Predictive Modeling and Pattern Mining for Mental Health Disorders Using Data Mining Techniques

Samruddhi Kamble
IIT Madras
Chennai, India
kamblesamruddhi85@gmail.com

Carson K. Leung
University of Manitoba
Winnipeg, MB, Canada
carson.leung@umanitoba.ca

Rahul Saxena
IIT Madras
Chennai, India
23f2004309@ds.study.iitm.ac.in

Abstract

Mental health issues impact a large portion of the global population, emphasizing the demand for analytical solutions that are both scalable and interpretable. The study utilizes a comprehensive dataset with over 290K+ records and 17 variables, capturing demographic, clinical, and attitudinal information related to mental health. This extensive data enables robust analysis of mental illness prevalence, burden, and societal perspectives. Identifying patterns and improving classification in large mental health data is challenging. This study uses clustering, frequent pattern mining, and predictive modeling to address this. Dimensionality reduction and unsupervised clustering via PCA and K-Means enable behavioral segmentation. The study mines high-quality association rules using vectorized computations and adaptive thresholds to generate L1–L3 itemsets. L1 itemsets are frequent individual items meeting a minimum support threshold (e.g., Treatment = Yes). L2 itemsets are frequent pairs of co-occurring items (e.g., Mood Swings = Yes and Family History = Yes), while L3 itemsets are frequent triplets that appear together (e.g., mood swings, family history, social withdrawal). These itemsets reveal important patterns and associations in the data. These patterns are used to enrich interpretable decision tree models for predicting coping difficulties. Results demonstrate improved model performance and clarity, supporting early mental health intervention strategies. What distinguishes this work is its intention to create meaningful tools for early mental health intervention. As someone who has personally navigated mental health challenges, I see data science as a bridge between silence and support a way to turn patterns into signals, and signals into action. This research is a step toward that vision: human-centered, technically rigorous and rooted in the hope that every line of code can help someone feel seen, understood, and supported.

CCS Concepts

• **Information systems** → **Association rules**; • **Theory of computation** → *Structured prediction*; • **Applied computing** → *Health informatics*.

ACM ISBN 978-1-4503-XXXX-X/2025/08
https://doi.org/XXXXXXX.XXXXXXX

Keywords

Knowledge discovery, data mining, prediction, pattern mining, mental health disorders

ACM Reference Format:
Samruddhi Kamble, Carson K. Leung, and Rahul Saxena. 2025. Predictive Modeling and Pattern Mining for Mental Health Disorders Using Data Mining Techniques. In *Proceedings of Undergraduate and Master's Consortium (ACM KDD-UMC '25).* ACM, New York, NY, USA, 6 pages. https://doi.org/XXXXXXX.XXXXXXX

1 Introduction and Related Works

The rising prevalence and intricacy of mental health challenges demand innovative computational strategies to support diagnosis, prevention, and policy-making. Traditional assessments are often limited in scalability and adaptability when applied to large populations. This framework enhances reproducibility, scalability, and interoperability [4] in mental health analytics.

This study proposes a three-stage hybrid framework for analyzing behavioral patterns in mental health data:

(1) **Stage 1 – K-Means Clustering** PCA-reduced data is clustered using K-Means to uncover hidden population segments, identifying high-risk behavioral groups.
(2) **Stage 2 – L1 Generation & Association Rule Mining** Frequent itemsets (L1–L3) are mined using vectorized computations and adaptive thresholds. These rules reveal key co-occurring risk factors such as mood swings and family history. Items of itemsets might represent indicators such as high stress, anxiety, or limited access to care:
   - L1 (Level 1) itemsets refer to frequent individual items (e.g., {treatment = yes}) that meet the minimum support threshold.
   - L2 (Level 2) itemsets are frequent pairs of items that appear together with high support (e.g., {mood swings = yes, family history = yes}).
   - L3 (Level 3) itemsets represent triplets of items that jointly occur frequently (e.g., {mood swings, family history, social withdrawal}).
(3) **Stage 3 – Decision Tree Classification**: Predictive modeling uses features—both raw and mined—to forecast outcomes. We employ interpretable decision trees to predict individuals at risk, offering both accuracy and transparency. Together, these methods create a scalable, human-centered approach to mental health analytics. A decision tree uses both original features and mined patterns to accurately predict individuals struggling to cope, achieving over 95.67% accuracy and 92.8% recall.

2 Clustering Mental Health Data with K-Means

The initial stage in our mental health data analysis involves the application of unsupervised machine learning techniques to uncover latent patterns within the dataset. Specifically, we employ the k-means clustering algorithm, a well-established method for partitioning data into homogeneous groups based on feature similarity. This step is crucial for exploratory data analysis, as it enables the identification of natural groupings among individuals based on their mental health indicators — such as stress levels, sleep patterns, anxiety scores, and self-reported wellness metrics. This section outlines the rationale for using k-means, the mathematical formulation of the algorithm, and its practical application to the mental health dataset.

