https://doi.org/10.48047/AFJBS.6.14.2024.12720-12743

AFJBS African Journal of Biological Sciences

Journal homepage: http://www.afjbs.com

Research Paper Open Access

A Combined Approach of Feature Selection and Machine Learning Techniques for the Real-Time Detection of Interpersonal Trust Issues

Arnab Chakraborty, Avijit Kumar Chaudhuri, Payel Sengupta, Sayak Das, Sulekha Das

mailforarnab@gmail.com , https://orcid.org/0009-0006-4561-4497, M.Tech. – Student, Computer Science & Engineering, Brainware University, Barasat, Kolkata – 125, West Bengal, India

c.avijit@gmail.com, Associate Professor Computer Science & Engineering, Brainware University, Barasat, Kolkata – 125, West Bengal, India

Payel9433@gmail.com, Assistant Professor, Computer Science & Engineering, Brainware University, Barasat, Kolkata – 125, West Bengal, India

sayakd006@gmail.com, https://orcid.org/0009-0003-4529-5391 ,M.Tech. – Student, Computer Science & Engineering, Brainware University, Barasat, Kolkata – 125, West Bengal, India

shu7773sea@gmail.com, Research Scholar, Computer Science and Engineering, GCECT, Kolkata, West Bengal, India

Volume 6, Issue 14, Sep 2024

Received: 15 July 2024

Accepted: 25 Aug 2024

Published: 05 Sep 2024

doi: 10.48047/AFJBS.6.14.2024.12720-12743

Abstract

Trust is the ability to have confidence in someone or depend on his/her word, person, or team. Thus, if any personality is labeled as having trust problems, then such a personality is seen as having dysfunctional trust for people and experiencing danger in interacting with others. The COVID-19 pandemic created tension, stress, and threats that affected interpersonal beliefs negatively. The conflict over interpersonal trust causes disturbances in individual and societal welfare. Hence, the only way is to work towards solving the trust issue to restore trust. In this article, the authors presented an approach that involved the use of a machine learning model in line with a real-life dataset to determine cases of problems relating to trust issues swiftly and accurately. In this research, an attempt has been made to construct a model with less number or features and this in turn reduces the computational requirement. To build the proposed model a four-stage method that entails data collection through face-to-face interviews, feature selection, various classification algorithms, and the comparison of the algorithms' performance was used. The dataset contains 91 independent variables and one dependent variable in the study. To perform feature selection several algorithms such as Information Gain, One-R, and Relief-F were used. These algorithms remove features from the lowest ranked of the dataset to higher ranked features in a recursive manner. Subsequently, the reduced dataset for training and testing of the classification algorithm was applied. The classification algorithms applied are Random Forest (RF), Random Tree (RT), Logistic Regression (LR), Multilayer Perceptron (MLP), and Support Vector Machine (SVM). The RF algorithm, with five features, gives 100% accuracy. To evaluate the performance, the authors employ several train-test split approaches and 10-fold cross-validation. The authors apply different statistical measures to assess the effectiveness of the classification algorithms. Last but not least; the suggested model has 100% accuracy in all classification matrices on a reduced feature set.

Index Terms: Trust Issues (TI), COVID-19, Information Gain (IG), One-R(OR), Random Forest (RF), Multilayer Perceptron (MLP), Post-Traumatic Stress Disorder (PTSD).

Introduction

Trusting is the act of believing in the integrity of another person. Trust issues can define as insufficient belief in people's integrity and having trouble feeling safe around others. Trust develops to be at least part of a heritable property, while mistrust seems to be an acquired behavioural response influenced by childhood nurturing and environment. Research in social psychology points to attachment style as having a strong influence on how trusting people are. For instance, people with an anxious attachment style seem to show less trust and more cognitive and behavioural jealousy in romantic relationships. Another factor is social isolation, which impacts the interpersonal trust issue, making it more difficult for people to trust others and form strong relationships. Due to Covid 19 pandemic, we all have faced a long social isolation period. Reduced face-to-face interaction, increased anxiety and stress, and changes in social norms, such as maintaining physical distance from others, form interpersonal Trust Issues (TI). Having trust issues isn't a standalone mental illness, but it may indicate potential symptoms, including attachment disorders, depression, anxiety, post-traumatic stress disorder (PTSD), and personality disorders. Generalized trust, the belief that most other people are trusted, is a thought that positively influences individuals and society in various ways (Dinesen, 2012). better physical and mental health, increased cooperation, well-being, and satisfaction with life highly attached to trust.

Inter-personal trust issues can manifest in any relationship, from romantic partners to work colleagues. Some may have suffered trust issues in every close relationship, while others only have trouble in one area. Trust issues may present in different contexts, such as romantic relationships, friendships, family members, and co-workers. The following table summarizes the signs of trust issues in different contexts:

Relationship types	Signs of trust issues
Romantic relationships	Checking social media or spying Hypersensitivity Avoiding commitment
Friendships	Over-focusing on the negative Not having friendships Lack of confidence and people-pleasing
Family members	Secret tracking apps Controlling behaviour Always being suspicious
Co-workers	Difficulty delegating Self-sabotaging Catastrophic thinking

Table 1: The signs of trust issues in different contexts

In this paper, the authors propose a machine learning model that detects the COVID-19 pandemic has impacted interpersonal trust in the general population and investigates whether trust is affected by pandemic-related strain, worries, and risk perception. To build the model, the authors follow a four-step way – data collection, applying features selection, using different classification algorithms, and comparing the performance of various algorithms. The authors try to build a model with fewer features, which leads to decreased computational complexity. This paper addresses the critical questions – (1) What are the crucial features that explain the cause of interpersonal trust issues? and (2) Which data mining tool yields higher accuracy?

