Received 27 September 2025, accepted 10 July 2026, date of publication 30 September 2026
www.computingonline.net / computing@computingonline.net

Print ISSN 1727-6209
Online ISSN 2312-5381
DOI 10.47839/ijc.25.3.4855

From Personality Prediction to Behavioral Subtyping: A Machine Learning Framework for Mental Health Assessment

ASMA CHEBLI[1], SARA DAAS[2] , TOUFIK HAFS[2]

[1] LRI laboratory, Faculty of Technology, Computer science department, Badji Mokhtar Annaba University, Annaba, Algeria, P. O. Box 12, 23000 Annaba, Algeria

[2]L.E.R.I.C.A laboratory , Faculty of Technology, Electronics department, Badji Mokhtar Annaba University, Annaba, Algeria, P. O. Box 12, 23000 Annaba, Algeria

Corresponding author: Asma CHEBLI (asma.chebli@univ-annaba.dz).

ABSTRACT Introversion and extroversion are personality traits that shape individuals behaviours; they equally influence habits, attitude, and cognition. Modern psychological theories define the introvert-extrovert spectrum as a continuous dimension of personality, rather than two distinct personality types. It is essential to recognize this continuity in order to better understand and support individual differences,and it remains difficult to capture these nuances using computational methods, as most existing studies rely on textual data or data from questionnaires, which can introduce bias.

In our study, we investigate personality classification using behavioral features, as behavioral data offer an unexploited path to understanding personality beyond self-reported questionnaires, by leveraging a dataset that includes key indicators for 2,900 individuals and 08 behavioral features. Our methodology integrates both supervised and unsupervised learning to address complementary objectives, where supervised classifiers are used to establish a predictive framework for the extraversion-introversion dichotomy, while unsupervised clustering is used to test whether the resulting latent structure corresponds to these binary labels. Through combining predictive and exploratory methods, our work contributes both to benchmarking behavioral personality classification, providing evidence that certain prediction errors may correspond to significant behavioral subtypes and empirical evidence for the computational recognition of ambiverts.

KEYWORDS Behavioral Data; Introversion–Extraversion Spectrum; Personality Prediction; Supervised Prediction, Unsupervised Discovery.

I. INTRODUCTION

Bergner [1] defined personality as a permanent set of traits (dispositions to act, think, or feel) and styles (characteristic ways of handling matters). The personality of an individual points out how that person differs from the "standard normal person" in their society. Thus, personality is defined by stable and distinguishing characteristics rather than by temporary states. This is precisely what has made the subject of reliable personality assessment highly compelling and significant, as it can provide valuable insights for diverse applications spanning from mental health monitoring to adaptive technologies.

Most personality assessments are carried out on psychometric questionnaires such as the Myers-Briggs Type Indicator (MBTI) or the Big Five inventories. These two popular tools are effective, yet costly, time-consuming, and intrusive, which limits their applicability in real-world settings. With the emergence of digital footprints, recent research has increasingly turned to computational methods for personality identification.

Numerous studies have examined personality identification, merging psychology and technology through machine learning applied to textual data, most often sourced from social media.

Khan et al [2] applied XGBoost with TF-IDF features to predict MBTI personality dimensions from social media texts, reporting accuracies above 99%. Sahono et al [3] classified introverted and extroverted users from the MBTI Twitter dataset, by combining TF-IDF and SVM, an accuracy of 84% was reported. Equally, Naz et al [4] used a Bi-LSTM enhanced with word embeddings to predict extraversion traits from MBTI questionnaire text with 92.5% accuracy.

Most recently, Serano[5]also focused on personality prediction by integrating a stacked model combining classical

classifiers and deep learning models on the MyPersonality dataset, achieving an accuracy of approximately 72.7%. Similarly, Chhabra [6] explored a hybrid deep learning architecture for personality detection, in which an implemented CNN+LSTM model achieved 96.8% accuracy in predicting the four MBTI dimensions. Another study by Al-Fallooji & Al-Azawei [7] predicts personality based on texts by combining SVM, logistic regression, and LightGBM using linguistic preprocessing. A 100% accuracy rate was reported across all MBTI dimensions on a dataset of 8,668 records.

These studies have effectively demonstrated the efficiency of linguistic features for personality identification, but remain limited by their reliance on textual data, which does not always accurately reflect real-world behavioral patterns.

Zhu et al [8] advanced this line of research using a hybrid CNN-LSTM model, achieving an F1 score of 97.16% on MBTI text data, demonstrating the effectiveness of deep sequential architectures for personality inference.

A systematic mapping of AI-based tools for personality and emotion recognition was provided by Islam et al [9]. This approach highlights the need for standardized validation across multimodal approaches such as textual, facial, and behavioral data.

Passive data collected by smartphones in the work of Yan et al [10] showed that data, particularly mobility and physiological signals, can be used to predict adaptive and maladaptive personality traits, highlighting the role of real-world behavioral characteristics.

Mobile sensors have been used by Sze et al [11] to collect data such as accelerometers and movement patterns, which were used to predict personality traits, reporting F1 scores of up to 0.78, further validating the use of behavioral signals for machine learning models.

Recent physiological and sensor-based approaches have extended personality prediction past text and questionnaires. Bhardwaj et al. [12] have used EEG signals with a Deep Long Short-Term Memory (DLSTM) network to predict personality traits, reporting an accuracy of 96.94%, which outperforms the baseline models ANN, LibSVM, KNN, and ensemble genetic programming. Similarly, Subramanian et al. [13] presented ASCERTAIN, a multimodal framework using commercially available physiological sensing devices to jointly identify personality and affect in 58 participants. Such studies reinforce the hypothesis that real-world, non-textual signals carry meaningful information about personality; however, similar to the present study, they rely on fairly small cohorts, highlighting that dataset size remains a frequent constraint in research on behavioral prediction of personality; a point we will revisit in Section III.C.

Fieri et al [14] outside of the text and has explored psychometric survey data combined with machine learning; for the predicted introversion-extraversion. Based on responses to a questionnaire from an open-source psychometric project; the study used a range of supervised classifiers and achieved 95% accuracy with Random Forest.

