IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 17 December 2023, accepted 21 December 2023, date of publication 1 January 2024,
date of current version 5 January 2024.

Digital Object Identifier 10.1109/ACCESS.2023.3348755

RESEARCH ARTICLE

Analyzing Social Exchange Motives With Theory-Driven Data and Machine Learning

KEVIN IGWE AND KEVIN DURRHEIM
Department of Psychology, Faculty of Humanities, University of Johannesburg, Johannesburg 2092, South Africa
Corresponding author: Kevin Igwe (igwekevin@gmail.com)

ABSTRACT This study investigated the capability of machine learning to analyze and predict individuals' motives within an experimental game context. Although humans predict the motives of others to respond appropriately, these motives often overlap and are difficult to tease apart in social exchange contexts. An act of reciprocated favor, for example, could equally be motivated by parochial altruism as by self-interest, and human attributions of motives are notoriously biased. Can machine learning effectively predict motives and offer insights into how individuals prioritize overlapping motives? We analyzed motives in an experimental social exchange game using a Multilayer Perceptron (MLP), Decision Tree (DT), Random Forest (RF), Gaussian Naïve Bayes (GNB), K-Nearest Neighbor (KNN), Support Vector Classifier (SVC), Logistic Regression (LR), and a novel combination of Clustering and Hidden Markov Model (C-HMM). The accuracy, precision, recall, and f1-score were compared in two phases: Phase 1, where individuals focus on a single motive, and Phase 2, where individuals consider multiple motives when making decisions in social exchange. With accuracies of 86.96%, 67.31%, and 70.74% for each class of motives tested in Phase 1, C-HMM outperformed the other models. LR demonstrated the best performance, with an accuracy of 45.57% in Phase 2. Further analysis shows that the strength of relationships with ingroup members is a reliable predictor of reciprocation motives. Our model can be extended to nudging prosocial behavior in human-agent collaborations.

INDEX TERMS Machine learning, motive prediction, social exchange, intent prediction.

I. INTRODUCTION

In the field of human-agent interaction, there is a compelling need for artificial agents to effectively navigate the complexity of human social behavior and elicit prosocial behavior among humans [1]. Among these complex aspects that govern human social interactions, intent and motives play central roles in driving decisions, shaping relationships, and influencing outcomes, especially in social exchange contexts in which individuals' motives for action influence the reaction of the observers [2].

Social exchange is a fluid and dynamic process wherein individuals are required to anticipate the motives of others prior to reaching a favorable decision – a decision that is optimized to assist the individual in achieving their objectives [3]. Individuals with interactive social exchanges can have 1. multiple obscure motives, such as fairness, selfishness, and ingroup favoritism; 2. where the importance of these motives evolves dynamically through interaction; and 3. within contexts characterized by overlapping behavioral demands that lack clear differentiation from one another [3]. Consider the myriad of obscure motives that underlie the act of giving your boss a gift. It may stem from self-interest, norms in the workplace, generosity, or a combination of these factors [4]. The motives are obscure and overlapping; the latter can help confuse observers, producing uncertainty about the motive behind an individual's actions. Humans exploit this uncertainty in their strategic acts. For example, giving your boss a gift while expecting promotion. This raises the question of whether machine learning can analyze and predict motives to reduce or eliminate uncertainty about the motives underlying individuals' actions in social exchanges.

The associate editor coordinating the review of this manuscript and approving it for publication was Yiqi Liu.

Machine learning models excel in solving objective tasks, such as image recognition [5], language translation, and pattern recognition [6]. Recent advancements in artificial intelligence have enabled more applications to tackle subjective tasks, such as intent prediction from text or speech data [7]. However, predicting motive from interaction traces presents a new challenge: While intent predominantly anticipates 'what' actions people or artificial agents will take [8], motives explore the 'why' behind these actions, revealing the underlying reason. Therefore, intent prediction is an individual's aptitude to perceive and understand the goals of another person, as opposed to predicting the reasons for pursuing those goals.

Whereas Schneider et al. [9] have further differentiated intents into "what" "why," and "how" categories based on the information estimated by agents, individuals, or systems, only a limited body of literature has focused on the "why" based intent. The "what" based intents deal with temporal patterns or goals to be achieved. For example, predicting the items customers on an e-commerce website will purchase. Conversely, the "how" based intent deals with the mechanisms in place to achieve the goal. An example of a "how" based intent is predicting the likely method of purchase, including options like online delivery, in-store purchase, and online purchase for in-store collection. The "why" based intent deals with underlying reasons, corresponding to motives. For example, why did a customer choose to purchase a specific product? Predicting the reasons for actions requires an understanding of certain aspects of an individual's mental state, which is an inner representation of features within a specific external context [10]. In turn, this allows us to predict how an individual will behave.

We applied machine learning to analyze motives using data collected in a simple experimental game context [11]. In this game, participants from two 7-player groups were tasked with allocating a single token to any player in each of 40 rounds. They had to make allocation decisions, mindful that players' tokens represented wealth, with token balances reflecting their relative wealth compared to others in each round of the game. Although the intergroup experimental game is less complex than a real-world context, it offers a valuable opportunity to observe and comprehend the dynamics of social interactions and to apply knowledge of theories about motives underlying the actions of individuals in social exchange environments.

The aim of this study is twofold: to compare the performance of machine learning models for predicting motives in a social exchange experimental game, and to present a machine learning analysis of how individuals prioritize motives in the game, as part of ongoing research on nudging prosocial behavior among humans in human-agent interaction. To achieve this aim, this study pursues the following objectives:

1) Apply well-established psychological theories to identify motives within a social exchange context.
2) Generate theory-driven features from the experimental game data for training the models.
3) Multilayer Perceptron (MLP), Decision Tree (DT), Random Forest (RF), Gaussian Naïve Bayes (GNB), K-Nearest Neighbor (KNN), Support Vector Classifier (SVC), and Logistic Regression (LR) were implemented to predict the motives for token allocation in the experimental game.
4) Implement a novel hidden Markov model that harnesses clustering mechanisms to improve motive prediction.
5) Evaluate the models based on the assumption that:
    a. Individuals in the game focus solely on a single motive when making an allocation decision.
    b. Individuals in the game consider multiple motives in making an allocation decision.

The structure of this paper is outlined as follows. We commence with a concise review of relevant literature, followed by an in-depth description of the current study. We then present the methods section that describes the dataset, features, selected algorithms, and novel Cluster Hidden Markov Model (C-HMM). Finally, we analyze the results and discuss our findings in the concluding section.

II. LITERATURE REVIEW

Intent prediction spans numerous fields of research. We highlight the applications of machine learning for predicting the "what" based intent using text and speech data, and review examples of recent literature that focus on using interaction traces instead of text. Lastly, we highlight the current trend on the "Why" based intent.

A. INTENT PREDICTION: FROM "WHAT" TO "WHY" BASED INTENT

Machine learning approaches, such as collaborative filtering, matrix factorization, and Multilayer Perceptron have been widely adopted to predict the goals or objectives of humans in research areas such as recommendation systems. For example, collaborative filtering methods leverage user-item interactions to uncover similarities between users [12], while matrix factorization techniques capture user-item relationships to predict item users like [13]. Deep learning models, especially those that combine neural collaborative filtering and other algorithms, such as multilayer perceptrons, have demonstrated superior performance in predicting user preferences and goals [14]. Natural Language Processing (NLP) and deep learning techniques have been applied to virtual assistants and chatbots to extract user intentions from textual or spoken input. Intent classification models, including recurrent neural networks (RNNs) and transformers, have proven to be effective in understanding user commands, and generating contextually relevant responses. Readers interested in intent prediction from text data can review [15] studies.

Predicting the "what" based intent is a valuable capability. However, it represents only a fraction of behavior, leaving the underlying reasons for actions – the "why" based intent unanswered [9], [16]. Understanding the reason for an action or pursuing a goal is crucial. Recent successes in goal prediction have primarily occurred in fields in which predictions are made from text or speech data. See [12] for a review. Although the proliferation of digital platforms has given rise to plenty of textual and speech data, it has also increased the availability of videos, images, and traces of interaction. Consequently, understanding the reasons for actions from the videos, images and interaction traces is equally important.

To this end, a few researchers [17], [18], [19] focused on the "why" based behavior intent, utilizing virtual data, or combining virtual and textual data. For example, Kofler et al. [19] aimed to predict "why" individuals create and post videos online. In [19], the authors trained machine learning algorithms that utilized both textual and nontext features for prediction. Possible intents, such as "Asking for an opinion" and "Expressing an opinion," were deduced by mining the web for videos and categorizing them based on their descriptions. These categories were presented to Amazon Mechanical Turk workers, who were asked to explain why they thought the video had been uploaded to validate their approach. While the work is not within the context of social exchange, achieving the best result (67% accuracy) reported in the study for predicting the "why" based intent is challenging. Similarly, Jia et al. [18] collected images featuring potential motives expressed in images obtained from online sources. These motives were obtained from human motives taxonomy developed through studies in the field of psychology. Both the motives and their corresponding image samples were then presented to Amazon Mechanical Turk workers, who filtered out non-representative taxonomy from the samples before utilizing them as motives for training machine learning algorithms. The work highlights the critical role of domain knowledge in psychology theories about motives underlying behaviors when training machine learning algorithms to predict motives.

Ignat et al. [17] used the combination of textual and visual information to predict the reasons for performing actions observed in online videos. Ignat, et al. [17] selected possible reasons for the observed actions from an online data repository that describes human actions and the possible reasons underlying them. In some cases, the reasons were verbally stated in the video. The videos and the selected reasons for the observed actions were then used to train machine learning algorithms to recognize reasons underlying actions in online videos. In addition to virtual and textual data, Zhang et al. [20] utilized audio information to offer an even richer context. For example, they used the speaker's expression and tone to identify emotions, such as joking and anger, to improve the machine learning intent prediction task. These studies [17], [18], [19], [20] demonstrate that a machine learning approach to predicting motives is a viable endeavor. Furthermore, they capitalized on the reach context provided by the combination of textual and visual data.