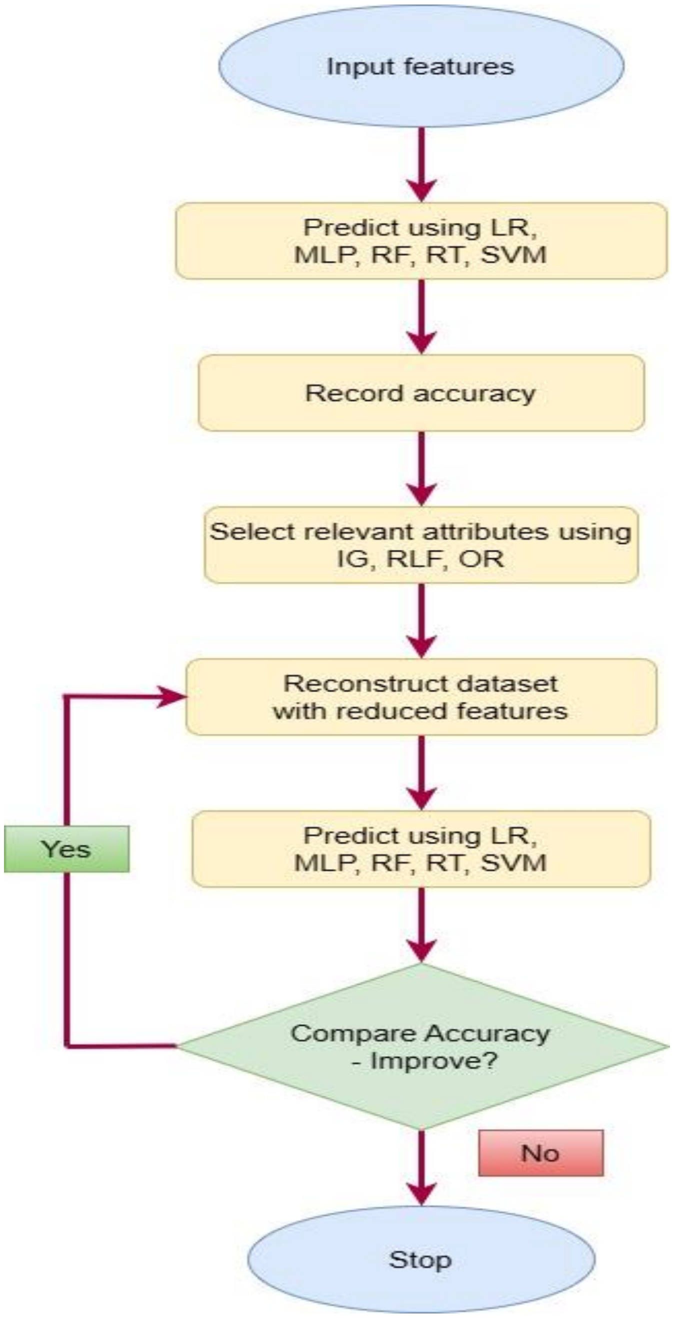

Figure 1: Multi-Stage Combining Feature Selection with Machine Learning Approach

The authors conducted face-to-face interviews with 207 people of different age groups based on questionnaires. The data set records 91 features of a candidate. The authors apply three different feature selection algorithms - Information-gain (IG), Relief-F(RLF), and One-R(OR), to reduce the unimportant features in three ways. It is a significant pre-processing function for classification, which helps to achieve higher accuracy of classification algorithm. The purpose of the Feature Selection is to rank the relevant features. A small subset of features reduces the computational complexity, which allows for faster classification and helps the human expert to concentrate on an appropriate subset of features. The authors used different well-defined Machine Learning Techniques (MLTs) like Support Vector Machines (SVM), Multilayer Perceptron (MLP), Logistic

Regression (LR), Random Forest (RF), Random Tree (RT) for predicting the Trust Issues (TI). The prediction procedure repeated for train-test ratios like 50%-50%, 66%-34%, and 80%-20%. After getting the accuracy, the authors used different performance measure techniques to compare the results of ML algorithms with subsets of features of our data set.

The authors proposed three sets of crucial features with high prediction accuracy of trust issues problems. Some features belong to all feature sets. These sets of features are essential for accurately identifying trust issue-related problems. The authors also proposed a set of classification models, which provide 100% accuracy based on a minimal set of features. Figure 1 flowchart depicted the Multi-Stage Combining Feature Selection with Machine Learning Approach

Relevant literature

Jiang et al. show in their study that interpersonal trust problems and mental health symptoms are correlated. Most people who suffer from trust issues have a high chance of experiencing symptoms of depression, anxiety, and stress. For women, the interrelation between interpersonal trust and mental health is more significant than for men. In the case of younger adults, mental illness sometimes creates trust issues. People who suffer from chronic health diseases, go through trauma, or have less social support are led to generate interpersonal trust issues (Jiang et al., 2021). The inter-personal trust is a significant predictor of mental health (Zhang et al., 2022).

Liu et al. analyse the role of interpersonal trust in the connection between social support and mental wellness. Specifically, social assistance positively affects the mental state when people possess higher interpersonal trust (Liu et al., 2023). Aloneness and interpersonal trust issues were significantly associated with increased mental stress among migrant people during the COVID-19 pandemic (Wang et al., 2023).

Lee et al. provide evidence that self-confidence and interpersonal trust are vital factors to improve the mental health of college students. The findings imply that therapies aiming to increase self-confidence and interpersonal trust may help lower social anxiety among college students (Lee et al., 2021).

Schnettler et al. study the importance of early childhood experiences in shaping social trust and social isolation later in life. The findings suggest that interventions aimed at improving family relationships and promoting social trust in childhood may help to prevent social isolation in adulthood (Schnettler et al., 2020).

Researchers are interested in using Machine Learning Techniques for medical diagnosis. In particular, mental health-related treatment is a prominent area for machine learning research. Table 2 shows a comparative review of previous works related to Machine Learning in mental health treatment.