All of these approaches remain tied to self-reported data, which can introduce biases and limitations of questionnaires.

Unlike these text- and survey-based approaches, we present a study based on a behavioral dataset (≈2,900 participants × 8 characteristics) from a Kaggle database Extrovert vs. Introvert Behavior Data. The database collects real-world indicators such as time spent alone, participation in social events, and activity on social media. We combined supervised classification (logistic regression, random forest, XGBoost) to evaluate the prediction of extroversion-introversion labels with unsupervised clustering (K-Means, Gaussian mixtures, agglomerative) to discover underlying personality subtypes.

Modern psychological theories emphasize that introversion and extroversion lie on a continuous spectrum rather than representing two mutually exclusive types[15]. This perspective implies that strict binary labels may overlook significant intermediate variations. As a result, we supplement supervised classification with unsupervised clustering in order to capture potential latent subtypes that transcend the dichotomy.

By moving from textual or survey data to behavioral data, and integrating prediction classification with discovery clustering , this study provides both a benchmark for behavioral personality classification and new evidence for the existence of hidden subtypes that extend beyond binary personality labels.

In our work, we address key questions: Do unsupervised clusters correspond to traditional labels? Are individuals misclassified by supervised models actually members of distinct and meaningful subgroups?

II. MATERIALS AND METHODS

This section describes the methodological framework of our study, including a presentation of the dataset, the preprocessing steps, the selected supervised and unsupervised modeling techniques, as well as the evaluation metrics used. The objective is to provide a description of the experimental process, from data preparation to model evaluation, in order to highlight how our approach combines predictive modeling and exploratory analysis of latent behavioral subtypes.

A. DATASET

The dataset used in our study is the Extrovert vs. Introvert Behavior Data available on Kaggle, which is intended to explore the spectrum of human personality. The dataset encompasses answers from 2,900 participants, with each individual entry featuring eight behavioral characteristics that reflect daily social engagement habits.

The eight behavioral characteristics included in this study are as follows: Time_spent_Alone (number of hours spent alone each day), Stage_fear (binary variable, self-reported), Social_event_attendance (frequency of participation in social events), Going_outside (weekly frequency of going outside), Drained_after_socializing (binary variable, self-reported), Friends_circle_size (number of close friends), Post_frequency (frequency of social media posts), and the target binary variable Personality (introvert/extrovert). The dataset is publicly available on Kaggle and does not contain demographic metadata (e.g., age, gender, culture, or geographic origin) about the participants, thus limiting the scope of subgroup or equity analyses that can be conducted.

To better understand the relationships between behavioral features and reveal strong and interpretable patterns across behavioral features, Figure 1 represents the correlation heatmap of behavioral variables in the dataset.. Citing as an example, time spent alone is negatively correlated with indicators of social activity such as participation in events (-0.73), outings (-0.75), size of circle of friends (-0.72), and frequency of posts (-0.73), confirming its role as a distinctive characteristic linked to introversion. Inversely, participation in social events exhibits a strong positive

correlation with going outside (0.75) and posting frequency (0.74), mirroring typical extroverted behaviors.

An initial overview prior to the learning phase helps to understand that systematic relationships effectively capture the behavioral contrasts underlying introversion and extroversion in the dataset. Simultaneously, moderate overlaps between variables (e.g., posting frequency related to both social events and friend circle size) help explain why some individuals may exhibit mixed or hybrid patterns, subsequently reflected in model misclassifications and clustering limitations.

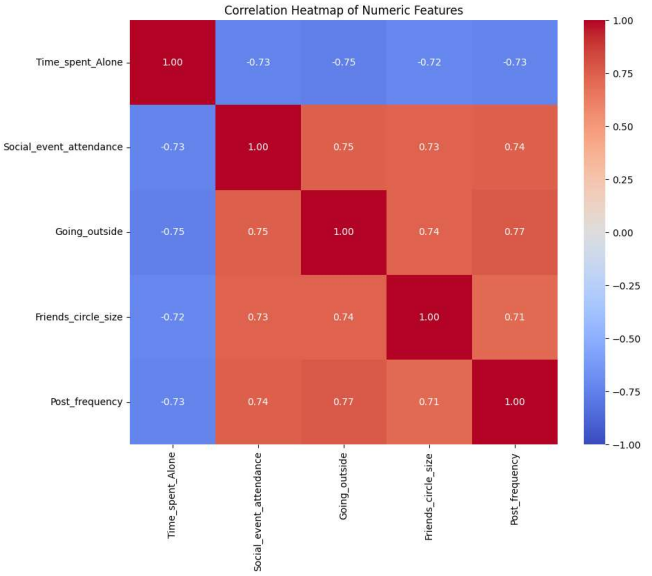


Figure 1. Correlation heatmap of behavioral variables in the dataset.

These features reflect both social activity (e.g., participation in events, number of friends)along with self-perception (e.g., stage fright, energy after socializing), providing a behavioral basis for predicting personality, and providing a binary label that indicates whether each participant is classified as an introvert or extrovert.

Figure 2 demonstrates the distribution of target labels of the dataset used (introvert and extrovert). The balance of the dataset reduces the risk of significant bias between classes in training. This supports the reliability of evaluation metrics like accuracy, and justifies the use of F1 score and precision-recall for a more comprehensive evaluation (see section E. Evaluation Metrics).

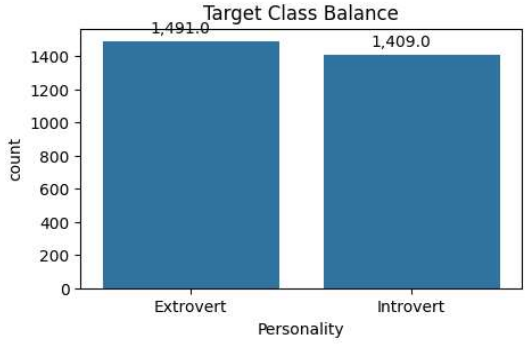


Figure 2. Target Class Balance.