While the combination of text and traces of interaction can provide a rich context for intent prediction, predicting users' purchase goals using only traces of interaction is a popular study focus in fields where clickstream data, a sequence of click events such as browsing a page, adding items to a cart, and buying items is a well-established source of behavioral information [21], [22], [23], [24]. For example, .Hatt and Feuerriegel [24] evaluated several machine learning algorithms, including the hidden Markov model, logistic regression, and a Markov modulated marked point process (M3PP), which considered the sequence of pages visited as well as the time spent on the pages to predict the risk of customer exiting without a purchase on an e-commerce website. The study found that M3PP outperformed the other tested models. While clickstream data is predominantly symbolic as opposed to conversational, it provides valuable insight for machine learning that predicts products a user may purchase or the goal of the users browsing ecommerce websites.

The prediction of "why" based intent using interaction traces like the clickstream and interaction network, however, remains relatively unexplored. This is because predicting the underlying reasons for actions without textual or speech data presents inherent challenges. Psychologists have long explored the analysis of "why" based behavioral intent, commonly referred to as motives [25]. Understanding psychological theories about motives in social interaction could help debunk the reasons underlying actions from interaction traces.

B. PREDICTING MOTIVES IN SOCIAL EXCHANGES

Research examining motives underlying actions has often relied on questionnaire responses to identify individuals' motives. This approach has been favored because of the latent and subjective nature of motives [26]. While questionnaires and surveys have been traditional tools for investigating motives, they have inherent self-report biases and are expensive and time-consuming, particularly in repeated interactions where motives evolve over time. Think about the motives underlying three quick and successive interactions from an individual. The motive underlying the first interaction may not necessarily be the same as that in the third.

Research has shown that motives can be derived from observations [27] and behavioral records in an experimental game [28], [29]. For example, Stoeckart et al. [29] conducted experiments to evaluate how individuals' implicit motives, specifically, the need for power, influenced their choices. The participants were presented with a repeated choice between two buttons. Each button press resulted in a distinct outcome: either the display of a submissive or dominant face. The need for power was interpreted as the desire to obtain an outcome in which the pressed button displayed a submissive face. This interpretation stems from the idea that individuals with a strong desire for power seek to control and impress others [30], [31] and would learn and repeat the action-outcome pair that resulted in the display of a submissive face.

Similarly, Bolle et al. [28] defined motives as response functions based on donation behavior, such as altruism and selfishness, which have been discussed extensively in the literature for decades. The study [28] aimed to identify different motives underlying individuals' donations using a modified version of the solidarity game [32]. Bolle et al. [28] investigated the social exchange dynamics involving two players, referred to as benefactors, and one player, the beneficiary. The benefactors received €10 each, whereas the beneficiary received no money. Benefactors can choose to give some to the beneficiary. To categorize individuals based on their underlying motives, the benefactors were asked how they would donate to the beneficiary, if the other benefactor has donated €x amount. The same question was asked repeatedly for different values of x, and the benefactors were categorized into defined motives for social exchanges, including altruism, warm glow, and selfishness. For example, selfish benefactors did not give anything, while altruistic benefactors gave less as the other benefactor gave more. The study used the log likelihood of giving and found that motives for giving were diverse, with altruism and envy being the most common, and only 40% of players consistently followed a utility function (motive).

While Bolle et al. [28] analyzed motives in a game context that relied on explicitly defined reward values, the open social exchange used in our study has no defined reward for actions. Also, there was no indication of whether the utility function can predict the players' 'unseen' motives. We suggest that analyzing and predicting motives using machine learning, a more complex algorithm, is likely to yield better results. By examining multiple motives in social exchanges, our study builds on and extends the work of Stoeckart et al. [29], who primarily focused on a single implicit motive. Furthermore, we broadened its scope by predicting compound motives within the context of social exchange games.

III. THE PRESENT STUDY

We applied machine learning to predict motives from interaction traces in an interactive social exchange experimental game. The start token of each player and their group identities were visible to each player at the beginning of each token allocation round [11]. In addition, the players' allocation decisions were revealed to each player at the end of each round. Each allocation decision can be influenced by one or more of the identified motives. We assumed that the motives underlying allocations in round $r + 1$ could be predicted by the outcomes (the allocation network) from round r.

We identify motives within the experimental game by leveraging theoretical knowledge about reciprocation [33], [34], support for the underdog or currying of favor with the rich [35], [36], and ingroup favoritism [37] in social exchanges.

- Reciprocation: Motives within interactive social exchanges are predominantly driven by self-interest [38] and served through reciprocation. Individuals make reciprocal exchanges as a means of building a self-benefiting partnership [33], especially within groups [34].
- Fairness vs. Power: Exchange with the underdog is rooted in notions of fairness and equity [36]. Conversely, individuals engage in exchange with the rich to solicit favor [35] and, feel powerful and domineering [30], [31]. The fairness and power motives can be expressed in the experimental game through token allocation to poor and wealthy individuals, respectively.
- Ingroup favoritism: Reputation motives are served through ingroup favoritism, also known as parochial altruism. Individuals identify with the ingroup and develop a sense of belonging, self-esteem, and pride based on their group membership. They favor their ingroup over the outgroup to enhance their self-concept, reputation, and economic gains [37].

We applied the above theoretical knowledge to identify motives in our game data using the observed characteristics of token allocation. Motives were read from the observable features of the recipients in each round. Is the allocation reciprocating favor from the recipient in the previous round? Was the selected recipient poor or rich? Was the selected recipient in the ingroup or outgroup? As described above, motives are not read perfectly or transparently from the behavior; therefore, we rely on theory to make the best guess about the motive informing each exchange. We do this in two ways. First, we consider each motive in isolation using allocation as a definite signal of the isolated motive. For example, considering ingroup favoritism independent of other motives, we classify ingroup allocation as parochial altruistic and outgroup allocations as not. Allocations to the rich are classified as driven by power-seeking motives, while allocations to the poor are classified as fairness-seeking motives. Allocations that reciprocated receipts in the previous round were classified as reciprocation motives.

This is an oversimplification because each allocation may be informed by multiple motives. Allocations to ingroup members might also be reciprocating and directed toward rich or poor members. Our second strategy for motive identification is to classify compound movies consisting of three indicators represented by a three-digit binary, for example, 100. Digit 1 indicates whether the allocation in round r reciprocates a receipt in round r-1 (yes = 1; no = 0). Digit 2 indicates whether the recipient's token balance at the end of round r-1 was above the mean token balance of all players (rich = 1) or below the mean (poor = 0). Digit 3 indicates whether the recipient was in the same group as the allocator (ingroup allocation = 0, outgroup allocation = 1).

The analyses were conducted in two phases. First, we adopted the vary-one-thing-at-a-time (VOTAT) behavioral strategy [39] and assumed that individuals focus on one motive each time they make an allocation decision. Thus, we analyzed these motives in isolation in Phase 1, where each motive is represented as a binary: 0 or 1. Second, we assume that each allocation can be informed by a combination of these motives. Thus, we analyzed the combinations as compound motives in Phase 2, where each compound motive is one of the eight three-digit binary combinations. Tables 1 and 2 show the motives in isolation and compound motives, respectively, as signaled by round *r* recipient features.

TABLE 1. Motive represented in isolation as binary.

Phase 1: motives in isolation			Explanation
Reciprocation	0	1	Non-reciprocal (i.e., 0) or reciprocal (i.e., 1) allocation
Poor vs Rich	0	1	Allocation to a poor (i.e., 0) or rich (i.e., 1) player
Ingroup favoritism	0	1	Allocation to an ingroup (i.e., 0) or an outgroup (i.e., 1) member

TABLE 2. Compound motives. The three-digit binary represent reciprocation, poor vs rich, and ingroup favoritism respectively.

Phase 2: Compound motives	Explanation
000	Non-reciprocal allocation to a poor ingroup member
001	Non-reciprocal allocation to a poor outgroup member
010	Non-reciprocal allocation to a rich ingroup member
011	Non-reciprocal allocation to a rich outgroup member
100	Reciprocating a poor ingroup member
101	Reciprocating a poor outgroup member
110	Reciprocating to a rich ingroup member
111	Reciprocating to a rich outgroup member

By leveraging the predictive capabilities of machine learning to analyze the motives in Phases 1 and 2, we gain insights into how individuals prioritize motives in the game. In essence, the performance of all tested machine learning algorithms would significantly improve when aligned with the phase that more accurately represents how individuals prioritize motives during the experimental game.

The contributions of this study are as follows: 1. It presents a case for predicting motives using interaction traces within social exchanges, 2. applied and evaluated various machine learning models for predicting motives in a social exchange context, and 3. provides insights into how individuals prioritize motives in social exchanges. Lastly, this study adds to the existing body of research focusing on predictive models that can be applied in behavioral economics, assisting in the design of nudges and interventions (in human-agent collaboration) to encourage prosocial behavior and improve public policy initiatives.

IV. METHOD

This section presents the 1. experimental data and the application of theoretical knowledge to engineer features for machine learning models (Fig. 1); 2. selected machine learning models, 3. proposed combination of clustering and the hidden Markov model, which will be compared to machine learning models, and 4. the evaluation metrics.