Table 2: A comparative review of previous works in application of MLTs in mental health prediction

Literature	Mental Health Prediction Using Machine Learning: Taxonomy, Applications, and Challenges (Chung et al., 2022)	Mental Disorder Detection: Bipolar Disorder Scrutinization Using Machine Learning (Kaur et al., 2020)	Predicting Mental Health Illness using Machine Learning Algorithms (Goddard et al., 2022)	Machine Learning Models based Mental Health Detection - IEEE Xplore (Muhammad et al., 2022)
Focus	Overview of machine learning in mental health prediction	Machine learning for bipolar disorder detection	Machine learning for predicting mental health illness	Machine learning framework for mental health disorder detection
Methods	Literature review	Support vector machine (SVM)	Logistic regression, SVM, decision trees	KNN, Naive Bayes, decision trees
Findings	Machine learning can be used to predict mental health disorders with a high degree of accuracy.	Machine learning can be used to detect bipolar disorder with an accuracy of 85%.	Machine learning algorithms can be used to predict mental health illness with high accuracy.	Machine learning framework can achieve an accuracy of up to 90% for detecting depression, anxiety, and stress.

As you can see, the papers all focus on using machine learning to detect or predict mental health disorders. However, they use different methods and have separate findings. The paper by Sadiq et al. (2022) provides a comprehensive overview of the use of machine learning in mental health prediction, while the paper by Kumar and Gupta (2022) focuses on using machine learning to detect bipolar disorder. The paper by Khan and Hussain (2022) investigates the application of machine learning to predict mental health illness, while the paper by Alzahrani and Almansour (2021) proposes a framework for using machine learning to detect mental health disorders. The paper by Saha (2022) discusses the use of machine learning to predict mental health disorders.

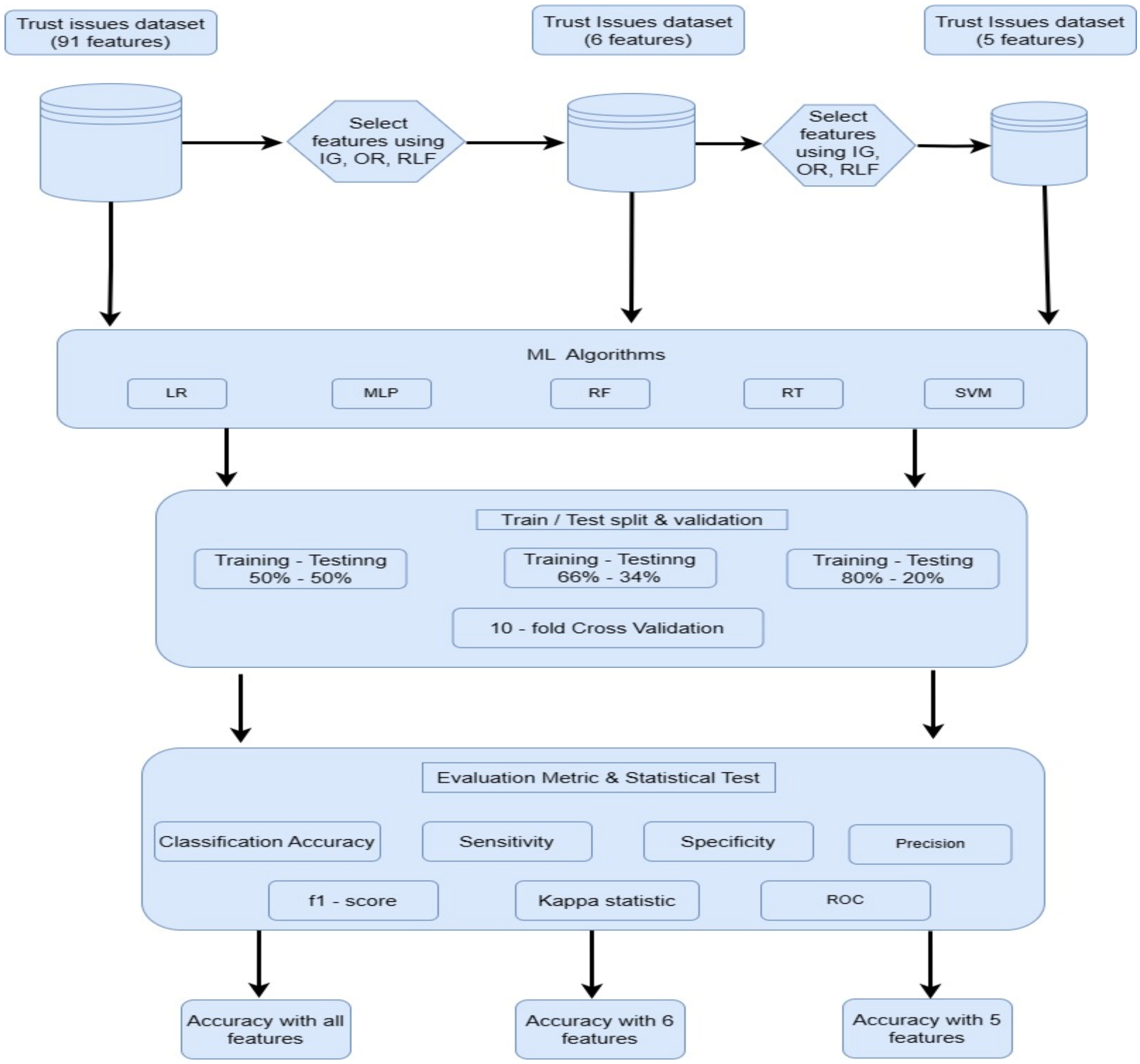

Figure 2: The Architecture for the Proposed System

Feature Selection Methods

Finding key info in a massive book? Two ways like choosing glasses: Wrapper method tries on different methods to pick the best "fit" for the task, like finding the juiciest parts of a fruit salad. Filter method skims each page fast, judging each piece on its own, like deciding if recipe ingredients are essential. In short, wrapper method meticulously chooses features, while filter method quickly assesses their individual importance. There are different ways to do this, like looking at information gain, relief-F, and one-R.