We note some limitations within the dataset that may affect reproducibility. First, two of the eight features (Stage_fear and Drained_after_socializing) consist of self-reported binary indicators as opposed to passively measured behavioral markers, implying that the dataset is not purely behavioral and may contain some self-reporting bias. Second, the source and collection methodology of the Kaggle dataset are not fully documented by its creators; hence, the sampled population, the collection period, and potential sampling biases cannot be independently verified.

B. DATA PREPROCESSING

One of the essential aspects of data preparation is data preprocessing. This refers to any processing carried out on raw data in order to prepare it for subsequent analysis or processing tasks. A number of preprocessing steps were carried out to ensure data quality prior to any use:

- Missing values have been processed using SimpleImputer: median imputation for numerical variables and modal imputation for categorical variables[16].
- Categorical variables (Stage_fear, Drained_after_socializing) have been encoded as dummy variables (a binary value 0 or 1) .
- Outliers in numerical features were capped using the interquartile range (IQR) method.
- The dataset was split into 80% training set and 20% testing set.

C. SUPERVISED MODELS

In line with previous studies that have applied machine learning to data from questionnaires or texts (e.g., Fieri et al [14]; Khan et al [2]), and based on the results achieved, supervised classifiers were selected to establish predictive benchmarks for introversion-extraversion classification. Unlike textual data, the characteristics in our study are structured behavioral indicators, which makes interpretable and robust algorithms particularly suitable.

We opted for three models for supervised prediction, starting with logistic regression, chosen as a simple but interpretable linear model [17], intended to provide a baseline for binary classification in order to assess the linear separability of behavioral traits. Next, we applied a random forest (RF), a set of decision trees [18], selected on the basis that RF is an effective tool for capturing nonlinear interactions between features, which is an important property given that behavioral traits can be highly complex. Finally, gradient boosting (XGBoost) [19] was included as a state-of-the-art ensemble method, widely used in personality prediction tasks, to further test the robustness and generalization ability of the classifiers.

D. UNSUPERVISED MODELS

As modern psychology conceptualizes introversion-extraversion as a spectrum rather than a rigid dichotomy, for in-depth experimentation we supplemented supervised prediction with clustering techniques, used in the aim of discovering hidden patterns in personality data.

Unsupervised learning is an essential exploratory tool, given that the goal of this work is to determine whether individuals “misclassified” in supervised learning form meaningful subgroups. K-Means [20], widely used for cluster partitioning, was adopted, with the number of clusters

optimized using the Elbow method and Silhouette scores. Gaussian mixture models (GMM) were also used to determine flexible memberships [21], reflecting the possibility that individuals may belong to multiple subtypes partially; which can be very revealing for personality prediction. Finally, agglomerative hierarchical clustering [22] was applied as a complementary approach to further validate the hypothesis, as it does not require specifying the number of clusters in advance and can highlight hierarchical relationships between the subgroups formed.

E. EVALUATION METRICS

Given that our framework combines supervised and unsupervised learning models, for a valid evaluation, it is necessary to apply a range of metrics adapted to each model to assess both the predictive performance of the classifiers and the validity of the discovered clusters.

The performance of supervised predictive models was evaluated through a combination of complementary metrics to ensure a robust assessment.

Accuracy (1) was carried out as a benchmark indicator of overall correctness, we also considered precision(2), recall(3), and F1 score(4), which reflect compromises between false positives and false negatives.

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

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$F1\ Score = 2 \times \frac{Recall \times Precision}{Recall+Precisi} \quad (4)$$

As recommended in previous studies on personality prediction, for further evaluating discriminative ability independently of threshold choice, the ROC-AUC measure was included.

In addition, confusion matrices were analyzed to interpret classification errors, a crucial step in determining whether misclassified individuals form meaningful subgroups in cluster analysis.

Prior to evaluating the results of the clustering, it is essential to determine the optimal number of clusters[23] [24] and then use internal and external indices to assess their validity. For all models, the silhouette coefficient was applied in order to quantify intra-group cohesion and inter-group separation, suiting it to our behavioral data with possibly overlapping characteristics. In addition, the Elbow method was specifically used for K-Means, where it provides a standard heuristic for estimating the optimal number of clusters by analyzing the sum of intra-cluster squares. For agglomerative clustering, inspection of the dendrogram complemented the silhouette analysis to determine an appropriate cluster structure, while GMM was primarily evaluated using the silhouette due to its probabilistic nature. To compare the discovered clusters with the binary reference labels, we used the Adjusted Rand index (ARI), which corrects for chance, normalized mutual information (NMI), which measures the information shared between predicted and actual partitions, and purity, which evaluates the dominance of a single class within clusters.

The combination of these measures ensured that the predictive performance of the classifiers and the theoretical validity of the latent clusters were rigorously evaluated, in line with both objectives of the study.

III. RESULTS AND DISCUSSION

In this section we present the experimental results obtained from supervised and unsupervised analyses, along with their interpretation in relation to the study objectives. First, the performance of supervised classifiers is reported in predicting introversion-extraversion based on the behavioral characteristics of the dataset used, supported by confusion matrices and ROC curves. Unsupervised learning is addressed next, we start the experiments by determining the optimal number of clusters using the Elbow, Silhouette, and dendrogram methods, a crucial step in clustering, before evaluating the performance of the models using external and internal indices. We discuss the findings of our experiments, highlighting the observation that certain classification errors correspond to coherent subgroups, providing insight into potential latent personality subtypes.

A. SUPERVISED PREDICTION

Table 1 reports the results of the three supervised classifiers for the introversion-extraversion prediction task. The best results were obtained by logistic regression, with an accuracy of 0.9296 and a balanced precision 0.9330 and recall 0.9262, which led to the highest F1 score 0.9292. Close behind, random forest produced comparable results (accuracy of 0.9283, F1 score of 0.9282, while XGBoost performed slightly lower (accuracy of 0.9204, F1 score of 0.9206).