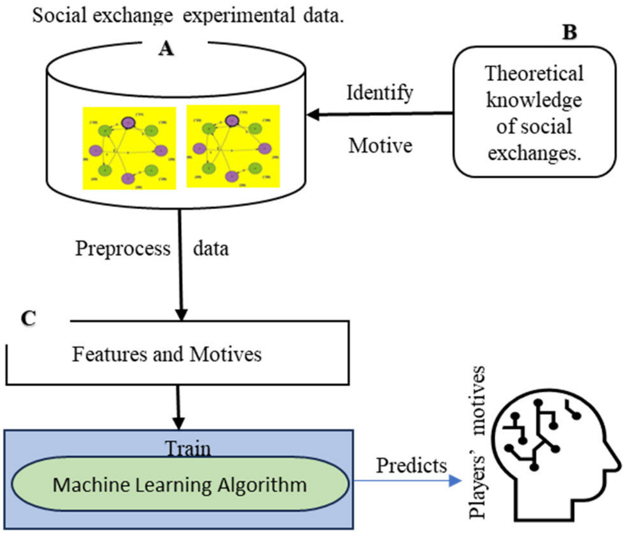

Social exchange experimental data.
A
B
Theoretical knowledge of social exchanges.
Identify
Motive
Preprocess data
C
Features and Motives
Train
Machine Learning Algorithm
Predicts
Players' motives

FIGURE 1. Basic steps in extracting theory-driven motives and features from game data for machine learning training and prediction of social exchange motives.

A. THE DATASET, FEATURES, AND PREPROCESSING

The dataset contains 7,644 observations (Table 3). The dataset contains records of the group identifiers, starting and ending token balances, and directed ties showing player-to-player allocations in each round. The first rounds of each game were not used to reduce noise in the data because the players were likely to allocate their tokens randomly. Next, we describe the features and preprocessing steps.

TABLE 3. Description of experimental game data. While the experiment contain other datasets, we selected those involving two groups.

Descriptive Statistics	Data (2013)	Data (2014)	Total Observations
Games by Condition	5 by 2 = 10	4by 1 = 4	14
Participants per game	14	14	14
Rounds per game	40	40	40
Total No. of allocation	5600	2240	7840
Excluding all round 1 (due to randomness)	5460	2184	n = 7644

B. FEATURES

We predict motives from features of the game context in which allocations are made, predicting motives on round r+1 (next round motives) from features in round r. These observable features differ from the motive attributes discussed above, which are based on the recipient's identity. Here, we identify features based on the identity of the giver or allocator. Our model predicts motives signaled by the characteristics of the recipient from the following allocator features:

- Status of allocator. The allocator's wealth relative to the wealth of all players in round r, as in (1).

$$Status = \frac{a}{\max(A)} \tag{1}$$

  where a is the token balance of an individual in a round, and A is a vector of the token balances of all players in that round. An individual is of high status if the status is greater than the average status in the round; otherwise, the individual is of low status.
- Group identity of the allocator. Either Group 1 or Group 2, constant across the game.
- Allocator previous behavior. Who the allocator gives tokens to in round r. (self = 0, ingroup = 1 or outgroup = 2).
- Reciprocator. Does the allocator have a tendency to reciprocate, measured by whether (i.e., 1) or not (i.e., 0) they reciprocate in round r.
- Ingroup ties. The allocator's receipts from the ingroup in round r. For each player, we measured the strength or bond with the ingroup and outgroup by determining the allocator's receipts from the ingroup and outgroup in each round. Thus, ingroup ties R_{in}^{r} are determined by (2).

$$R_{in}^{r} = \frac{T_{\text{in}}}{N_{\text{in}}} \tag{2}$$

  where T_{in} is the number of tokens received from ingroup members and N_{in} is the number of ingroup members in round r, given that the game rules specify one token allocation per round. Note that $R_{in}^{r} = 1$ (very strong bond) when T_{in}=N_{in} and R_{in}^{r}= 0 (no bond) when T_{in}= 0.
- Outgroup ties. Allocator receipts from outgroup in round r. An individual may have a strong bond with both ingroup and outgroup members. Consequently, outgroup ties are determined by (3). where T_{in} is replaced by T_{out} and N_{in} is replaced by N_{out}.

$$R_{out}^{r} = \frac{T_{\text{out}}}{N_{\text{out}}} \tag{3}$$

- Privilege. Comparison (ratio) of allocator status with the previous receiver in round r. This was measured relative to the wealth of the player to whom the individual had allocated a token in the previous round. Privilege in round r was calculated using Equation (4).

$$P = \frac{\text{Status}_{r}^{a}}{\text{Status}_{r}^{r}} \tag{4}$$

  where $Status_{r}^{r}$ is the status of the player to whom an individual allocated a token in the previous round, and $Status_{r}^{a}$ is the status of the allocator in round r. We set P to $Status_{r}^{a}$ where $Status_{r}^{r} = 0$.

An observation comprises the above seven features along with the corresponding motive (Phase 1) or compound motive (Phase 2). We refer to the first four features as observable because they were visible to the players, and the last three features are theory-driven because they were formulated based on theoretical knowledge of intergroup exchange behavior.

C. SELECTED MACHINE LEARNING ALGORITHMS

While various types of machine learning algorithms, such as classical optimization [40] and nature-inspired [41] exist, we selected the implemented algorithms because of their widespread applications across various intent prediction domains and to explore diverse data analysis approaches. In addition, predicting motives from interaction traces in a social exchange context is relatively unexplored. Thus, it is important to first test well-known basic algorithms. Consequently, we selected the Multilayer Perceptron (MLP) neural network, Decision Tree (DT) classifier, Random Forest (RF) classifier, Gaussian Naïve Bayes (GNB), K-Nearest Neighbor (KNN), Support Vector Classifier (SVC), and Logistic Regression (LR). Furthermore, the performance of these algorithms was benchmarked against a random guessing approach.

1) ARTIFICIAL NEURAL NETWORK

The MLP [42], [43] neural network has three core layers: an input layer, one or more hidden layers, and an output layer. Each hidden layer consists of data processing nodes, called neurons. The number of layers and neurons in each layer can vary depending on the problem. Whereas neurons in adjacent layers have fully weighted connections, neurons within the same layer are not connected. The back-propagation algorithm is commonly employed to train MLPs. During the training process, computations were performed to generate outputs for each input and existing weight. Consequently, the weights were adjusted based on the difference between the output of the network and the intended target output.

2) DECISION TREE

DT uses a rule-based tree structure to split data into predefined classes. The formulation of decision rules for data splitting depends on the specific attributes and characteristics of the dataset. The decision tree learns these rules, identifies distinct subsets of data, and subsequently employs these rules for individual instances within the dataset to predict the target class. Several algorithms have been proposed for constructing decision trees. These include Iterative Dichotomizer 3 (ID3) [44], C4.5 [45], and Classification and Regression Tree (CART).

3) RANDOM FOREST

RF uses bagging to enhance model precision by combining the capabilities of multiple decision trees. It uses the average or median output of the decision trees for continuous targets, and the mode of the decision trees for discrete class labels. Given M decision trees in a RF, the output $F_{rf}^{M}\left(X^{i}\right)$ of the RF is calculated in (5), where X^{i} is the ith instance of a dataset $X = x_1, x_2, x_3, \ldots, x_n$ having n features, and $T_{tree}\left(x^{i}\right)$ is the output of a decision tree that takes the *ith* instance of the dataset.

$$F_{rf}^{M}\left(X^{i}\right)=y=\begin{cases}\frac{1}{M}(\sum_{m=1}^{M}T_{tree}(x^{i})), & y \text{ discrete}\\ \max(count(T_{tree}\left(x^{i}\right)), & y \text{ continoues}\end{cases} \tag{5}$$

4) GAUSSIAN NAÏVE BAYES

The GNB classifier is based on Bayes' theorem, which is calculated using the conditional probability. Given the *ith* instance of *n*-dimensional feature, the GNB classifier calculates the probability that the instance belongs to a specific class based on the Gaussian distribution of class characteristics. After calculating the probability of an instance belonging to all classes, the class with the highest probability value was considered the predicted class. GNB computes the output relatively fast in large datasets. Consequently, it has been applied to numerous classification problems, including the prediction of sleep behavior [46]. Naïve Bayes assumes that the class to be predicted is conditionally independent, as represented in (6), where C_i is the *ith* class represented by *n*-dimentional vector $X = x_1, x_2, x_3 \ldots, x_n$.

$$P(X)=\frac{P(X)\,P(C_i)}{P(X)} \tag{6}$$

5) K-NEAREST NEIGHBOR

The KNN algorithm [47] compute the class of an instance of a dataset based on the similarity or distance measure. KNN assigns an instance to a class label that is closely related or close to the given instance. Thus, KNN is a nonparametric classifier that uses decision rules to assign a class label to an instance in a dataset.

6) LOGISTIC REGRESSION

LR is a classification algorithm that is often used to benchmark other algorithms. It uses the logit function to estimate the parameters of the model. Given $X = x_1, x_2, x_3 \ldots, x_n$ features, LR is formulated by (7), where $b_1x_1 + b_2x_2 + \ldots + b_nx_n$ is the liner regression of output $\hat{y}$ on X.

$$F_{lr}=\frac{1}{1+e^{-z}}=\frac{1}{1+e^{-(b_1x_1+b_2x_2+\ldots+b_nx_n)}} \tag{7}$$

7) SUPPORT VECTOR CLASSIFIER

The SVC aims to generate an optimal separation surface that classifies the dataset, thereby maximizing the generalization capabilities of the model. Whereas empirical risk minimization aims to minimize the mean squared error on the given dataset, SVC uses the Structural Risk Minimization (SRM) principle to find a hyperplane that separates the dataset with a greater margin. SVC was originally developed for binary classification; however, it has been developed for multiclass problems [48] using a one-vs-all approach.

D. SYSTEM SPECIFICATIONS AND PARAMETERS

Machine learning algorithms were implemented using the Scikit-Learn packages [49] in Python.

TABLE 4. Parameters selected via the random search.

