Classification

The XGBoost model showed promising results in table no 6, with test accuracy close to train accuracy.
Introvert | Extrovert Classification Report
Train accuracy: 88%
Test accuracy: 78%
F1-score: 70% for 'Introvert', 87% for 'Extrover

Table 1: Introvert | Extrovert Classification Report using Logistic Regression

	Precision	Recall	f1-score		precision	Recall	f1-score
IE Train Data				**IE Test Data**			
0	0.81	0.95	0.87	**0**	0.8	0.95	0.87
1	0.64	0.27	0.38	**1**	0.62	0.26	0.37
accuracy			0.79	**accuracy**			0.78
macro avg	0.73	0.61	0.63	**macro avg**	0.71	0.6	0.62
weighted avg	0.77	0.79	0.76	**weighted avg**	0.76	0.78	0.75

Table 2: Train Test Accuracy for a person with the common trait of Introvert using Random forest

Type	INFP	INFJ	ISTP	INTP	INTJ	ISTJ	ISFJ	ISFP
Train Accuracy	0.98987	0.99623	0.99892	0.94693	0.95878	0.99919	0.99704	0.99919
Test Accuracy	0.91766	0.89629	0.97234	0.87052	0.87932	0.98743	0.9912	0.99434

Table 3: Train Test accuracy for persons with the common trait of Extrovertusing Random forest

Type	ENFP	ENFJ	ESTP	ENTP	ENTJ	ESTJ	ESFJ	ESFP
Train Accuracy	0.99165	0.99892	0.99327	0.97656	0.99973	1	1	1
Test Accuracy	0.94406	0.98492	0.98303	0.91703	0.97486	0.99371	0.99811	0.99623

Table 4: Train and Test Confusion Matrix for K=1 using K-nearest neighbor

K=1					
Train Dataset			**Test Dataset**		
Predicted	**0**	**1**	**Predicted**	**0**	**1**
Actual			**Actual**		
0	**41229**	**106**	**0**	**1434**	**18863**
1	**7**	**29722**	**1**	**757**	**13949**

Table 5: Train and Test Confusion Matrix for K=2 using K-nearest neighbor

K=2					
Train Dataset			**Test Dataset**		
Predicted	**0**	**1**	**Predicted**	**0**	**1**
Actual			**Actual**		
0	41234	101	**0**	1696	18601
1	1403	28326	**1**	1003	13703

Table 6: Introvert | Extrovert Classification Report using X-GBoost Classification

	Train Dataset			**Test Dataset**		
	Precision	**Recall**	**f1-score**	**precision**	**Recall**	**f1-score**
0	0.93	0.56	0.7	0.59	0.26	0.36
1	0.88	0.99	0.93	0.8	0.94	0.87
Accuracy			0.88			0.78
macro avg	0.91	0.77	0.81	0.7	0.6	0.62
weighted avg	0.89	0.88	0.87	0.75	0.78	0.75

4.7 Summary

The XGBoost model outperformed Logistic Regression and KNN models, aligning with expectations due to its ensemble-based approach.

Conclusion

This paper presents a comprehensive study on enhancing automated personality prediction by integrating various machine learning models. Utilizing the Myers-Briggs Type Indicator (MBTI) dataset, we focused on the Judging-Perceiving (J/P) dichotomy, a challenging aspect of personality prediction. Our approach combined multiple machine learning techniques, including Logistic Regression, Random Forest, K-Nearest Neighbors (KNN), and XGBoost to evaluate their individual and collective performance.

The results demonstrate that ensemble models, particularlyXGBoost, significantly outperform traditional single-model approaches in predicting the J/P dichotomy. This improvement is attributed to the model's ability to handle data complexity and imbalance more effectively. Advanced preprocessing techniques such as TF-IDF vectorization and Synthetic Minority Over-sampling Technique (SMOTE) further enhanced the model's accuracy and robustness.

This research addresses the limitations identified in previous studies, such as data leakage and inadequate model performance, and provides a robust framework for automated personality prediction. The findings highlight the potential of combining diverse modelling strategies to achieve more accurate and reliable predictions, offering valuable insights for applications in psychology, marketing, human resources, and beyond.

In conclusion, this study contributes to the field of personality prediction by demonstrating the effectiveness of integrated machine-learning models. The proposed methodology improves predictive accuracy and provides a deeper understanding of the influence of linguistic features on personality traits. Future research should explore the integration of cutting-edge technologies like Transformers and BERT, which may further enhance the performance and applicability of automated personality prediction systems. Additionally, applying these methods to a broader range of datasets will help validate their generalizability and robustness, paving the way for more advanced and practical applications in various domains.