Information Gain

Information Gain, our data-sorting superhero, tackles messy facts! Imagine a mission: find the key details. This super-powered detective uses "entropy," a measurement unit of data chaos, like a villain detector. The smaller the chaos after comparing pre- and post-sorting "entropy," the hotter the clue! Information Gain picks the details that clean up the most, guiding us to the data's hidden gems, making us data heroes ourselves! Ready to unleash this sorting power on your dataset? Let's find the golden nuggets! (Chaudhuri et al. 2021).

Relief-F

Imagine Relief-F as your detective's AI sidekick. It picks random suspects, analyses clues for differences between groups, and flags the most distinct ones. Like a super-powered fingerprint scanner, it highlights the clues that crack the case, even with messy info. No more two-group limitations, Relief-F tackles diverse crowds with ease. It's your secret weapon for solving mysteries, no matter how complex (Ray et al., 2021).

One R

One R Ranker, your detective partner, tackles every clue (numbers included!). It crafts simple rules, groups them neatly, and flags missing pieces. By ranking clues based on rule flubs, it highlights the most helpful suspects. This oldie but goodie sets the standard for other detective tools, a trusty sidekick for comparing and boosting your mystery-solving skills.

Choice of data mining models

Our research dives deep into five popular machine learning methods, from Random Trees to SVMs. We dissect their strengths, weaknesses, and performance across diverse situations. Our goal? To unlock their full potential and make them work better, in more ways, for everyone. It's a journey towards smarter, more powerful AI.

Random forest

Imagine a group of mini-experts, each with a simple sorting rule. RF gathers them all, throws a bunch of data their way, and lets them vote on how to categorize it. The majority rules, winning with confidence based on how many mini-experts agree. This powerful team, built from diverse perspectives, excels at sorting things into the right categories (Chaudhuri et al., 2021).

Random tree

Imagine Sherlock Holmes leading a team of detectives, each sniffing out clues to solve a medical mystery. That's Random Trees, a super-smart way to spot diseases. These "detective trees" use diverse clues and flexible thinking, beating other methods in accuracy. Researchers agree: Random Trees are the Moriarty of smart computer guesses, solving cases other methods can't crack!.

Multilayer perceptron

Imagine MLP as a superhero in the world of machines learning. It's like a super-smart helper that can handle lots of different jobs and teach machines how to learn by themselves. MLP is really good at figuring out the tricky connections between what goes in and what comes out. Even when there's a ton of information to deal with, MLP can handle it, although it might take a bit longer when there's a whole bunch of data. It's like having a patient and trustworthy teacher for machines, making learning easy, especially when things get complicated (Das et al. 2024).

Logistic regression

Logistic regression, leveraging a sigmoid function, predicts binary outcomes (e.g., success/failure) based on diverse data factors. Employing statistical modelling, it quantifies relationships between variables and outcomes, enabling informed decision-making. Think of it as a data-driven tool for classifying and predicting with precision (Deb et al., 2023).

Support vector machine

Support Vector Machines (SVMs), conceived in the 1990s, are powerful machine learning algorithms adept at finding optimal decision boundaries, or hyperplanes, within data. This ability allows them to effectively separate distinct data classes, akin to classifying cats and dogs. Notably, SVMs excel at achieving a crucial balance: delivering accurate predictions while avoiding overfitting the data. Consequently, they have become invaluable tools in diverse fields, particularly aiding medical professionals in disease diagnosis and researchers in analysing conditions like cerebral palsy. In essence, SVMs function as versatile AI instruments for driving insightful predictions across various disciplines.

Performance metrics

To see how well the proposed work is doing, we use something called a Confusion Matrix. This special chart helps us check how good a learning model is. We look at four things to measure its performance:

True Positive (TP): This is the number of people with trust issues correctly identified as having trust issues.

False Positive (FP): It's the number of people without trust issues mistakenly labelled as having trust issues.

True Negative (TN): This is the number of people correctly identified as not having trust issues.

False Negative (FN): It's the number of people incorrectly labelled as having trust issues when they don't.

These terms help us see how well the model can tell the difference between people with trust issues and those without.

Accuracy: Accuracy is like checking how many things you got right out of all the things you guessed.

$$ACCURACY = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision: Precision is like figuring out how many of the people you said have trust issues really do, compared to everyone you said might have trust issues (Kar et all. 2024).

$$PRECESSION = \left(\frac{TP}{FP + TP} \right)$$

Recall or Sensitivity: Recall or sensitivity is like checking how many people with trust issues were found by a system, compared to all the people who actually have trust issues.

$$RECALL = \left(\frac{TP}{FN + TP} \right)$$

Specificity: Specificity is like checking how well a test can correctly tell if someone doesn't have trust issues. It looks at the number of people without trust issues compared to those the test correctly identified as not having trust issues (Das et al. 2024)

$$SPECIFICITY = \left(\frac{TN}{TN + FP} \right)$$

F1 score: The F1 Score helps to strike a balance between being accurate about identifying the right things and not missing out on important things. It's like finding a middle point and is calculated using a special method that considers both accuracy and completeness.

$$F1SCORE = \frac{2(PRECISION \times RECALL)}{PRECISION + RECALL}$$

AUC-ROC curve: The AUC-ROC Curve is like a scorecard for how well a model can distinguish between two things, such as figuring out if someone has trust issues or not. It shows how good the model is at telling them apart. The ROC part is about finding the right balance between correctly spotting when someone has trust issues (True Positives) and avoiding mistakenly thinking someone has trust issues when they don't (False Positives). The AUC number is like a grade, and the closer it is to 1, the better the model is at making this distinction (Das et al. 2024).

$$TPR = \frac{TP}{TP + FN}$$

$$FPR = \frac{FP}{FP + TN}$$

Kappa statistics: Kappa Statistics is like a tool to see how much two people agree when they're giving ratings or judgments. It helps us know if their agreement is more than what we might expect by random chance, giving us a better idea of how well they agree, not just by luck (Das et al. 2024).

$$K_{STAT} = \frac{A_{OBS} - A_{EXP}}{N - A_{EXP}}$$

Dataset description

The dataset presented has been manually compiled by observing the behavioral dynamics of individuals during the COVID-19 pandemic. It encompasses information related to 207 individuals, covering a wide range of attributes such as age, various symptoms, and medical background. In total, there are 90 features or attributes, with a specific target group identified as "Trust issues."