Algorithms	Parameters
MLP	Solver: sgd learning_rate: adaptive hidden_layer_sizes: (10, 30, 20) Alpha: 0.001 Activation: relu
RF	n_estimators: 1000 min_samples_split: 2 min_samples_leaf: 4 max_depth: 5 Bootstrap: True
SVC	C: 20 kernel : rbf gamma : 0.5
DT	criterion: gini max_depth: 5
LR	Penalty: l2 C: 0.3907
GNB	var_smoothing: 0.2812
KNN	n_neighbors: 42

The algorithms were executed on a Windows 11 Pro 64-bit machine, with Intel® Core™ i7 with 4G RAM and 8 CPUs running at 2.8GHz speed. For each selected algorithm, we use a random search over 20 iterations to select the best parameter. Each iteration used a 3-Fold cross-validation to minimize the effect of imbalanced classes and the relatively small size of the dataset. Table 4 lists the main parameters used in each algorithm.

E. EVALUATION METRICS

The models were evaluated using the accuracy scores. We counted the number of correctly predicted motives. However, the accuracy is not a true reflection of a model's performance for imbalanced classes. To ensure a more accurate measure, we calculated weighted precision, recall and f1-scores from the multiclass confusion matrix detailed in [50]. Precision measures the actual number of samples belonging to a class among the total number of samples identified as belonging to that class. The value ranged from 0 or 0% (no identification) to 1 or 100% (perfect identification). Recall measures the model's ability to discriminate samples that do not belong to a particular class. Again, the value ranged from 0 or 0% (no discrimination) to 1 or 100% (perfect discrimination). F1-score measures the balance between the precision and recall. The value ranged from 0 to 1. A higher value indicated a better score.

Each model was executed via 10-fold cross-validation to reduce the impact of imbalanced classes [51]. Within this iterative approach, the dataset was divided into ten distinct groups, with each iteration involving the training of a new model on nine of these groups and its evaluation on the remaining one. Each group was used once for evaluation and nine times for training of a new model. Consequently, this yielded a total of 10 iterations, resulting in a computation of 10 fits, on which the evaluation metrics were applied for both the training and test data. Research [52] has shown that 10-fold works well for most datasets. In addition, we generated ten random guesses and computed the average accuracy, precision, recall, and f1-score.

Furthermore, we employed a widely used oversampling technique, the synthetic minority over-sampling technique for nominal and continuous (SMOTENC) datasets, as described in [53], to evaluate the impact of artificially increasing the number of samples in a minority class. Given a data sample, this technique generates new samples using information about the nearest neighbors. SMOTENC was applied to the nine groups used for training and not to the test data, thereby preventing data leakage – a situation in which information about the test data is unintentionally shared with the model during training.

Considering the preliminary result obtained during parameter tuning and the difficulty in predicting motives [19], we adopted the suggestion by Zaki et al. [54] that clustering behaviors can ensure high performance of a hidden Markov model in predicting tasks. Hence, we developed a combination of the clustering and a hidden Markov model (C-HMM) and compared its performance with that of the selected machine learning algorithms. C-HMM is described below.

F. COMBINATION OF CLUSTERING AND A HIDDEN MARKOV MODEL (C-HMM)

We performed a cluster analysis to identify patterns of feature co-occurrence in individuals at each allocation point. Thus, each cluster group observation is based on their feature co-occurrence. The partitioning around the medoids (PAM) clustering algorithm in R [55] was applied with Gower distance [56] as the distance measure. Although other distance measures, such as Euclidean and Manhattan [57], can be used, the Gower distance is very useful and performs well in a domain with mixed data types, categorical and non-categorical data [56], [58]. We used silhouette width, also referred to as the silhouette coefficient [59], to determine the optimal number of clusters. The silhouette width measures the within-cluster cohesion and separation distance between clusters. The silhouette width of a data sample ranges from -1 to 1, where a large s (near 1) implies good clustering, a small s (near 0) implies that the data sample lies between clusters, and a negative s implies that the data sample is placed in the wrong cluster. The optimal number of clusters was determined by averaging across five runs for each value of k clusters, k ranging from 2 to 20.

Fig. 2 illustrates that the optimal number of clusters for the C-HMM is six, as evidenced by the average silhouette width of 0.632. In Fig. 3, we project the observations in each cluster onto a two-dimensional plane using a t-distributed stochastic neighbor embedding (t-SNE) [60]. While the current study refrained from explicitly interpreting each cluster, statistical insights, such as the motive distribution across clusters, can be harnessed by the HMM during the training of initial, transition, and emission probabilities to enhance motive prediction.

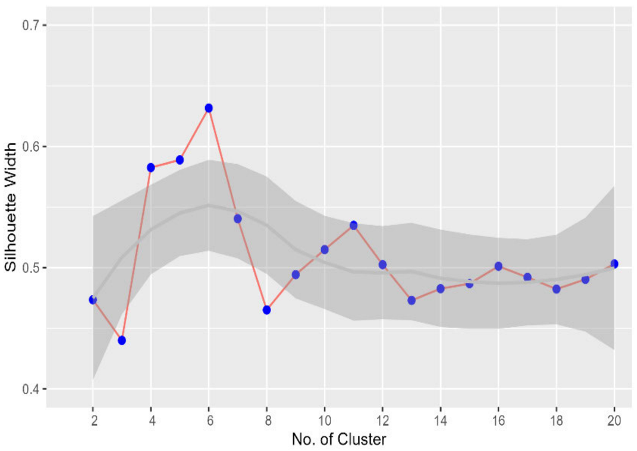


FIGURE 2. Silhouette width for determining the optimal number of clusters. Higher numbers imply a more optimal number of clusters.

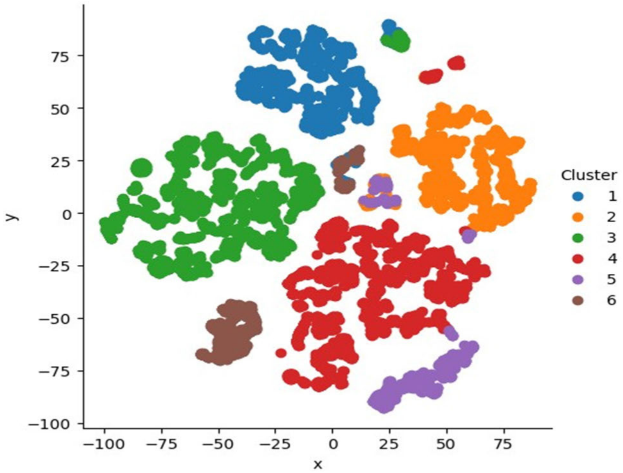


FIGURE 3. Projection of the observations in each cluster onto a two-dimensional plain, showing cluster separations. We used perplexity of 25, and 1000 steps as the t-SNE parameters.

C-HMM was developed to consider the round-by-round structure of the game. It was implemented to predict motives from players' past behaviors. Traditionally, a hidden Markov model has hidden states in which observables are conditioned. See [61] for a review. In our study, we define a state in the C-HMM as the cluster s to which observation o belongs. In this context, an observation represents the motive (or compound motive) underlying the allocation made by the individual in that specific round. Thus, we implemented a C-HMM with the following parameters, where $s \in S$ is a vector of all possible states (clusters), and $o \in O$ is the vector of all possible observations (isolate motives or compound motives).

- s: State (cluster) of the allocator at round r.
  o: Motive (or compound motive) of the allocator at round r.

- α : vector of length 6 showing the probabilities of starting from each state in S.
- β : a 6-by-6 transition matrix showing the probabilities of moving from one state to another.
- γ : 6-by-n emission matrix showing the probability of each observation o, given each state s. Here, n is 2 for the isolate motive and 8 for the compound motive.

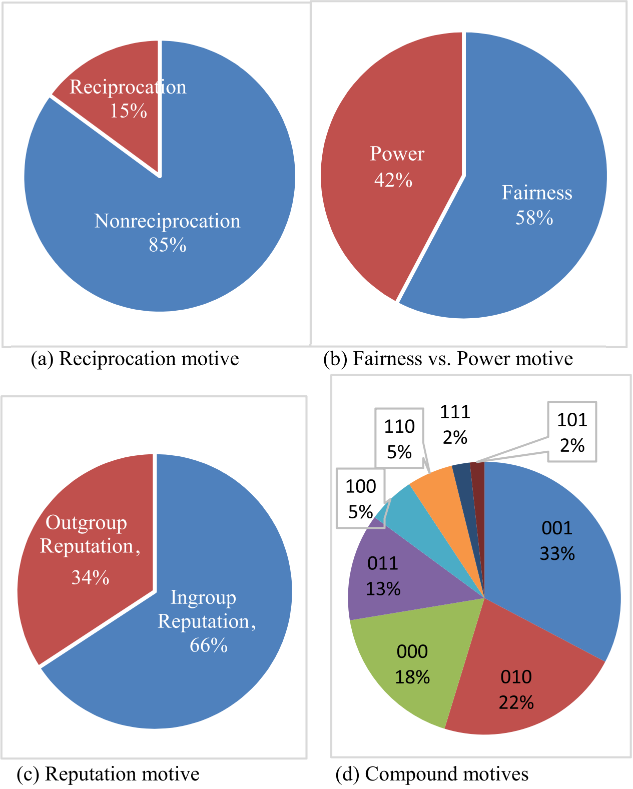
Given $\theta = (\alpha, \beta, \gamma)$, and the sequence of s and o of a player in the previous rounds, we predict the next motive by computing the most likely state from the previous state and then computing the probability of observing each motive given that state. Thus, this study implemented a hidden Markov model that uses round-forward chaining cross-validation for training and testing. Round-forward chaining starts by using data from rounds 1 to r to train the hidden Markov model, which is tested by predicting the exchange decisions in round $r + 1$. Next, it includes the prediction from round $r + 1$ in the training set, and predicts round $r + 2$. This process continued until the last round was predicted. The C-HMM was retrained after each round of the game, and the transition and emission probabilities changed per round. Although this process is computationally more expensive than the traditional hidden Markov model, which is trained once, it was implemented to accommodate the dynamics of the exchanges over time. All trainings were performed using the Baum-Welch expectation-maximization algorithm described in the seminal work of .Rabiner [62]. Owing to the stochastic nature of the model, ten runs were performed for each round, and the result was averaged over rounds to compute the final accuracy.