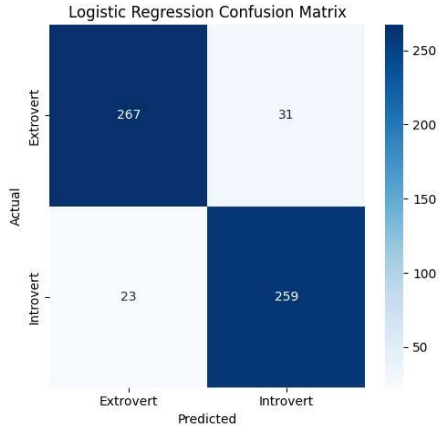
Although the first aspect of our study is the prediction of introversion-extraversion and the metrics in Table 1 establish the strong predictive ability of our models, the complementary results provided by the confusion matrices are particularly valuable and of interest for our study, as they reveal misclassified participants and the reasons for misclassification, an aspect directly related to our study of latent personality subtypes.

Table 1. Supervised Model Classification Results .

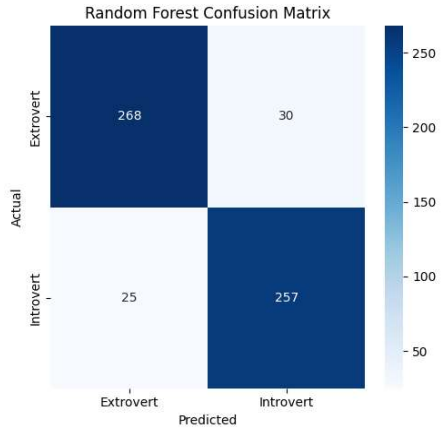
Model	Accuracy	Precision	Recal	F1-score
Logistic Regression	0.9296	0.9330	0.9262	0.9292
Random Forest	0.9283	0.9290	0.9279	0.9282
XGBoost	0.9204	0.9181	0.9237	0.9206

The confusion matrices in Figure 3 (a-c) indicate that the three classifiers correctly predicted most participants, with logistic regression Figure 3(a) 267 extroverts, 259 introverts and the random forest Figure 3(b) 268 extroverts, 257 introverts giving nearly the same results, while XGBoost Figure 3(c) got 261 correct predictions for both classes. Classification errors remained small but consistent(20 to 30 errors per model). In order to characterize this consistency, we examined the overlap among individuals misclassified by all three classifiers. A significant fraction of the misclassified individuals were

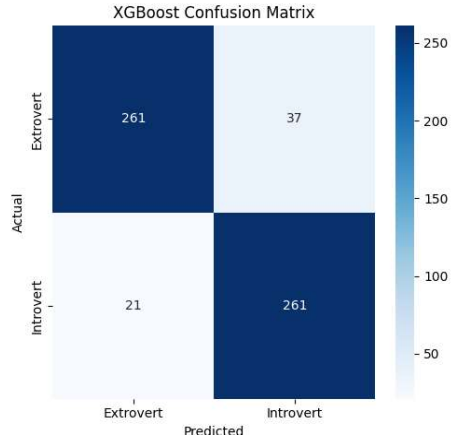
misclassified by all three, indicating that these errors are centralized in a stable subset of the sample rather than being randomly distributed across the models. This common subset showed that the predicted class probabilities consistently clustered around the decision threshold of 0.5 rather than being distributed across the probability distribution, a pattern consistent with genuine behavioral ambiguity rather than random model noise. These hybrid cases highlight the limitations of binary labels and led us to conduct an unsupervised analysis to determine whether these "difficult to classify" participants form coherent latent groups.



(a)



(b)



(c)

Figure 3. :Confusion matrices for Unsupervised models. (a) Logistic Regression, (b) Random Forest, (c) XGBoost .

The ROC curves in Figure 4 display a high discrimination capability for all three supervised models, with AUC values of 0.9261 for logistic regression, 0.9445 for random forest, and 0.9493 for XGBoost. These results demonstrate that ensemble methods outperform the linear baseline slightly, yet all classifiers consistently achieve a near perfect separation between introverts and extroverts. The significant rise of the ROC curves towards the upper left corner indicates a low false positive rate, with high true positive rates, thus indicating robust classification performance at all thresholds. It is worth mentioning that robustness confirms the predictive value of behavioral characteristics, thus validating and boosting their use as an alternative to textual or survey data for the implementation of personality assessment tools.

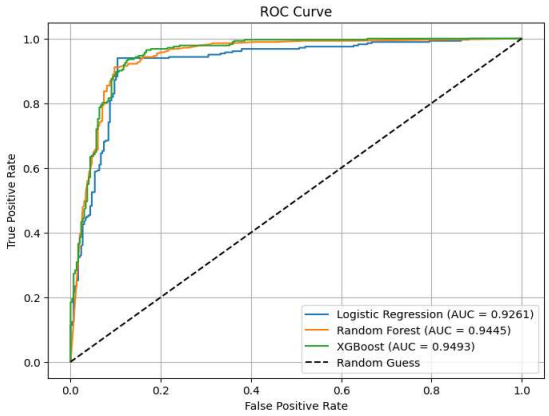


Figure 4. ROC curves of supervised models for introversion–extroversion classification.

Beyond the above predictive performance, we next had to ask: which features most useful for actually performing this classification? To shed light on this question, we inspected feature importance of the Random Forest model in Figure 5. We discovered that time spent alone, social event attendance, and size of one's friend circle were the most discriminative, closely mirroring psychological theories of introversion/extroversion.However, it is also worth noting that these very features can generate conflicting signals in certain individuals for example, someone who frequently attends social events but also spends extended time alone. Such profiles help explain why some participants were consistently misclassified across models and suggest the presence of intermediate personality patterns. In this sense, feature importance not only validates the plausibility of the classifiers but also provides a behavioral rationale for the emergence of ambiverts, whose mixed signals position them between traditional categories.

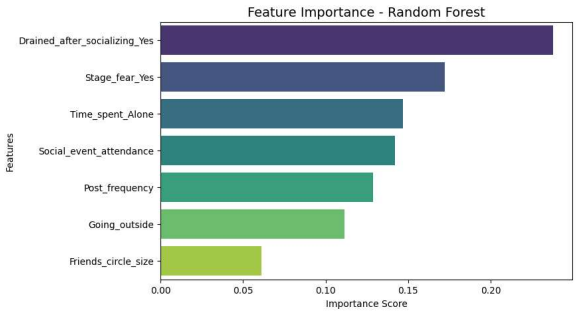


Figure 5. Random Forest features importance for behavioral data.