Due to privacy concerns, some individuals have chosen not to provide answers to certain attributes, resulting in missing values in the dataset. To address this, missing values have been replaced with mean values to ensure data integrity and avoid repetition.

Attribute No.	Attribute Name	Description	Total Data (Count)	Missing Data (Count)	Data Type

1	Age	Participant's age in years	207	0	Integer
2	Gender	Participant's gender (Male=1, Female=0)	207	0	Boolean
3	Teacher	Occupation as a teacher (Yes=1, No=0)	207	0	Boolean
4	Homemaker	Occupation as a homemaker (Yes=1, No=0)	207	0	Boolean
5	College Student	Enrolment as a college student (Yes=1, No=0)	207	0	Boolean
6	School Student	Enrolment as a school student (Yes=1, No=0)	207	0	Boolean
7	Business Person	Occupation as a business person (Yes=1, No=0)	207	0	Boolean
8	Unemployed	Unemployment status (Yes=1, No=0)	207	0	Boolean
9	Health Service Provider	Occupation in health services (Doctor, Nurse, Health staff) (Yes=1, No=0)	207	0	Boolean
10	Service Provider	Occupation as a service provider (Delivery boy, Freelancer, Private staff) (Yes=1, No=0)	207	0	Boolean
11	Other Occupations	Any other occupation not listed (Yes=1, No=0)	207	1	Boolean
12	Marital Status: Single	Marital status as single (Yes=1, No=0)	207	0	Boolean
13	Marital Status: Married	Marital status as married (Yes=1, No=0)	207	0	Boolean
14	Marital Status: Divorced	Marital status as divorced (Yes=1, No=0)	207	0	Boolean
15	Marital Status: Separated	Marital status as separated (Yes=1, No=0)	207	0	Boolean
16	Marital Status: Widowed	Marital status as widowed (Yes=1, No=0)	207	0	Boolean

17	Relationship Status: In a Relationship	Relationship status (Yes=1, No=0)	207	0	Boolean
18	Anxiety	Presence of anxiety (Yes=1, No=0)	207	0	Boolean
19	Depression	Presence of depression (Yes=1, No=0)	207	0	Boolean
20	PTSD	Presence of post-traumatic stress disorder (Yes=1, No=0)	207	0	Boolean
21	Eating Disorders	Presence of eating disorders (Anorexia, Bulimia, Binge eating, etc.) (Yes=1, No=0)	207	0	Boolean
22	Neurodevelopmental Disorders	Presence of neurodevelopmental disorders (ADHD, Autism, Learning disabilities, Intellectual disabilities, etc.) (Yes=1, No=0)	207	0	Boolean
23	Other Problems	Presence of other unspecified problems (Yes=1, No=0)	207	0	Boolean
24	Past Trust Issues	History of trust issues (Yes=1, No=0)	207	3	Boolean
25	Current Trust Issues	Current experience of trust issues (Yes=1, No=0)	207	0	Boolean
26	Frequency: Once	Trust issues occurred only once (Yes=1, No=0)	207	0	Boolean
27	Frequency: Frequently	Trust issues occur frequently (Yes=1, No=0)	207	0	Boolean
28	Duration: <1 Month	Trust issues duration less than 1 month (Yes=1, No=0)	207	0	Boolean
29	Duration: 1-6 Months	Trust issues duration between 1 and 6 months (Yes=1, No=0)	207	0	Boolean
30	Duration: 6-12 Months	Trust issues duration between 6 months and 1 year (Yes=1, No=0)	207	0	Boolean

31	Duration: >1 Year	Trust issues duration more than 1 year (Yes=1, No=0)	207	0	Boolean
32	Insecurity	The feeling of insecurity (Yes=1, No=0)	207	0	Boolean
33	Irritability	Experience of irritability (Yes=1, No=0)	207	0	Boolean
34	Loneliness	Experience of loneliness (Yes=1, No=0)	207	0	Boolean
35	Confusion	Experience of confusion (Yes=1, No=0)	207	0	Boolean
36	Hallucination	Experience of hallucination (Yes=1, No=0)	207	0	Boolean
37	Self-Harm Thoughts	Thoughts of self-harm or suicide (Yes=1, No=0)	207	0	Boolean
38	Aggression	Experience of aggression (Yes=1, No=0)	207	0	Boolean
39	Nervousness	Experience of nervousness or excessive sweating (Yes=1, No=0)	207	0	Boolean
40	Lack of Concentration	Difficulty in concentration or lack of confidence (Yes=1, No=0)	207	0	Boolean
41	Overdoing	Engaging in activities more than needed (Yes=1, No=0)	207	0	Boolean
42	Breathing Difficulty	Difficulty in breathing or shortness of breath (Yes=1, No=0)	207	0	Boolean
43	Fatigue	Experience of fatigue (Yes=1, No=0)	207	1	Boolean
44	Insomnia	Experience of insomnia (Yes=1, No=0)	207	0	Boolean
45	Restlessness	Experience of restlessness (Yes=1, No=0)	207	1	Boolean
46	Loss of Appetite	Loss of appetite (Yes=1, No=0)	207	0	Boolean