V. RESULT AND DISCUSSION

This study investigated the performance of Multilayer Perceptron, Decision Tree, Random Forest, Gaussian Naïve Bayes, K-Nearest Neighbors, Support Vector Classifier, and Logistic Regression, in terms of accuracy, precision, recall, and f1-score, in predicting players' motives in a social exchange experiment. We present the results in three steps: First, we provide a preliminary analysis of the results, justifying the use of methods suitable for imbalanced classes, and the inclusion of theory-driven features for predictive tasks. Next, we present detailed results for predicting motives in Phases 1 and 2, using the features and parameters informed by the preliminary analysis. Finally, we present and compare the performance of C-HMM with that of the other models in Phases 1 and 2. We report the average performance of the models on the test data over 10-fold cross-validation, unless stated otherwise. Whereas the weighted average in Scikit-Learn is suitable for imbalanced classes, it results in an accuracy metric equivalent to recall. Consequently, we omitted recall scores from the report.

A. PRELIMINARY ANALYSIS

The class distribution in Fig. 4 shows that reciprocation, with a 15% occurrence, is the least common motive for token allocation. The fairness-motivated token allocation was more than the need for power by 16%, while ingroup reputation occurred more than reputation within the outgroup. Overall, the identified motives within the dataset were not equally distributed, confirming the need for a performance metric suitable for imbalanced classes.



(a) Reciprocation motive (b) Fairness vs. Power motive

(c) Reputation motive (d) Compound motives

FIGURE 4. (a)-(c) show the distribution of motives in Phase 1, indicating imbalanced classes. (d) The distribution of compound motives in Phase 2 indicates imbalanced classes. The three-digit binary representation is presented in Table 2.

We investigated the usefulness of theory-driven features by comparing the accuracy and f1-score of the models trained with and without theory-driven features. The inclusion of theory-driven features enhanced the performance of the seven selected algorithms in both phase 1 and 2. As shown in Fig. 5, theory-driven features enhanced the accuracy of each model except for SVC, which was also the least performing model in Phase 2. This is not surprising, as SVC is known to underperform when dealing with imbalanced classes [63]. Both observable and theory-driven features were used in all the results presented.

B. RESULT OF PREDICTING ISOLATE AND COMPOUND MOTIVES

1) PHASE 1: ISOLATE MOTIVE

Fig. 6 compares the average accuracy of the models in predicting the reciprocation motive with and without the SMOTENC oversampling technique applied to the training data. Notably, the accuracy of each model was better without the application of SMOTENC, as shown in Fig. 6.

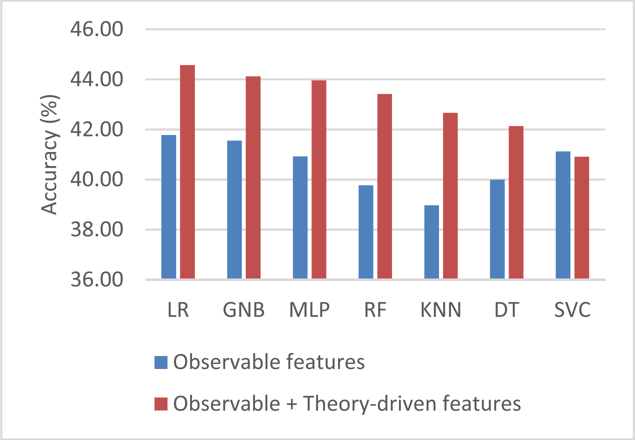


FIGURE 5. Models' accuracies with and without theory-driven features. Except for SVC, the accuracies of the models were better with the inclusion of theory-driven features.

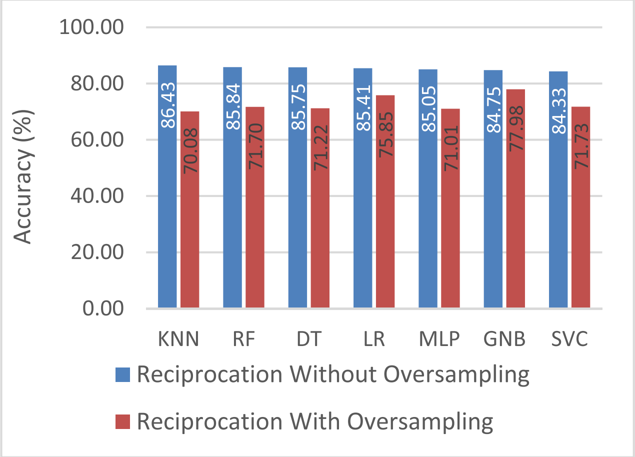


FIGURE 6. Models' accuracies with and without the application of SMOTENC oversampling technique, showing better accuracies in predicting Reciprocation without the application of SMOTENC.

TABLE 5. The results of predicting reciprocation.

Model	Accuracy (%)	Precision (%)	f1_score (%)	std dev of Accuracy
KNN	86.43	83.06	83.22	0.030
RF	85.84	81.88	81.55	0.025
DT	85.75	82.71	83.30	0.031
LR	85.41	82.10	82.35	0.015
MLP	85.05	81.44	81.16	0.020
GNB	84.75	81.31	81.75	0.028
SVC	84.33	81.16	81.92	0.022

TABLE 6. The results of predicting ingroup favoritism.

Model	Accuracy (%)	Precision (%)	f1_score (%)	std dev of Accuracy
RF	65.32	65.10	63.80	0.069
MLP	65.25	65.41	63.28	0.080
LR	64.78	65.00	62.20	0.070
DT	64.35	64.49	63.11	0.072
GNB	63.58	63.85	59.81	0.068
KNN	63.51	63.23	60.86	0.070
SVC	61.63	61.10	59.84	0.050

Although SMOTENC has been known to enhance model performance in predictions involving imbalanced classes, there have been cases where its application led to a worsening of model performance. See [64] for an example. In our study, SMOTENC did not enhance model performance. Therefore, we present the detailed results obtained without the application of SMOTENC.

Tables 5 – 7 show the detailed results, including the models' precision, and f1-score for predicting the reciprocation motive, fairness versus power motive, and ingroup favoritism, respectively. In Table 5, the difference between the accuracy scores of the best (KNN, 86.43%) and least (SVC, 84.33%) well-performing models is 2.10%, demonstrating competitive performance. Therefore, it makes sense to compare the f1-scores, where KNN, RF, and DT have the highest scores. The f1-scores demonstrate the efficacy of these three algorithms in predicting the reciprocation motive within the game.

As shown in the preliminary analysis, the class distribution for the power versus fairness motive was relatively balanced. This suggests that model accuracy can serve as a reliable indicator of a model's fit. Notably, RF exhibits better accuracy and f1-score than the other models, as shown in Table 6. The difference between the RF accuracy (65.32%) and accuracy (61.63) of the least well-performing model (SVC) was 3.69%. This difference is slightly larger than that observed between the best and least performing models in predicting reciprocation. This suggests that predicting fairness versus power motive is slightly more challenging for the models.

In general, the Support Vector Classifier (SVC) consistently ranks as the least well-performing model across all predictive tasks. This is not surprising because SVC tends to underperform when dealing with imbalanced classes and features that exhibit overlapping characteristics across different classes [63]. Moreover, SVC overfits the data, as evidenced by the substantial difference (11.59%) in Table 7 between SVC training accuracy (74.64%) and test accuracy (63.05%), in contrast to the minor difference (1.51%) between LR training accuracy (68.63%) and test accuracy (67.12%) in predicting ingroup favoritism.

TABLE 7. The results of predicting poor vs rich.

Model	Training Accuracy (%)	Test Accuracy (%)	Precision (%)	f1_score (%)	std dev of Accuracy
LR	68.63	67.12	66.30	65.05	LR
GNB	68.08	66.59	65.18	63.27	GNB
MLP	69.37	65.89	64.41	62.81	MLP
KNN	70.63	65.03	62.48	60.73	KNN
DT	69.96	64.85	62.80	60.56	DT
RF	70.58	64.74	63.51	60.74	RF
SVC	74.64	63.05	60.77	59.95	SVC

Using the best models, KNN, RF, and LR in Tables 5 – 7, we conducted feature importance tests to gain insight into the significance of each feature for the three predictive tasks.

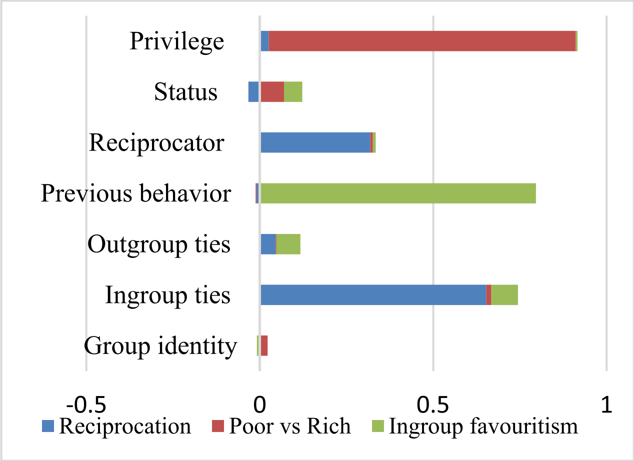


FIGURE 7. Feature importance graph, showing that Ingroup ties, Privilege and Previous allocation behavior are more important in predicting Reciprocation, Poor vs Rich and Ingroup favoritism respectively.