References:

[1] Wenyu Li1 ,Chengpeng Wu1 , Xin Hu1 , Jingjing Chen2 , Shimin Fu3 , Fei Wang1 4 , Dan Zhang1 4(2022). Quantitative Personality Predictions From a Brief EEG Recording. IEEE Transactions onAffective Computing, doi: 10.1109/taffc.2020.3008775and Intelligent Systems

[2] Wiebke, Bleidorn., Patrick, L., Hill., Mitja, D., Back., Jaap, J., A., Denissen., Marie, Hennecke., Christopher, J., Hopwood., Markus, Jokela., Christian, Kandler., Richard, E., Lucas., Maike, Luhmann., Ulrich, Orth., Jenny, Wagner., Cornelia, Wrzus., Johannes, Zimmermann., Brent, W., Roberts. (2019). The policy relevance of personality traits. American Psychologist, doi: 10.1037/AMP0000503

[3] Yağmur, Güçlütürk., Umut, Güçlü., Marc, Pérez., Hugo, Jair, Escalante., Xavier, Baró., Carlos, Andujar., Isabelle, Guyon., Julio, C., S., Jacques, Junior., Meysam, Madadi., Sergio, Escalera., Marcel, A., J., van, Gerven., Rob, van, Lier. (2017). Visualizing Apparent Personality Analysis with Deep Residual Networks. doi: 10.1109/ICCVW.2017.367

[4] Madhura, Jayaratne., Buddhi, Jayatilleke. (2020). Predicting Personality Using Answers to Open-Ended Interview Questions. IEEE Access, doi: 10.1109/ACCESS.2020.3004002

[5] Hao, L., Chundong, W., Qingbo, H. (2023). A novel personality detection method based on high-dimensional psycholinguistic features and improved distributed Gray Wolf Optimizer for feature selection. Information Processing and Management, doi 10.1016/j.ipm.2022.103217

[6] Hans, Christian., Derwin, Suhartono., Andry, Chowanda., Kamal, Z., Zamli. (2021). Using a pre-trained language model and model averaging, text-based personality prediction from multiple social media data sources. Journal of Big Data, doi: 10.1186/S40537-021-00459-1

[7] Beck, E. D., & Jackson, J. J. (2022). A meta-analysis of personality prediction: Robustness and boundary conditions. Journal of Personality and Social Psychology, 122(3), 523–553. https://doi.org/10.1037/pspp0000386

[8] Mark, D., Lescroart. (2022). Limits of prediction accuracy on randomly selected natural images for model evaluation. Journal of Vision, doi: 10.1167/jov.22.3.52

[9] María, Fernanda, Vergara, Mengual. (2022). Automatic personality prediction: an enhanced method using ensemble modelling. Neural Computing and Applications, doi: 10.1007/s00521-022-07444-6

[10] Onno, Pepijn, Kampman., Elham, J., Barezi., Dario, Bertero., Pascale, Fung. (2018). Investigating Audio, Video, and Text Fusion Methods for End-to-End Automatic Personality Prediction. doi: 10.18653/V1/P18-2096

[11] Gal, Lavi., Jonathan, Rosenblatt., Michael, Gilead. (2022). A prediction-focused approach to personality modelling. Dental science reports, doi: 10.1038/s41598-022-16108-3

[12] Yash, Mehta., Samin, Fatehi., Amir Mohammad, Kazameini., Clemens, Stachl., Erik, C., Sauleh, E. (2020). Bottom-Up and Top-Down: Predicting Personality with Psycholinguistic and Language Model Features. doi: 10.1109/ICDM50108.2020.00146

[13] Isha, G., Manasvi, J., Prashant, J. (2023). Smart-Hire Personality Prediction Using ML. doi: 10.1109/ICDT57929.2023.10151367

[14] Nuo, H., Sijia, Li., Feng, Huang., Yeye, Wen., Yue, Su., Linyan, Li., Xiaoqian, Liu., Tingshao, Zhu. (2023). How social media expression can reveal personality. Frontiers in Psychiatry, doi: 10.3389/fpsyt.2023.1052844