47	Intrusive Thoughts	Experience of intrusive thoughts (Yes=1, No=0)	207	0	Boolean
48	Guilt	Experience of guilt without reason (Yes=1, No=0)	207	0	Boolean
49	Sudden Weight Change	Sudden weight gain or loss (Yes=1, No=0)	207	0	Boolean
50	Excessive Behaviours	Excessive crying, eating, or sleeping (Yes=1, No=0)	207	0	Boolean
51	Hopelessness	Feeling of hopelessness or loss of interest in activities (Yes=1, No=0)	207	0	Boolean
52	Other Symptoms	Presence of other symptoms (Yes=1, No=0)	207	3	Boolean
53	Substance Abuse	Drug and alcohol misuse (Yes=1, No=0)	207	0	Boolean
54	Family Pressure	Experience of excessive family pressure (Yes=1, No=0)	207	1	Boolean
55	Long-term Stress	Experience of severe or long-term stress (Yes=1, No=0)	207	1	Boolean
56	Social Isolation	Experience of social isolation or loneliness (Yes=1, No=0)	207	2	Boolean
57	Poor Housing	Experience of homelessness or poor housing (Yes=1, No=0)	207	0	Boolean
58	Job Loss	Experience of unemployment or losing a job (Yes=1, No=0)	207	0	Boolean
59	Victim of Crime	Being the victim of a violent crime (Yes=1, No=0)	207	0	Boolean
60	Study Stress	Excessive mental stress from studies (Yes=1, No=0)	207	3	Boolean
61	Childhood Trauma	History of childhood abuse, trauma, or neglect (Yes=1, No=0)	207	0	Boolean

62	Social Disadvantage	Experience of social disadvantage, poverty, or debt (Yes=1, No=0)	207	0	Boolean
63	Long-term Caregiver	Being a long-term carer for someone (Yes=1, No=0)	207	0	Boolean
64	Self-obsession Insecurity	Insecurity due to excessive self-obsession (Yes=1, No=0)	207	0	Boolean
65	Bereavement	Experience of bereavement (losing someone close) (Yes=1, No=0)	207	0	Boolean
66	Chronic Health Condition	Having a long-term physical health condition (Yes=1, No=0)	207	0	Boolean
67	Domestic Violence	Experience of domestic violence, bullying, or other abuse as an adult (Yes=1, No=0)	207	0	Boolean
68	Discrimination	Experience of discrimination and stigma, including racism (Yes=1, No=0)	207	0	Boolean
69	Serious Incident	Being involved in a serious incident where life was feared for (Yes=1, No=0)	207	0	Boolean
70	Physical Causes	Presence of physical causes like a head injury or a neurological condition (Yes=1, No=0)	207	0	Boolean
71	Other Causes	Presence of other unspecified causes (Yes=1, No=0)	207	14	Boolean
72	Visited Professional	Has the participant ever visited a professional regarding mental health? (Yes=1, No=0)	207	0	Boolean

73	Problem Diagnosed	Is the problem diagnosed by a professional? (Yes=1, No=0)	207	0	Boolean
74	Problem Resolved	Is the problem resolved after professional consultation? (Yes=1, No=0)	207	0	Boolean
75	Ongoing Treatment	Is the participant still in treatment? (Yes=1, No=0)	207	6	Boolean
76	Treatment Duration: <1 Month	Treatment duration less than 1 month (Yes=1, No=0)	207	0	Boolean
77	Treatment Duration: 1-6 Months	Treatment duration between 1 and 6 months (Yes=1, No=0)	207	0	Boolean
78	Treatment Duration: 6-12 Months	Treatment duration between 6 months and 1 year (Yes=1, No=0)	207	0	Boolean
79	Treatment Duration: >1 Year	Treatment duration more than 1 year (Yes=1, No=0)	207	0	Boolean
80	Problem Fixed by Professional	Problem fixed by consulting with a doctor, psychiatrist, or therapist (Yes=1, No=0)	207	0	Boolean
81	Problem Fixed by Friends/Relatives	Problem resolved with the help of friends or relatives (Yes=1, No=0)	207	0	Boolean
82	Problem Fixed by Self-Motivation	Problem fixed by self-realization or self-motivation (Yes=1, No=0)	207	0	Boolean
83	Family Mental Disorder History	History of mental disorder in the family (Yes=1, No=0)	207	0	Boolean
84	Grandfather's Mental Health	Did the participant's grandfather have any kind of mental illness? (Yes=1, No=0)	207	0	Boolean
85	Grandmother's Mental Health	Did the participant's grandmother have any kind of mental illness? (Yes=1, No=0)	207	0	Boolean

86	Father's Mental Health	Did the participant's father have any kind of mental illness? (Yes=1, No=0)	207	0	Boolean
87	Mother's Mental Health	Did the participant's mother have any kind of mental illness? (Yes=1, No=0)	207	0	Boolean
88	Brother's Mental Health	Did the participant's brother have any kind of mental illness? (Yes=1, No=0)	207	0	Boolean
89	Sister's Mental Health	Did the participant's sister have any kind of mental illness? (Yes=1, No=0)	207	0	Boolean
90	Other Family Members' Mental Health	Did any other family member have any kind of mental illness? (Yes=1, No=0)	207	3	Boolean

Table 3: Data Description

Results and discussion

Based on the TI dataset, the authors predict the trust issue–related problem through a comparative study between the results of various MLTs and a combination of different FS methods with MLTs. Information Gain, Relief-F, and One-R are feature selection methods used in this study and use advanced ML algorithms like LR, MLP, SVM, RF, and RT for the classification. The 10-fold Cross-validation and Split percentages, including 50%-50%, 66%-34%, and 80%-20%, are used to test the classification models.

To justify the outcomes of blending techniques FS and MLTs, the authors apply various performance metrics – accuracy, Specificity, Sensitivity, Precision, and statistical measures like Kappa statistics and ROC.

MLP shows the highest accuracy rate, 72% with all features, under a 10-fold cross-validation test, and other MLTs like LR, SVM, RF, and RT with all features and for 10-fold-cross-validation test show accuracy results of 63%, 64%, 71%, and 54% respectively. The authors used feature selection methods to increase the accuracy rate with a minimum number of attributes. Information Gain, Relief F, and One R feature selection methods provide three different types of rankings of attributes. The authors recursively reduce less-important features and apply MLTs to identify the most significant features.

The authors found that all three sets of six features and five features provide above 95% accuracy. RF and RT show 99% and 100% accuracy with six and five features for all ranked sets, respectively.

Classification algorithm with information gain features selection

The authors use the Information Gain method to rank the attributes. The authors found that all used classification models give the best result with six and five attributes set. The attribute "Bereavement (losing someone close to you)" has no significant contribution to the results. So, the authors exclude the attribute from the data set.