Notably, we found that 'ingroup ties' was the strongest predictor of reciprocation motive (Fig. 7). This finding aligns with the existing literature, suggesting that individuals expect to reciprocate more often by ingroup members than outgroup members [65], [66].

Additionally, we observed that 'previous allocation' serves as the most reliable predictor of ingroup favoritism, whereas 'privilege' stands out as the top predictor of the fairness versus power motive. This observation was not surprising, as 'privilege' captures a player's previous disposition toward others' status. The results confirm that players' motives for fairness or power do not undergo drastic changes over time.

2) PHASE 2: COMPOUND MOTIVE

The models have lower accuracies when predicting compound motives compared than when predicting isolated motives. Similar to what was observed in Phase 1, SMOTENC did not improve the accuracy of the models. Instead, the models performed well on the training data but were significantly lower on the test data. Hence, we report the results obtained without using SMOTENC.

Table 8 shows that LR outperformed the other models in terms of the accuracy score and ranked second in terms of the f1-score. We did not expect these outcomes because LR is less complex than other algorithms. However, simpler models such as LR are less prone to overfitting, whereas complex models are more likely to overfit the data.

We compared the results to a random guess to ensure that the outcomes were not random and to demonstrate that, although the accuracies were low, they were significantly higher than a random guess. The difference between the average accuracy of 10 random guesses and that of LR was 32.08%, affirming that the algorithms indeed learned from the provided data.

Additionally, we assessed the contribution of each feature using LR, the best performing model. Fig. 8. illustrates that 'status' is more influential in predicting compound motives, while 'outgroup ties' is less important in the prediction by LR.

TABLE 8. The results of predicting compound motives.

Model	Accuracy (%)	Precision (%)	f1_score (%)	std dev of Accuracy
LR	44.56	44.26	41.17	0.072
GNB	44.12	44.13	41.44	0.070
MLP	43.96	42.49	39.96	0.060
RF	43.42	39.79	37.54	0.065
KNN	42.66	40.60	38.54	0.063
DT	42.13	38.98	36.60	0.053
SVC	40.91	39.52	37.81	0.052
Random Guess	12.49	20.52	14.47	0.003

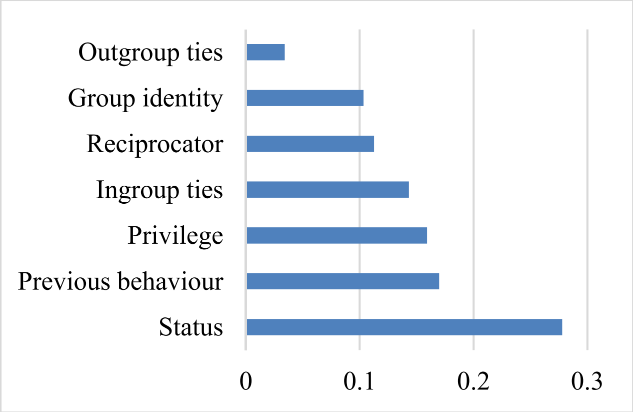


FIGURE 8. Feature importance graph, showing that status is the most important feature in predicting compound motives.

C. PERFORMANCE COMPARISON WITH C-HMM

We report the results of the C-HMM implemented to obtain better predictions. Table 9 shows the performance of the C-HMM in Phase 1. C-HMM predicted the reciprocation motive with 86.96% accuracy, ranking first and better than KNN by 0.53%. Similarly, C-HMM predicted fairness versus power motive with 67.31% accuracy and ingroup favoritism with 70.74% accuracy, ranking first in both predictions with differences of 1.99% and 3.62%, respectively, compared to the best model reported in Tables 6 and 7.

TABLE 9. The C-HMM results in phase 1 and 2.

Motive	C-HMM Accuracy (%)	C-HMM Precision (%)	C-HMM f1_score (%)	std dev of C_HMM Accuracy
Reciprocation	86.96	79.61	67.85	0.59
Poor vs Rich	67.31	66.38	65.93	1.289
Ingroup favoritism	70.74	67.09	67.14	1.781
Compound Motives	43.49	25.52	28.53	2.271

We investigated the accuracy of the C-HMM to gain deeper insight into the performance of the model. Fig. 9 confirms that C-HMM exhibited a bias toward the larger class when predicting the reciprocation motive, resulting in consistently high accuracy that did not show improvement across rounds. Additionally, the standard deviations of accuracy, represented by the width of the shadows on the graphs, were greater for ingroup favoritism than for the other motives, indicating a more varied accuracy per run.

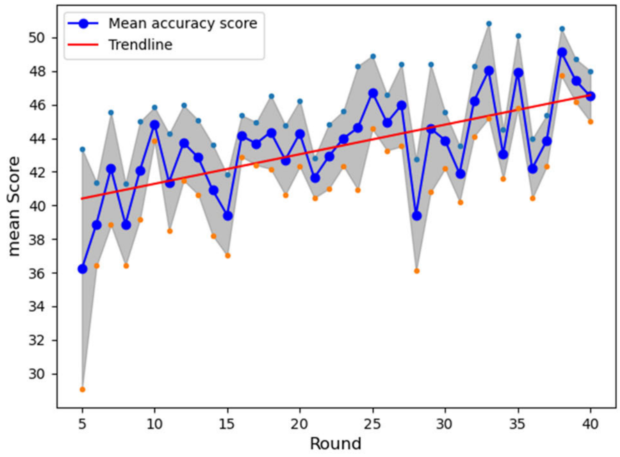


FIGURE 9. The C-HMM accuracies, showing improvements in predicting motives over rounds in Phase 1.

The accuracy of C-HMM, as shown in Table 9, ranks 4th and is higher than that of RF, KNN, DT, and SVC in predicting compound motives. Notably, the standard deviation of the C-HMM was larger and distinct from the range of standard deviations exhibited by the other models (Table 8). Additionally, the f1-score of C-HMM is considerably lower than that of the other models, resulting in C-HMM being ranked as the least well-performing model based on the f1-score.

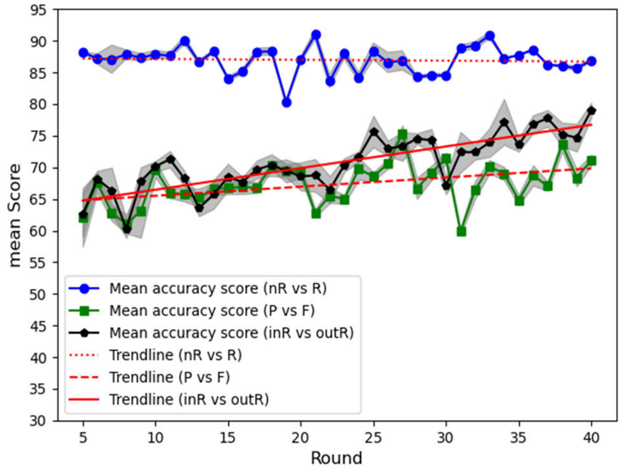


FIGURE 10. The C-HMM accuracy graph, showing that the model's accuracy in predicting compound motives improved over rounds.

Again, we conducted further analysis of C-HMM. As shown in Fig. 10, the accuracy of the C-HMM displays a consistent improvement across rounds, accompanied by decreasing standard deviations. Although the standard deviations were above 1.0, the decreasing pattern indicated improvement with more data.

D. THE IMPACT OF DATASET SIZE ON MODEL PERFORMANCE

Dataset size has effect on the performance of many machine learning algorithms [67]. For example, SVC underperforms when trained with a small dataset size. Using compound motives, we evaluated the impact of dataset size on the accuracy of the machine learning algorithms to gain insight on how the algorithms might perform with increased dataset size.

We generated eight distinct training set sizes: 637, 1456, 2275, 3094, 3913, 4732, 5551, and 6370. Using 3-fold cross-validation for each training set size, we compute and graph the average training and testing accuracy for each model.

Fig 11 shows the accuracy of each model in predicting compound motives using various training set sizes. The figure indicates that increases in the dataset size slightly improved the accuracy of each model. For example, SVC improved from 31% accuracy when the training set size was 637 to 40% with a training set size of 6370, reflecting a 9% improvement. However, models like LR and KNN exhibited improvements of less than 5%. Except for SVC, an increase in the dataset size will have minimal impact on the accuracy of the models, as evidenced by the training and testing accuracy lines showing little variations and almost converging. Therefore, we conclude that increasing the datasets can improve the accuracy of the models, especially that of the SVC.

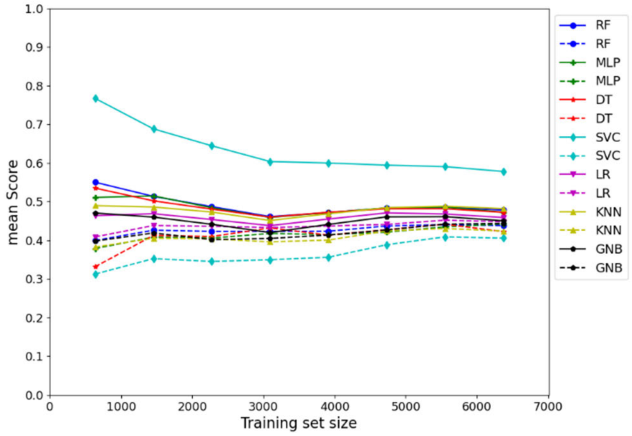


FIGURE 11. The average training and testing accuracy (dotted lines) of the models at various dataset sizes.

VI. CONCLUSION

Understanding the motives underlying others' actions is pivotal for informed decision-making within a social context. However, the complex interplay between uncertainty and overlapping motives makes it difficult for humans to predict. This study compared the performance of a Multilayer Perceptron, Decision Tree, Random Forest, Gaussian

Naïve Bayes, K-Nearest Neighbors, Support Vector Classifier, Logistic Regression, and a combination of clustering and hidden Markov model (C-HMM) to analyze and predict motives in a social context, represented by an experimental game. The motives, namely, Reciprocation, Fairness vs. Power, Ingroup favoritism, theoretically deduced from secondary data derived from game-like experiments, were succinctly represented as binary (Phase 1) or three-digit binary (Phase 2).