[15] Speer, A. B., Christiansen, N. D., Robie, C., & Jacobs, R. R. (2022). Measurement specificity with modern methods: Using dimensions, facets, and items from personality assessments to predict performance. *Journal of Applied Psychology, 107*(8), 1428–1439. https://doi.org/10.1037/apl0000618

[16] Damien, L., Joseph, A., Gyorda., Erika, L., Moen., Nicholas, C., Jacobson. (2022). Using passive sensor data to probe social structure associations with personality changes: A synthesis of network analysis and machine learning. PLOS ONE, doi: 10.1371/journal.pone.0277516

[17] Nigam, A., Nigam, B., Bhaisare, C., & Arya, N. (2012). *Classifying the Bugs Using Multi-Class Semi-Supervised Support Vector Machine*. In Proceedings of the International Conference on Pattern Recognition, Informatics and Medical Engineering (PRIME-2012), 21-23 March 2012, IEEE.

[18] Nigam, B., Ahirwal, P., Salve, S., & Vamney, S. (2011). *Document Classification Using Expectation Maximization with Semi-Supervised Learning*. International Journal on Soft Computing (IJSC), 2(4), 37-44. DOI: 10.5121/ijsc.2011.2404.

[19] Sucheta, L., Jeffrey, S., Saltz. (2023). The Need for an Enhanced *Process Methodology for Ethical Data Science Projects. Ethics, doi: 10.1109/ETHICS57328.2023.10155007*

[20] Fernando, Martínez-Plumed., Lidia, Contreras-Ochando., Cèsar, Ferri., José, Hernández-Orallo., Meelis, Kull., Nicolas, Lachiche., María, José, Ramírez-Quintana., Peter, A., Flach. (2021). CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories. IEEE Transactions on Knowledge and Data Engineering, doi: 10.1109/TKDE.2019.2962680

[21] Mihaela, Cazacu., Emilia, Titan. (2020). Adapting CRISP-DM for Social Sciences. Brain, doi: 10.18662/BRAIN/11.2SUP1/97

[22] Vivek, Bhakta, Mathema., Kassaporn, Duangkumpha., Kwanjeera, Wanichthanarak., Narumol, Jariyasopit., Esha, Dhakal., Nuankanya, Sathirapongsasuti., Chagriya, Kitiyakara., Yongyut, Sirivatanauksorn., Sakda, Khoomrung. (2022). CRISP: a deep learning architecture for

GC × GC–TOFMS contour ROI identification, simulation and analysis in imaging metabolomics. Briefings in Bioinformatics, doi: 10.1093/bib/bbab550

[23] Sen, Cheng. (2013). The CRISP theory of hippocampal function in episodic memory. Frontiers in Neural Circuits, doi: 10.3389/FNCIR.2013.00088Sen, Cheng. (2013). The CRISP theory of hippocampal function in episodic memory. Frontiers in Neural Circuits, doi: 10.3389/FNCIR.2013.00088

[24] Dolores, G.-T., Dolores, G.-T., Ann, C, Bilbrey., Ester, Carolina, Apesoa-Varano., Rita, Ghatak., Katherine, K., Kim., Fawn, A., Cothran. (2020). Conceptual Framework to Guide Intervention Research Across the Trajectory of Dementia Caregiving. Gerontologist, doi: 10.1093/GERONT/GNZ157

[25] Avinash, Gaikwad., Avinash, Gaikwad., Ashwin, Nandagiri., Ashwin, Nandagiri., David, L., Potter., Reza, Nosrati., Anne, E., O'Connor., Anne, E., O'Connor., Sameer, Jadhav., Julio, Soria., Ranganathan, Prabhakar., Moira, K, O'Bryan. (2021). CRISPs Function to Boost Sperm Power Output and Motility. Frontiers in Cell and Developmental Biology, doi: 10.3389/FCELL.2021.693258

[26] Sarang Narkade, Understanding Confusion matrix,2018,https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62

[27] En, Jun, Choong., Kasturi, Dewi, Varathan. (2021). Predicting judging-perceiving of Myers-Briggs Type Indicator (MBTI) in an online social forum. PeerJ, doi: 10.7717/PEERJ.11382

[28] Nomura, A., Almeida, M., Johnson, S., &Pruinelli, L. (2021). Pain Information Model and Its Potential for Predictive Analytics: Applicability of a Big Data Science Framework. Journal of Nursing Scholarship, 53(3), 315-322.

[29] Nils, M. (2020). Analysis of an Intelligence Dataset. Journal of Intelligence, doi10.3390/JINTELLIGENCE8040039