Attributes
Other occupations, Other problems, Difficulty in breathing or shortness of breath, Homelessness or poor housing, Bereavement (losing someone close to you), More than 6 months and less than a year in treatment, Trust issue

Table 4: Trust Issue Dataset with 6 Features and 1 Target Variable using Information Gain Features Selection

Attributes
Other occupations, Other problems, Difficulty in breathing or shortness of breath, Homelessness or poor housing, More than 6 months and less than a year in treatment, Trust Issues

Table 5: Trust Issue Dataset with 5 Features and 1 Target Variable using Information Gain Features Selection

Performance Metrics	91 Features		6 Features		5 Features	
	ML Techniques	Maximum Score	ML Techniques	Maximum Score	ML Techniques	Maximum Score
Accuracy	MLP	0.73	RF, RT	1.00	MLP, RF, RT	1.00
Sensitivity	MLP	0.80	LR, MLP, SVM, RF, RT	1.00	LR, MLP, SVM, RF, RT	1.00
Precision	RF	1.00	LR, MLP, SVM, RF, RT	1.00	MLP, RF, RT	1.00
Specificity	RF	1.00	LR, MLP, SVM, RF, RT	1.00	MLP, RF, RT	1.00
f1-score	RF	0.83	LR, MLP, SVM, RF, RT	1.00	MLP, RF, RT	1.00
ROC	RF	0.78	LR, MLP, SVM, RF, RT	1.00	MLP, RF, RT	1.00
Kappa	MLP	0.39	LR, MLP, SVM, RF, RT	1.00	MLP, RF, RT	1.00

Table 6: Improvement in Performance due to Information Gain FS from 91 to 5

Classification algorithm with One R feature selection

One R FS method generates a different rank of all attributes with respect to the target variable. The authors found that the top six and five features show the best results with all classification models. 100% accuracy gain from RF and RT classifier with five attributes.

Table 15 shows the accuracy rate with all features and six and five features. The six and five feature datasets are different from the Information Gain datasets. Under the 10-fold cross-validation test, with these six and five feature datasets, all classification algorithm provides almost 100% accuracy. The following tables display different performance matrices under the One R features selection method.

Attributes
Facing these problems less than 1 month, Facing these problems more than 1 month and less than 6 months, Facing these problems more than 6 months and less than 1 year, Facing these problems more than a year, Insecurity, Other family members who had mental illness, Trust Issues

Table 7: Trust Issue Dataset with 6 Features and 1 Target Variable using One R Features Selection

Attributes
Facing these problems more than 1 month and less than 6 months, Facing these problems more than 6 months and less than 1 year, Facing these problems more than a year, Insecurity Other family members who had mental illness, Trust Issues

Table 8: Trust Issue Dataset with 5 Features and 1 Target Variable using One R Features Selection

Performance Metrics	91 Features		6 Features		5 Features	
	ML Techniques	Maximum Score	ML Techniques	Maximum Score	ML Techniques	Maximum Score
Accuracy	MLP	0.73	LR, MLP, SVM, RF, RT	1.00	MLP, RF, RT, SVM	1.00
Sensitivity	MLP	0.77	LR, MLP, SVM, RF, RT	1.00	MLP, RF, RT	1.00
Precision	RF	1.00	LR, MLP, SVM, RF, RT	1.00	MLP, RF, RT	1.00
Specificity	RF	1.00	LR, MLP, SVM, RF, RT	1.00	MLP, RF, RT	1.00
f1-score	RF	0.83	LR, MLP, SVM, RF, RT	1.00	MLP, RF, RT	1.00
ROC	RF	0.88	LR, MLP, SVM, RF, RT	1.00	MLP, RF, RT	1.00
Kappa	MLP	0.39	LR, MLP, SVM, RF, RT	1.00	MLP, RF, RT	1.00

Table 9: Improvement in Performance due to One-R FS from 91 to 5

Classification algorithm with Relief-F feature selection

The relief-F FS method generates a different ranked list of attributes to the target variable. The authors found that the top six and five features show the best results with all classification models. 100% accuracy gain from RF and RT classifiers with six and five attributes.

Attributes
Other occupations, Other problems, Facing these problems more than 1 month and less than 6 months, Insomnia, Homelessness or poor housing, Do your father had any kind of mental illness ?, Trust Issues

Table 10: Trust Issue Dataset with 6 Features and 1 Target Variable using Relief F Features Selection

Attributes
Other occupations, Others problems, Facing these problems more than 1 month and less than 6 months, Homelessness or poor housing, Do your father had any kind of mental illness?, Trust Issues

Table 11: Trust Issue Dataset with 5 Features and 1 Target Variable using Relief F Features Selection

Performance Metrics	91 Features		6 Features		5 Features	
	ML Techniques	Maximum Score	ML Techniques	Maximum Score	ML Techniques	Maximum Score
Accuracy	MLP	0.73	LR, MLP, SVM, RF, RT	1.00	MLP, RF, RT, SVM	1.00
Sensitivity	MLP	0.77	LR, MLP, SVM, RF, RT	1.00	MLP, RF, RT	1.00
Precision	RF	1.00	LR, MLP, SVM, RF, RT	1.00	MLP, RF, RT	1.00
Specificity	RF	1.00	LR, MLP, SVM, RF, RT	1.00	MLP, RF, RT	1.00
f1-score	RF	0.83	LR, MLP, SVM, RF, RT	1.00	MLP, RF, RT	1.00
ROC	RF	0.88	LR, MLP, SVM, RF, RT	1.00	MLP, RF, RT	1.00
Kappa	MLP	0.39	LR, MLP, SVM, RF, RT	1.00	MLP, RF, RT	1.00

Table 12: Improvement in Performance due to Relief-F FS from 91 to 5

Conclusion

The prediction accuracy of MLTs can be at various levels. Different factors affect the accuracy level, like the type and size of the dataset, the type of ML techniques (supervised and unsupervised), and the irrelevant features in the dataset. Careful feature selection leads to high-accuracy outcomes from ML techniques. So, a combination of feature selection and ML techniques makes a high level of accuracy.