This analysis aimed to provide insights into how individuals prioritize motives in social exchanges. Phase 1 assumed that individuals concentrate on a single motive during allocation decisions, whereas Phase 2 suggests that individuals in the game consider multiple motives in making an allocation decision. Based on these assumptions, we trained eight machine learning algorithms. We analyzed how individuals prioritize motives within the experiment by comparing the performance of all the tested machine learning algorithms when aligned with Phase 1 and Phase 2.

The dataset, preprocessed from the game context, contained four observable features and three theory-driven features. The Python programming language and Scikit-Learn machine learning library were utilized to implement and compare the selected algorithms. These algorithms were trained and tested using 10-fold cross-validation to mitigate the effects of imbalanced class distributions and small dataset size.

The C-HMM performed better than the selected machine learning models in terms of accuracy across all predictions in Phase 1. Logistic regression outperformed the other selected algorithms in predicting motives in Phase 2, whereas the performance of C-HMM was ranked fourth based on accuracy. Notably, in Phase 2, the results revealed that status was a stronger predictor of compound motives than the other features. Additionally, the strength of relationships with the ingroup had a more significant influence on motives than mere group membership. The performance of the models was best in Phase 1. We conclude that Phase 1 more accurately represents how individuals prioritize motives in the experimental game.

Interestingly, we discovered that the behavior of individuals in the game was in line with theoretical assumptions [65], [66] that individuals expect that ingroup members will reciprocate a favor more than outgroup members. This is evident from the results of Phase 1, where the feature, 'Ingroup ties' emerged as the best predictor of reciprocation motive, against our expectation that previous 'Reciprocator' would be a better predictor of reciprocation motive.

In line with many behavioral studies that use machine learning to predict motives, including those that rely on self-reporting, a critique that can be made of this study is whether the ground truth is accurate. Self-reporting introduces inherent biases, including impression management, such as faking and lying [68]. This raised the question: why should we place trust in individuals' self-disclosures? Consequently, self-reporting does not serve as a benchmark for accurately identifying motives. In our study, we combined previous moves and the knowledge of well-established theories on social exchanges to identify motives signaled by the recipient's identity. While this approach mitigated the trust issues associated with self-reporting, it introduced a dependency on, and necessitated trust in, psychological theories about motives in social exchanges.

The insights provided by theories on social exchange behaviors help predict motives in the experimental game. Conversely, our approach can help social scientists identify the factors that predict motives. The C-HMM algorithm introduced in this study performed competitively well with other selected machine learning algorithms. However, its performance indicated that it was susceptible to an imbalanced dataset. An improvement would be to incorporate a search strategy that penalizes the algorithm when it is biased toward a class.

The ability to predict the underlying motives of actions is crucial for ongoing efforts to enhance human-agent collaboration and foster equitable social dynamics. For example, agents that predict motives can effectively nudge prosocial behavior based on predicted outcomes. This study thus makes a valuable contribution to the existing literature on human-agent collaboration in interactive social exchanges.

REFERENCES

[1] A. Paiva, F. Santos, and F. Santos, "Engineering pro-sociality with autonomous agents," in *Proc. AAAI Conf. Artif. Intell.*, 2018, vol. 32, no. 1, pp. 7994–7999.

[2] R. W. Carlson and J. Zaki, "Good deeds gone bad: Lay theories of altruism and selfishness," *J. Experim. Social Psychol.*, vol. 75, pp. 36–40, Mar. 2018.

[3] S. Suzuki and J. P. O'Doherty, "Breaking human social decision making into multiple components and then putting them together again," *Cortex*, vol. 127, pp. 221–230, Jun. 2020.

[4] R. Oliveira, P. Arriaga, F. P. Santos, S. Mascarenhas, and A. Paiva, "Towards prosocial design: A scoping review of the use of robots and virtual agents to trigger prosocial behaviour," *Comput. Hum. Behav.*, vol. 114, Jan. 2021, Art. no. 106547.

[5] H. Jiang, Z. Diao, T. Shi, Y. Zhou, F. Wang, W. Hu, X. Zhu, S. Luo, G. Tong, and Y.-D. Yao, "A review of deep learning-based multiple-lesion recognition from medical images: Classification, detection and segmentation," *Comput. Biol. Med.*, vol. 157, May 2023, Art. no. 106726.

[6] D. Khurana, A. Koli, K. Khatter, and S. Singh, "Natural language processing: State of the art, current trends and challenges," *Multimedia Tools Appl.*, vol. 82, no. 3, pp. 3713–3744, Jan. 2023.

[7] V. S. Sadasivan, A. Kumar, S. Balasubramanian, W. Wang, and S. Feizi, "Can AI-generated text be reliably detected?" 2023, *arXiv:2303.11156*.

[8] E. Bonchek-Dokow and G. A. Kaminka, "Towards computational models of intention detection and intention prediction," *Cogn. Syst. Res.*, vol. 28, pp. 44–79, Jan. 2014.

[9] M. F. Schneider, M. E. Miller, T. C. Ford, G. Peterson, and D. Jacques, "Intent integration for human-agent teaming," *Syst. Eng.*, vol. 25, no. 4, pp. 291–303, Jul. 2022.

[10] P. Johnsonlaird, "Mental models in cognitive science," *Cognit. Sci.*, vol. 4, no. 1, pp. 71–115, Mar. 1981.

[11] K. Durrheim, M. Quayle, C. G. Tredoux, K. Titlestad, and L. Tooke, "Investigating the evolution of ingroup favoritism using a minimal group interaction paradigm: The effects of inter- and intragroup interdependence," *PLoS ONE*, vol. 11, no. 11, Nov. 2016, Art. no. e0165974.

[12] S. Farshidi, K. Rezaee, S. Mazaheri, A. H. Rahimi, A. Dadashzadeh, M. Ziabakhsh, S. Eskandari, and S. Jansen, "Understanding user intent modeling for conversational recommender systems: A systematic literature review," 2023, *arXiv:2308.08496*.

[13] Y.-N. Chen, M. Sun, and A. I. Rudnicky, "Matrix factorization with domain knowledge and behavioral patterns for intent modeling," in *Proc. NIPS Workshop Mach. Learn. SLU Interact.*, 2015, pp. 1–7.

[14] X. He, L. Liao, H. Zhang, L. Nie, X. Hu, and T.-S. Chua, "Neural collaborative filtering," in *Proc. 26th Int. Conf. World Wide Web (WWW)*, vol. 2017, pp. 173–182.

[15] M. Hamroun and M. S. Gouider, "A survey on intention analysis: Successful approaches and open challenges," *J. Intell. Inf. Syst.*, vol. 55, no. 3, pp. 423–443, Dec. 2020.

[16] C. Kofler, M. Larson, and A. Hanjalic, "User intent in multimedia search: A survey of the state of the art and future challenges," *ACM Comput. Surv.*, vol. 49, no. 2, pp. 1–37, Jun. 2017.

[17] O. Ignat, S. Castro, H. Miao, W. Li, and R. Mihalcea, "WhyAct: Identifying action reasons in lifestyle vlogs," 2021, *arXiv:2109.02747*.

[18] M. Jia, Z. Wu, A. Reiter, C. Cardie, S. Belongie, and S.-N. Lim, "Intentonomy: A dataset and study towards human intent understanding," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 12981–12991.

[19] C. Kofler, S. Bhattacharya, M. Larson, T. Chen, A. Hanjalic, and S.-F. Chang, "Uploader intent for online video: Typology, inference, and applications," *IEEE Trans. Multimedia*, vol. 17, no. 8, pp. 1200–1212, Aug. 2015.

[20] H. Zhang, H. Xu, X. Wang, Q. Zhou, S. Zhao, and J. Teng, "MIntRec: A new dataset for multimodal intent recognition," in *Proc. 30th ACM Int. Conf. Multimedia*, Oct. 2022, pp. 1688–1697.

[21] B. Requena, G. Cassani, J. Tagliabue, C. Greco, and L. Lacasa, "Shopper intent prediction from clickstream e-commerce data with minimal browsing information," *Sci. Rep.*, vol. 10, no. 1, p. 16983, Oct. 2020.

[22] X. Wang and S. Kadioglu, "Dichotomic pattern mining with applications to intent prediction from semi-structured clickstream datasets," 2022, *arXiv:2201.09178*.

[23] A. Rashid, M. S. Farooq, A. Abid, T. Umer, A. K. Bashir, and Y. B. Zikria, "Social media intention mining for sustainable information systems: Categories, taxonomy, datasets and challenges," *Complex Intell. Syst.*, vol. 2021, pp. 1–27, Apr. 2021.

[24] T. Hatt and S. Feuerriegel, "Early detection of user exits from clickstream data: A Markov modulated marked point process model," in *Proc. Web Conf.*, Apr. 2020, pp. 1671–1681.

[25] D. G. Winter, O. P. John, A. J. Stewart, E. C. Klohnen, and L. E. Duncan, "Traits and motives: Toward an integration of two traditions in personality research," *Psychol. Rev.*, vol. 105, no. 2, pp. 230–250, 1998.

[26] Y. Li and Y. Peng, "What drives gift-giving intention in live streaming? The perspectives of emotional attachment and flow experience," *Int. J. Hum.-Comput. Interact.*, vol. 37, no. 14, pp. 1317–1329, Aug. 2021.

[27] A. Berhenke, A. L. Miller, E. Brown, R. Seifer, and S. Dickstein, "Observed emotional and behavioral indicators of motivation predict school readiness in head start graduates," *Early Childhood Res. Quart.*, vol. 26, no. 4, pp. 430–441, Oct. 2011.