In Figure 5, the ranking of feature importance generated by the Random Forest algorithm demonstrates which behavioral features are, overall, the most significant across the entire dataset; however, the ranking does not explain what causes an individual in a borderline or misclassified category to receive a particular prediction. This would require the use of instance-level methods, such as SHAP or LIME, which are beyond the scope of this study and will be the subject of future work (Section IV).

It is also important to consider the imbalance between observable indicators and those self-reported by individuals: Social_event_attendance, Going_outside, Friends_circle_size, and Post_frequency are externally observable, whereas Stage_fear and Drained_after_socializing are self-reported. Since the former dominate the ranking by importance, our interpretation of ambivert profiles could be biased toward visible social activity, potentially underestimating internal states that are more difficult to measure but equally relevant to ambiversion; an interpretability bias inherent in the set of characteristics itself rather than in the modeling approach.

To contextualize our results relative to prior work, Table 2 summarizes accuracy/F1-score and data modality across representative studies.

Table 2. Comparison with representative prior studies on personality classification.

Study	Data Modality	Method	Accuracy	F1-score
Khan et al. [2]	Text (social media)	XGBoost + TF-IDF	>99%	—
Sahono et al. [3]	Text (Twitter/MBTI)	SVM + TF-IDF	84%	—
Fieri et al. [12]	Survey (psychometric)	Random Forest	95%	—
Sze et al. [11]	Behavioral (mobile sensors)	ML classifiers	—	0.78
Bhardwaj et al. [add citation]	Physiological (EEG)	Deep LSTM	96.94 %	—
Our study	Behavioral (self-report + activity)	Logistic Regression	92.96 %	0.9292

It is important to highlight, though, that raw classification accuracy is not the main objective of this study. In contrast to the studies summarized in Table 3, which treat the prediction of introversion-extraversion as a purely binary classification task and evaluate success exclusively in terms of accuracy or F1 score, our study intentionally combines supervised prediction and unsupervised discovery to challenge the validity of the binary label itself.

We present the accuracy achieved by our behavioral models (Table 1) as a necessary benchmark rather than a final goal: it establishes that our behavioral characteristics contain a predictive signal sufficient to make subsequent analysis of classification errors and clustering meaningful, rather than being the result of poor model performance.

The main contribution of this work lies in what is revealed by the approximately 2% of individuals systematically misclassified by all three classifiers, as well as their systematic placement near the decision boundaries of the clusters (Section III.B), regarding the existence of coherent ambivert profiles, an issue that none of the previous studies cited in Table 2 has sought to address.

B. UNSUPERVISED DISCOVERY

Prior to evaluating the clustering results, the optimal number of clusters has been determined using the Elbow method, the Silhouette score, and Dendrogram(as explained in section 2.4). These methods have been selected as they offer complementary insights into the cohesion and separation of clusters. This step is represented in Figures 6 to 8, indicating that the data set is globally well represented by two groups, corresponding to the introvert-extrovert dichotomy.

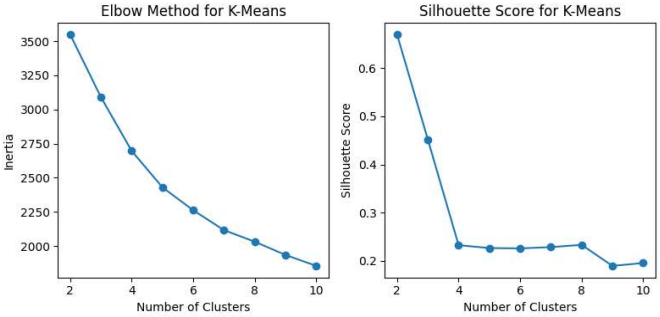


Figure 6. Determining optimal number of clusters for K-Means using the Elbow Method and Silhouette Score.

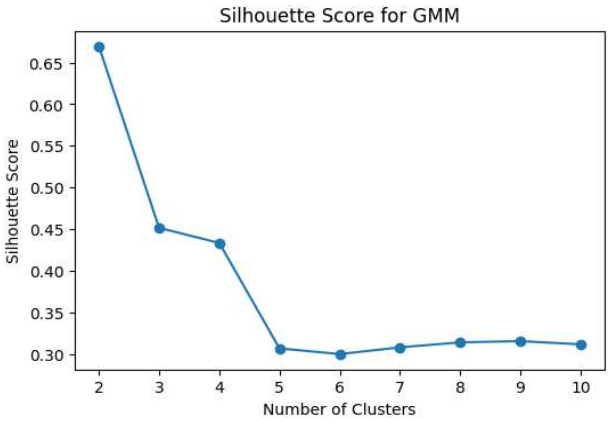


Figure 7. Determining optimal number of clusters for GMM using Silhouette Score.

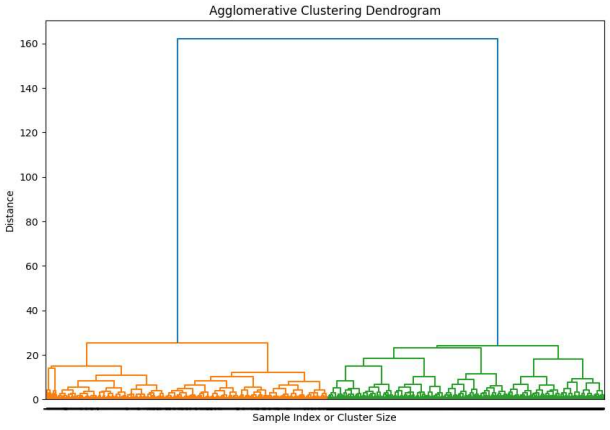


Figure 8. Determining optimal number of clusters for Agglomerative Clustering using Dendrogram.