In this study, the authors identify the significant factors which have maximum contribution to predicting trust issues. Results of the research show a high prediction accuracy using most MLTs with a subset of features. There are no single feature selection techniques that can preferred over others. So, this fact insists the authors apply multiple feature selection techniques. Different feature selection techniques provide multiple significant feature sets. In medical diagnosis, multiple- sets of factors help to improve the diagnosis process.

In the medical field, error play a pivotal role. False positive (FP) and False negative (FN) these two errors are equally important. So, not being diagnosed will lead to fatality, and treatments without disease will have adverse effects. The performance metrics help us to evaluate and justify the performance of MLTs. The authors obtained a 1.0 score for sensitivity, specificity, ROC, and Kappa statistics.

In this research, Random Forest and Random Tree delivered 100% accuracy with five features. SVM and MLP provide 92% accuracy with five features, and LR provides 81% accuracy. But with six features, all MLTs had delivered consistently above 95% accuracy. Thus, the authors conclude that RF and RT had delivered consistently better results with six and five feature sets than other MLTs.

The research deduces the answer to the question, what are the crucial factors that explain the cause of interpersonal trust issues? The answer is the reduced feature set. The reduced feature set helps medical professionals to understand which risk factors are associated with trust issue-related problems. The study also addresses another crucial question, "Which data mining tool yields high accuracy?". The authors proposed a combination of feature selection and machine learning techniques to predict trust issues, lead to a reduction in errors, and optimize the performance metrics.

References

Kaur, Amritpal, Kumar, Pardeep, Singh, Harinder and Singh, Manminder. (2022). Journal of Big Data, Mental Disorder Detection: Bipolar Disorder Scrutinization Using Machine Learning., 9(1).

Chaudhuri, A. K., Banerjee, D. K., & Das, A. (2021). International Journal of Intelligent Systems and Applications. A dataset centric feature selection and stacked model to detect breast cancer. 13(4), 24.

Chaudhuri, A. K., Sinha, D., Banerjee, D. K., & Das, A. (2021). Network Modeling Analysis in Health Informatics and Bioinformatics. A novel enhanced decision tree model for detecting chronic kidney disease. 10, 1-22.

Chung, J. and Teo, J. (2022). Applied Computational Intelligence and Soft Computing. Mental health prediction using machine learning: taxonomy, applications, and challenges. 2022(1), p.9970363.

Das, S., Chaudhuri, A. K., & Ghosh, P. (2024). Scope Journal. Leveraging Reinforcement Learning for Enhanced Cancer Detection: A Comprehensive Review. vol. 14(01). pp. 1634 – 1655.

Das, S., Chaudhuri, A. K., Ghosh, P., & Sikdar, S. (2024). Bioscene. A Novel CNN Model for Early Prediction of Brain Cancer. vol. 21(02). pp. 04 – 23.

Das, S., Chaudhuri, A. K., Ghosh, P., & Sikdar, S. (2024). African Journal of Biological Sciences. A Novel Robust CNN Model for MRI Brain Tumor Classification. Vol. 6(05). Pp. 7114 – 7138.

Das, S., Kar, S. P., Sil, S., Molla, A. R., Rajak, R., & Chaudhuri, A. K. (2024). In AI-Driven Innovations in Digital Healthcare: Emerging Trends, Challenges, and Applications IGI Global. A Multifaceted Approach to Understanding Mental Health Crises in the COVID-19 Era: Using AI Algorithms and Feature Selection Strategies. pp. 97-119.

Jiang, W, Ma, X, & Jia, J. (2021). Journal of Affective Disorders. Interpersonal trust and mental health: A systematic review and meta-analysis. pp. 337-350.

Kar, S. P., Molla, A. R., Das, S., Rajak, R., Sil, S., & Chaudhuri, A. K. (2024). In Medical Robotics and AI-Assisted Diagnostics for a High-Tech Healthcare Industry IGI Global. Identification of Insecurity in COVID-19 Using Machine Learning Techniques. pp. 239-256.

Lee, S. Y., & Jung, Y. J. (2021). International Journal of Environmental Research and Public Health. The Relationship Between Self-Esteem, Interpersonal Trust, and Social Anxiety of College Students, 18(20)(10904).

Liu, H., & He, Q. (2023). Journal of Interpersonal Violence. The role of interpersonal trust in the relationship between social support and mental health. 38(1-2)105-124.

E. Goddard, Michael, Hatcher, Simon J., and White, Martin. (2022). Journal of Mental Health. Predicting Mental Health Illness using Machine Learning Algorithms. 31(2).

Muhammad, Sajjad, Iqbal, Muhammad Mohsin, Umair, Muhammad, Aslam, Imran and Nabeel, Muhammad. (2022). IEEE Xplore. Machine Learning Models based Mental Health Detection.

Ray, A., & Chaudhuri, A. K. (2021). Machine Learning with Applications. Smart healthcare disease diagnosis and patient management: Innovation, improvement and skill development. 3, 100011.

Schnettler, J. E., Eng, P. T., & Harack, B. M. (2020). Personality and Social Psychology Bulletin. The Role of Interpersonal Trust in the Relationship Between Childhood Family Environment and Later Social Isolation.46-2, pp. 251-262.

Wang, W., Yang, F., & Zhang, M. (2022). Frontiers in Psychology. Loneliness and Trust Issues Reshape Mental Stress of Expatriates During Early COVID-19: A Structural Equation Modelling Approach. 13(794217).

Zhang, M, Zhang, W, & Liu, Y. (2022). Frontiers in Psychology. The relationship between interpersonal trust and mental health in Chinese adolescents: A cross-sectional study. vol. 13-2022(829307).

Deb, A, Samadder, B, Chowdhury, S, Das, S, Banarjee, S (2023). International Journal of Engineering Technology and Management Sciences. Measuring Mental Health Condition using Logistic Regression. 7(2), pp. 327–338. https://doi.org/10.46647/ijetms.2023.v07i02.040.