[28] F. Bolle, Y. Breitmoser, J. Heimel, and C. Vogel, "Multiple motives of pro-social behavior: Evidence from the solidarity game," *Theory Decis.*, vol. 72, no. 3, pp. 303–321, Mar. 2012.

[29] P. F. Stoeckart, M. Strick, E. Bijleveld, and H. Aarts, "The implicit power motive predicts action selection," *Psychol. Res.*, vol. 81, no. 3, pp. 560–570, May 2017.

[30] E. M. Fodor, O. Schultheiss, and J. Brunstein, "Power motivation," in *Implicit Motives*, O. C. Schultheiss and J. C. Brunstein, Eds. New York, NY, USA: Oxford Univ. Press, 2010, pp. 3–29.

[31] I. H. Frieze and B. S. Boneva, "Power motivation and motivation to help others," in *The Use and Abuse of Power*. London, U.K.: Psychol. Press, 2015, pp. 75–89.

[32] R. Selten and A. Ockenfels, "An experimental solidarity game," *J. Econ. Behav. Org.*, vol. 34, no. 4, pp. 517–539, Mar. 1998.

[33] R. E. Kranton, "Reciprocal exchange: A self-sustaining system," *Amer. Econ. Rev.*, vol. 20, pp. 830–851, Sep. 1996.

[34] T. Yamagishi and T. Kiyonari, "The group as the container of generalized reciprocity," *Social Psychol. Quart.*, vol. 63, no. 2, p. 116, Jun. 2000.

[35] J. Chen and T. Igarashi, "Unequal but not separate: Emergence of rich–poor cooperation in resource exchange," *Asian J. Social Psychol.*, vol. 26, no. 4, pp. 431–444, Dec. 2023.

[36] J. Kim, S. T. Allison, D. Eylon, G. R. Goethals, M. J. Markus, S. M. Hindle, and H. A. McGuire, "Rooting for (and then Abandoning) the underdog," *J. Appl. Social Psychol.*, vol. 38, no. 10, pp. 2550–2573, Oct. 2008.

[37] H. Tajfel and J. C. Turner, "An integrative theory of intergroup conflict," in *Psychology of Intergroup Relations*, W. G. Austin and S. Worchel, Eds. Chicago, IL, USA: Nelson-Hall, 1979, pp. 56–65.

[38] J. Nadler and M.-H. McDonnell, "Moral character, motive, and the psychology of blame," *Cornell L. Rev.*, vol. 97, p. 255, Jan. 2011.

[39] S. Greiff, S. Wüstenberg, and F. Avvisati, "Computer-generated log-file analyses as a window into students' minds? A showcase study based on the PISA 2012 assessment of problem solving," *Comput. Educ.*, vol. 91, pp. 92–105, Dec. 2015.

[40] K. Igwe, M. Olusanya, and A. Adewumi, "On the performance of GRASP and dynamic programming for the blood assignment problem," in *Proc. IEEE Global Humanitarian Technol. Conf. (GHTC)*, Oct. 2013, pp. 221–225.

[41] K. Igwe and N. Pillay, "Automatic programming using genetic programming," in *Proc. 3rd World Congr. Inf. Commun. Technol. (WICT)*, Dec. 2013, pp. 337–342.

[42] M. W. Gardner and S. R. Dorling, "Artificial neural networks (the multilayer perceptron)—A review of applications in the atmospheric sciences," *Atmos. Environ.*, vol. 32, nos. 14–15, pp. 2627–2636, Aug. 1998.

[43] S. Haykin, *Neural Networks: A Comprehensive Foundation*. Upper Saddle River, NJ, USA: Prentice-Hall, 1998.

[44] H. Zhang and R. Zhou, "The analysis and optimization of decision tree based on ID3 algorithm," in *Proc. 9th Int. Conf. Modeling, Identificat. Control (ICMIC)*, Jul. 2017, pp. 924–928.

[45] H. Yuliansyah, R. A. P. Imaniati, A. Wirasto, and M. Wibowo, "Predicting students graduate on time using C4. 5 algorithm," *J. Inf. Syst. Eng. Bus. Intell.*, vol. 7, no. 1, pp. 67–73, 2021.

[46] M. M. Rahman, A. Mohaimenul Islam, J. Miah, S. Ahmad, and M. Mamun, "SleepWell: Stress level prediction through sleep data. Are you stressed?" in *Proc. IEEE World AI IoT Congr. (AIIoT)*, Jun. 2023, pp. 0229–0235.

[47] P. Cunningham and S. J. Delany, "K-nearest neighbour classifiers—A tutorial," *ACM Comput. Surv.*, vol. 54, no. 6, pp. 1–25, Jul. 2022.

[48] H. Wang, G. Li, and Z. Wang, "Fast SVM classifier for large-scale classification problems," *Inf. Sci.*, vol. 642, Sep. 2023, Art. no. 119136.

[49] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, and M. Blondel, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12 no. 10, pp. 2825–2830, 2012.

[50] A. Tharwat, "Classification assessment methods," *Appl. Comput. Informat.*, vol. 17, no. 1, pp. 168–192, Jan. 2021.

[51] R. A. Hamad, L. Yang, W. L. Woo, and B. Wei, "Joint learning of temporal models to handle imbalanced data for human activity recognition," *Appl. Sci.*, vol. 10, no. 15, p. 5293, Jul. 2020.

[52] B. G. Marcot and A. M. Hanea, "What is an optimal value of k in k-fold cross-validation in discrete Bayesian network analysis?" *Comput. Statist.*, vol. 36, no. 3, pp. 2009–2031, Sep. 2021.

[53] R. Blagus and L. Lusa, "Joint use of over- and under-sampling techniques and cross-validation for the development and assessment of prediction models," *BMC Bioinf.*, vol. 16, no. 1, pp. 1–10, Dec. 2015.

[54] J. F. Zaki, A. Ali-Eldin, S. E. Hussein, S. F. Saraya, and F. F. Areed, "Traffic congestion prediction based on hidden Markov models and contrast measure," *Ain Shams Eng. J.*, vol. 11, no. 3, pp. 535–551, Sep. 2020.

[55] Team R Core. (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. [Online]. Available: https://www.R-project.org

[56] J. C. Gower, "A comparison of some methods of cluster analysis," *Biometrics*, vol. 23, no. 4, p. 623, Dec. 1967.

[57] G. Gan, C. Ma, and J. Wu, *Data Clustering: Theory, Algorithms, and Applications*. Philadelphia, PA, USA: SIAM, 2020.

[58] Ö. Akay and G. Yüksel, "Clustering the mixed panel dataset using Gower's distance and k-prototypes algorithms," *Commun. Statist.-Simul. Comput.*, vol. 47, no. 10, pp. 3031–3041, Nov. 2018.

[59] D.-T. Dinh, T. Fujinami, and V.-N. Huynh, "Estimating the optimal number of clusters in categorical data clustering by silhouette coefficient," in *Proc. Int. Symp. Knowl. Syst. Sci.* Cham, Switzerland: Springer, 2019, pp. 1–17.

[60] M. Wattenberg, F. Viégas, and I. Johnson, "How to use t-SNE effectively," *Distill*, vol. 1, no. 10, p. e2, Oct. 2016.

[61] B. Mor, S. Garhwal, and A. Kumar, "A systematic review of hidden Markov models and their applications," *Arch. Comput. Methods Eng.*, vol. 28, no. 3, pp. 1–12, 2021.

[62] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, no. 2, pp. 257–286, Jan. 1989.

[63] R. Batuwita and V. Palade, "Class imbalance learning methods for support vector machines," in *Imbalanced Learning: Foundations, Algorithms, and Applications*. Wiley, 2013, pp. 83–99.

[64] R. van den Goorbergh, M. van Smeden, D. Timmerman, and B. Van Calster, "The harm of class imbalance corrections for risk prediction models: Illustration and simulation using logistic regression," *J. Amer. Med. Inform. Assoc.*, vol. 29, no. 9, pp. 1525–1534, Aug. 2022.

[65] D. Balliet, J. Wu, and C. K. W. De Dreu, "Ingroup favoritism in cooperation: A meta-analysis," *Psychol. Bull.*, vol. 140, no. 6, pp. 1556–1581, 2014.

[66] T. Yamagishi, N. Jin, and T. Kiyonari, "Bounded generalized reciprocity: Ingroup boasting and ingroup favoritism," *Adv. Group Processes*, vol. 16, no. 1, pp. 161–197, 1999.

[67] A. Althnian, D. AlSaeed, H. Al-Baity, A. Samha, A. B. Dris, N. Alzakari, A. A. Elwafa, and H. Kurdi, "Impact of dataset size on classification performance: An empirical evaluation in the medical domain," *Appl. Sci.*, vol. 11, no. 2, p. 796, Jan. 2021.

[68] D. L. Paulhus and S. Vazire, "The self-report method," in *Handbook of Research Methods in Personality Psychology*, vol. 1. New York, NY, USA: Guilford Press, 2007, pp. 224–239.

KEVIN IGWE received the B.Sc. (Hons.) and M.Sc. degrees in computer science and the Ph.D. degree in psychology from the University of KwaZulu-Natal, South Africa, in 2013, 2016, and 2022, respectively.

He is currently a Postdoctoral Research Fellow (PDRF) with the Department of Psychology, Faculty of Humanities, University of Johannesburg, Johannesburg, South Africa. His research interests include artificial intelligence (AI) and machine learning (ML) to promote prosocial behavior and mental health.

KEVIN DURRHEIM is currently a Distinguished Professor in psychology with the Faculty of Humanities, University of Johannesburg, Johannesburg, South Africa, where he heads the UJ Methods Laboratory, promoting open access to open science in South Africa. He has broad interests in the field of social psychology of intergroup relations, and a program of research related to interaction dynamics and social change.