In determining the optimal number of clusters, two clusters were consistently identified for all methods. Although this result confirms that the dataset can be split into two general groups: introverts and extroverts; a more comprehensive and relevant overview requires further consideration of the evaluation metrics. Table 3 shows that this alignment is not perfect. These three metrics capture distinct aspects of cluster quality. Purity (0.9172) is high, measuring how dominated each cluster is by a single ground-truth class, but it is lenient: it stays high even if a consistent minority of "wrong" points are simply absorbed into the majority cluster. The adjusted Rand index (0.6958) corrects for chance agreement between clustering and ground truth; a value near 0.70 indicates substantial but incomplete agreement, consistent with two clusters that capture the dominant split while leaving a non-trivial share of individuals ambiguously assigned. The normalized mutual information (0.5927), similarly lower than purity, reinforces that part of the clustering structure is not explained by the binary label alone. The gap between high purity and moderate ARI/NMI is itself diagnostic: clusters are label-dominant overall, but a systematic subset of individuals is not confidently assigned by either the classifier or the clustering, precisely the candidate ambiverts discussed in Section III.A. The Silhouette score (0.6725) reflects this at the individual level: while the mean indicates reasonably well-separated clusters, a tail of low per-point scores corresponds to the same borderline cases. Cross-referencing cluster membership with the misclassification subset described in Section III.A, we observed that a notable proportion of consistently misclassified individuals also fell within the lowest silhouette-score range in the K-Means solution; simultaneously the hardest cases for supervised classifiers to label and the least cleanly separated cases in unsupervised clustering. This convergence between two independent methodological approaches provides stronger evidence for the existence of a coherent ambivert subgroup than either of these methods could provide on its own.

Table 3.Unsupervised Model Clustering Results .

Model	Adjusted Rand Index (ARI)	Normalized Mutual Information (NMI)	Purity	Silhouette Score
K-Means Clustering	0.6958	0.5927	0.9172	0.6725
Gaussian Mixture Models (GMM)	0.6958	0.5927	0.9172	0.6725
Agglomerative Clustering	0.6958	0.5927	0.9172	0.6725

Although these subgroups did not exhibit sufficient robustness to be identified as distinct clusters by the applied methods, their recurrent emergence is theoretically significant and warrants careful consideration.

The dataset of approximately 2,900 participants described by eight behavioral characteristics provides strong evidence of the introvert-extrovert dichotomy, yet may not offer sufficient granularity to reveal finer substructures in the form of distinct clusters.

Significantly, the number of misclassified individuals was relatively small but consistent across all supervised models: for example, logistic regression misclassified 54 participants (≈1.9% of the dataset), Random Forest misclassified 55 participants (≈1.9%), and XGBoost misclassified 58 participants (≈2.0%). While these proportions may seem minor, the fact that the same individuals were consistently misclassified across different classifiers; and had a tendency to cluster near decision boundaries in the unsupervised analysis; suggests that they represent consistent hybrid profiles rather than random noise.

Taking into consideration one of the recent studies in psychological research, which has been recognizing for some time that many individuals do not clearly fit into the categories of introvert or extrovert, but occupy an intermediate position on a spectrum, often referred to as ambiversion [25][26][27]. Our results reinforce this perspective: participants who are consistently misclassified by supervised models and located near the cluster boundaries in unsupervised analysis likely correspond to these ambivert profiles. Instead of mirroring a model error, these cases show there are meaningful intermediate subtypes, in line with theories that see introversion-extroversion as a spectrum rather than a rigid dichotomy.

C. LIMITATIONS AND GENERALIZATION

While our results confirm the value of behavioral indicators for predicting personality and for computer recognition of ambiverts, certain limitations restrict the generalizability of the proposed framework. First, the dataset provides no demographic metadata (age, gender, culture, or nationality), which makes it difficult to verify whether the observed behavioral patterns and identified ambivert profiles are generalizable to different populations or whether they reflect the norms of a specific cultural or social context in which the data were collected.

This concern is not purely hypothetical: Khwaja et al. [28] reported that the relationship between behavioral indicators derived from mobile sensors and personality traits varies considerably from one country to another and across cultural contexts, indicating that models trained on the behavioral norms of a single population are not automatically reliable when applied to others. This limitation applies directly to the present dataset, whose cultural origin and composition are not documented, and by extension to the subgroup of ambiverts we identify, whose behavioral signature may itself be culturally contingent rather than universal.

Second, two of the eight characteristics are based on self-reports from the participants (Stage_fear, Drained_after_socializing), reintroducing an element of the same self-report bias that this study otherwise seeks to overcome by using behavioral data. Given that these two traits appear among the most predictive in our analysis of trait significance using the Random Forest method (Section III.A, Figure 5), the ambivert profiles identified may be partly shaped by self-perception biases rather than purely observable behavior, a nuance that the current dataset does not allow us to untangle.

On a more fundamental level, the binary “introvert/extrovert” distinction is in itself a simplification that may not reflect the underlying psychological reality: it fails to

account for a truly intermediate profile. Therefore, our analysis of "classification errors" relies on a label that may be poorly calibrated for a portion of the population, rather than on a "ground truth" that the models simply failed to learn.

Third, the size of the dataset (2,900 participants) and the granularity of the characteristics (8 variables) limit the resolution at which latent subtypes may be isolated statistically; as discussed in Section III.B, this likely explains why clustering methods identified only two dominant groups, even though the classification error analysis revealed consistent patterns of hybrid profiles. A larger and more behaviorally granular dataset will likely be necessary before ambiverts can emerge as a fully distinct computational cluster, rather than as a pattern deduced indirectly from the disagreement among classifiers and the proximity of cluster boundaries, as is the case in the present study.

Finally, it is important to note that external validation and predictive performance, while essential methodological requirements, were not the primary objective of this study; the model has not been externally validated on an independent dataset from a different source or population, and we do not present it as a predictive tool suitable for deployment. Accordingly, neither its predictive performance nor its approach to identifying ambiverts can be immediately applied to other behavioral datasets or to real-world implementation contexts. Rather, the value of this study resides in demonstrating that "ambivert" type profiles emerge consistently across different independent supervised classifiers and are corroborated by the proximity of the boundaries of unsupervised clusters, suggesting that these profiles reflect a genuine behavioral phenomenon rather than an isolated model error.

Future research should prioritize datasets annotated demographically and interculturally, along with external validation, both to establish the generalizability of the supervised and unsupervised results presented here, as well as to determine whether the ambivert profiles identified in this study reflect a stable and consistent personality subtype from an intercultural perspective, or whether they are an artifact.

IV. CONCLUSION

This study presented a combined supervised and unsupervised learning framework to explore personality through behavioral data. Results demonstrated that supervised models can reliably predict introversion-extraversion, while clustering reveals latent subgroups that go beyond binary classification.

Supervised models such as logistic regression, random forest, and XGBoost achieved high predictive accuracy, confirming the value of simple behavioral characteristics as alternatives to textual or psychometric data, and comparable to prior text, survey, and sensor-based approaches (Table 2); though this benchmarking serves as validation rather than the study's primary aim. In complement, unsupervised clustering revealed that individuals consistently misclassified by supervised models tended to form coherent subgroups near cluster boundaries.

Our study has highlighted the existence of ambiverts, a personality profile that falls between introversion and extroversion and has been recognized in psychology yet rarely implemented in computational studies.

Although the database used is not sufficiently large or granular to isolate ambiverts as distinct clusters, their constant occurrence across all methods is theoretically significant. These profiles provide empirical support for modern psychological theories rather than random noise; they conceptualize introversion-extraversion as a spectrum rather than a rigid dichotomy. This evidence should nonetheless be interpreted with caution given the dataset's lack of demographic and cultural metadata and the absence of external validation (Section III.C).

This paper extended research on personality prediction beyond categorical classification by explicitly recognizing ambiversion in a machine learning framework, using supervised prediction and unsupervised discovery. Future research will build on the findings of this study by leveraging advanced AI methods such as transformer-based models for an enriched feature representation, multimodal learning that integrates behavioral, textual, and physiological data, and for a better interpretation of the nature of intermediate subtypes the use of explainable AI (XAI) techniques.A larger and more diverse and demographically annotated datasets may also enable ambiverts to emerge as distinct computational categories and confirm their consistency across cultures. Ultimately, integrating ambiversion into AI-based personality modeling provides promising prospects for digital phenotyping and personalized mental health interventions.

References

[1] R. M. Bergner, "What is personality? Two myths and a definition," *New Ideas Psychol.*, vol. 57, p. 100759, 2020, https://doi.org/10.1016/j.newideapsych.2019.100759.

[2] A. S. Khan, H. Ahmad, M. Zubair, F. Khan, A. Arif, and H. Ali, "Personality classification from online text using machine learning approach," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 3, 2020, https://doi.org/10.14569/IJACSA.2020.0110358.

[3] M. N. Sahono, H. A. Pradana, R. P. Sari, M. Gunawan, R. F. Rahmat, and A. G. Wibowo, "Extrovert and introvert classification based on Myers–Briggs Type Indicator (MBTI) using Support Vector Machine (SVM)," in *Proc. Int. Semin. Appl. Technol. Inf. Commun. (iSemantic)*, Semarang, Indonesia, 2020, pp. 572–577, https://doi.org/10.1109/iSemantic50169.2020.9234288.

[4] A. Naz, H. U. Khan, S. Alesawi, O. I. Abouola, A. Daud, and M. Ramzan, "AI knows you: Deep learning model for prediction of extroversion personality trait," *IEEE Access*, vol. 12, pp. 159152–159175, 2024, https://doi.org/10.1109/ACCESS.2024.3486578.

[5] J. Serrano-Guerrero, B. Alshouha, M. Bani-Doumi, F. Chiclana, F. P. Romero, and J. A. Olivas,"Combining machine learning algorithms for personality trait prediction," Egyptian Informatics Journal, vol. 25, p. 100439, 2024. https://doi.org/10.1016/j.eij.2024.100439.

[6] P. Chhabra, S. Kumar, G. Chhabra, R. Tanwar, A. Virmani, and A. Dwivedi, "Personality trait classification from online text using hybrid deep learning techniques," in S. Tanwar, P. K. Singh, M. Ganzha, and G. Epiphaniou, Eds., *Proc. 5th Int. Conf. Comput., Commun., Cyber-Secur.* (Lecture Notes in Networks and Systems, vol. 991). Singapore: Springer Nature, 2024, pp. 689–701, https://doi.org/10.1007/978-981-97-2550-2_49.

[7] A. S. Al-Fallooji and A. Al-Azawei, "Predicting users' personality on social media: A comparative study of different machine learning techniques," *Karbala Int. J. Mod. Sci.*, vol. 8, no. 4, pp. 617–630, 2022, https://doi.org/10.33640/2405-609X.3262.

[8] Z. Zhu, "Advanced prediction of Myers–Briggs personality traits using hybrid CNN-LSTM models and textual data," *Intell. Decis. Technol.*, vol. 19, no. 4, pp. 2772–2787, 2025, https://doi.org/10.1177/18724981251319630.

[9] A. Islam, N. F. Mohd Noor, and S. S. Abdul Rahman, "Systematic mapping study of tools to identify emotions and personality traits," *Discov. Artif. Intell.*, vol. 5, no. 1, p. 58, 2025, https://doi.org/10.1007/s44163-025-00271-3.

[10] R. Yan, Y. Zhang, Y. Ma, Q. Wu, and T. Zhu, "Exploratory machine learning modeling of adaptive and maladaptive personality traits from passively sensed behavior," *Future Gener. Comput. Syst.*, vol. 132, pp. 266–281, 2022, https://doi.org/10.1016/j.future.2022.02.010.

[11] W. Y. S. Sze, M. P. Herrero, and R. Garriga, "Personality trait inference via mobile phone sensors: A machine learning approach," *arXiv*, 2024, doi: 10.48550/arXiv.2401.10305.
[12] H. Bhardwaj, P. Tomar, A. Sakalle, W. Ibrahim, "EEG-based personality prediction using fast Fourier transform and DeepLSTM model," *Computational Intelligence and Neuroscience*, 2021. https://doi.org/10.1155/2021/6524858.
[13] R. Subramanian, J. Wache, M. K. Abadi, R. L. Vieriu, S. Winkler, & N. Sebe, "ASCERTAIN: Emotion and personality recognition using commercial sensors," *IEEE Transactions on Affective Computing*, vol. 9, issue 2, pp. 147–160, 2018. https://doi.org/10.1109/TAFFC.2016.2625250.
[14] B. Fieri, J. La'la, and D. Suhartono, "Introversion-extraversion prediction using machine learning," *Int. J. Inf. Vis.*, vol. 7, no. 4, p. 2154, 2023, https://doi.org/10.62527/joiv.7.4.1019.
[15] D. Petric, "Introvert, extrovert and ambivert," 2019, doi: 10.13140/RG.2.2.28059.41764/2.
[16] A. Chebli and C.-G. Haba, "Coping with missing values in medical data sets: A proposal of an approach based on soft clustering and measures of central tendency," in *Proc. E-Health Bioeng. Conf. (EHB)*, Iasi, Romania, 2022, pp. 1–4, https://doi.org/10.1109/EHB55594.2022.9991417.
[17] S. Domínguez-Almendros, N. Benítez-Parejo, and A. R. Gonzalez-Ramirez, "Logistic regression models," *Allergol. Immunopathol.*, vol. 39, no. 5, pp. 295–305, 2011, https://doi.org/10.1016/j.aller.2011.05.002.
[18] A. Parmar, R. Katariya, and V. Patel, "A review on random forest: An ensemble classifier," in J. Hemanth, X. Fernando, P. Lafata, and Z. Baig, Eds., *Proc. Int. Conf. Intell. Data Commun. Technol. Internet Things (ICICI) 2018* (Lecture Notes on Data Engineering and Communications Technologies, vol. 26). Cham, Switzerland: Springer, 2019, pp. 758–763, https://doi.org/10.1007/978-3-030-03146-6_86.
[19] C. Bentéjac, A. Csörgő, and G. Martínez-Muñoz, "A comparative analysis of gradient boosting algorithms," *Artif. Intell. Rev.*, vol. 54, no. 3, pp. 1937–1967, 2021, https://doi.org/10.1007/s10462-020-09896-5.
[20] M. Ahmed, R. Seraj, and S. M. S. Islam, "The k-means algorithm: A comprehensive survey and performance evaluation," *Electronics*, vol. 9, no. 8, p. 1295, 2020, https://doi.org/10.3390/electronics9081295.
[21] M.-S. Yang, C.-Y. Lai, and C.-Y. Lin, "A robust EM clustering algorithm for Gaussian mixture models," *Pattern Recognit.*, vol. 45, no. 11, pp. 3950–3961, 2012, https://doi.org/10.1016/j.patcog.2012.04.031.
[22] F. Murtagh and P. Contreras, "Algorithms for hierarchical clustering: An overview," *WIREs Data Min. Knowl. Discov.*, vol. 2, no. 1, pp. 86–97, 2012, https://doi.org/10.1002/widm.53.
[23] Y. Jung, H. Park, D.-Z. Du, and B. L. Drake, "A decision criterion for the optimal number of clusters in hierarchical clustering," *J. Glob. Optim.*, vol. 25, no. 1, pp. 91–111, 2003, https://doi.org/10.1023/A:1021394316112.
[24] D. M. Saputra, D. Saputra, and L. D. Oswari, "Effect of distance metrics in determining K-value in K-means clustering using elbow and silhouette method," in *Proc. Sriwijaya Int. Conf. Inf. Technol. Appl. (SICONIAN 2019)*, Palembang, Indonesia, 2020, Atlantis Press, https://doi.org/10.2991/aisr.k.200424.051.
[25] S. Butt, G. Sidorov, and A. Gelbukh, "Interpretation of Myers–Briggs Type Indicator personality profiles based on ambivert continuum scale," *Expert Syst. Appl.*, vol. 264, p. 125689, 2025, https://doi.org/10.1016/j.eswa.2024.125689.
[26] A. Grant, "Rethinking the Extraverted Sales Ideal: The Ambivert Advantage," *Psychological Science*, vol. 24, no. 6, pp. 1024–1030, 2013, https://doi.org/10.1177/0956797612463706.
[27] K. Moore, We Are All Ambiverts Now: Introverts, Ambiverts, and Extroverts in the C-Suite, 1st ed. Oxford: Taylor & Francis Group, 2025. https://doi.org/10.4324/9781003612216-1.
[28] M. Khwaja, S. S. Vaid, S. Zannone, G. M. Harari, A. A. Faisal, and A. Matic, "Modeling personality vs. modeling personalidad: In-the-wild mobile data analysis in five countries suggests cultural impact on personality models," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 3, no. 3, Art. no. 88, pp. 1–24, 2019, https://doi.org/10.1145/3351246.

***Asma CHEBLI** is an Associate Professor at the Computer Science; Faculty of Technology at Badji Mokhtar-Annaba University, Algeria. Her research interests include machine learning, pattern recognition, case-based reasoning, and semi-supervised learning. She has published works addressing challenges in medical data analysis, such as handling missing values and improving computer-aided diagnosis systems. Her recent research focuses on explainable AI and its application to healthcare, aiming to enhance trust and interpretability in predictive models. She has expertise in clustering algorithms, fuzzy clustering, active learning, and case-base maintenance, and is committed to bridging theoretical advances with practical solutions in the medical domain.*

***SARA DAAS** is an Associate Professor at the Electronics Department at Badji Mokhtar-Annaba University, Algeria. She is a member of the LERICA lab and has published papers in the areas of image processing, signal/image/video processing, pattern recognition, and biomedical engineering. Her research focuses particularly on biometric recognition systems (including multimodal biometrics and ECG-based identification), feature extraction, and image segmentation. She applies deep learning methods for improving accuracy, security, and robustness in recognition tasks. Sara Daas is interested in the intersection of biometrics and medical/healthcare applications.*

***TOUFIK HAFS** is a Full Professor at the Electronics Department, Faculty of Technology, at Badji Mokhtar Annaba University, Algeria. His research interests focus on artificial intelligence, biometric recognition, and secure information systems. He has contributed extensively to the development of algorithms integrating physiological and behavioral characteristics for reliable biometric authentication. Professor Hafs has published numerous works in international journals and conferences and has supervised several master's and doctoral theses. His recent research also explores the application of machine learning and explainable AI in the medical field, particularly for predictive modeling and decision support. He actively collaborates with academic and industrial partners to bridge the gap between theory and practice.*