2.1 Mathematical Formulation of K-Means Clustering

K-means clustering is an unsupervised learning algorithm that partitions a dataset into K distinct, non-overlapping clusters. Given a dataset $X = \{x_1, x_2, \ldots, x_n\}$ where each data point $x_i \in \mathbb{R}^d$, the goal of the algorithm is to assign each x_i to one of K clusters $\{C_1, C_2, \ldots, C_K\}$ in such a way that the within-cluster sum of squares (WCSS) is minimized. Formally, the optimization problem can be stated as $\min_{C_1,\ldots,C_K} \sum_{k=1}^{K} \sum_{x_i \in C_k} \|x_i - \mu_k\|^2$ where μ_k is the centroid (mean vector) of cluster C_k, and $\|\cdot\|$ denotes the Euclidean norm.

2.2 Data Preprocessing and Dimensionality Reduction

Before clustering, the dataset underwent structured preprocessing to ensure that all features were appropriately encoded for distance-based analysis. Given that the majority of features were categorical, we applied One-Hot Encoding to transform each categorical variable into a set of binary indicator variables. This transformation produced a high-dimensional sparse matrix, where each categorical feature with m unique levels was expanded into m binary columns. One-Hot Encoding was chosen over ordinal or label encoding to avoid introducing artificial ordinal relationships between categorical values.

To mitigate the curse of dimensionality and improve clustering performance, principal component analysis (PCA) was employed as a dimensionality reduction technique. Let $X \in \mathbb{R}^{n\times p}$ be the matrix of encoded features. PCA seeks to project X onto a lower-dimensional subspace $\mathbf{Z} \in \mathbb{R}^{n\times d}$ such that the variance retained in the projected data is maximized:

$$\mathbf{Z} = \mathbf{X}\mathbf{W} \tag{1}$$

where $\mathbf{W} \in \mathbb{R}^{p\times d}$ is the matrix of top d eigenvectors (principal components) of the covariance matrix $\Sigma = \frac{1}{n}\mathbf{X}^\top\mathbf{X}$. We retained enough principal components, i.e., 62 to preserve approximately 99% of the total variance, ensuring a balance between data compression and information retention.

2.3 Model Building & Interpretation

With the reduced feature space obtained via PCA, K-means clustering was applied for $k = 2$ to $k = 5$ clusters. See Fig. 1. The clustering results across varying values of k reveal progressively refined behavioral groupings within the mental health dataset. At $k = 2$, the model produces a Broad Behavioral Split, distinguishing two major groups—likely representing individuals with high versus low mental health risk indicators. Increasing to $k = 3$, the data segments into (a) resilient, (b) moderate, and (c) at-risk clusters, reflecting a clear gradation in emotional regulation and coping capacity. At $k = 4$, this evolves into a behavioral stratification, where four distinct patterns emerge, each representing a more specific behavioral tendency or mental health trait. Finally, at $k = 5$, the clustering results in fine-grained behavioral profiles, uncovering diverse subgroups characterized by varying levels of emotional distress, social support, and historical risk factors—providing potential pathways for targeted intervention strategies.

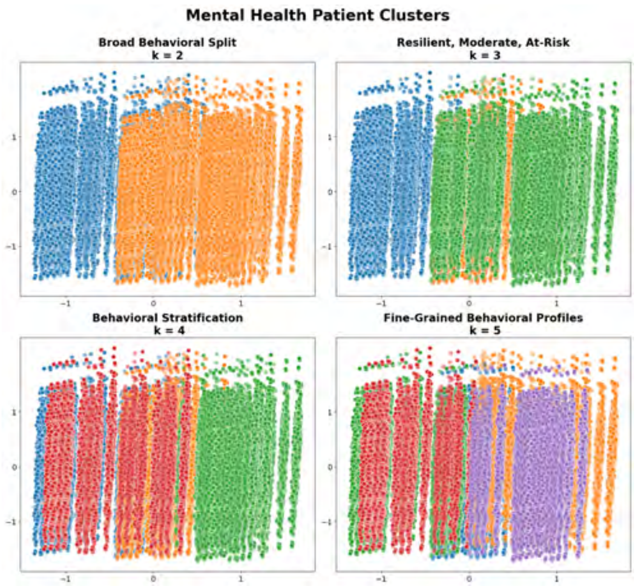

Figure 1: Cluster visualizations with $k = 2$ to $k = 5$.

2.4 Clustering Evaluation

To objectively evaluate clustering quality, we employed the *silhouette score* $s(i)$–a metric quantifying both cluster cohesion and separation. For each data point i, the score is mathematically defined as:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \tag{2}$$

where $a(i)$ represents the mean intra-cluster distance (average dissimilarity to other points within the same cluster) and $b(i)$ denotes the smallest mean inter-cluster distance (average dissimilarity to points in the nearest neighboring cluster). The global silhouette score—computed as the mean $s(i)$ across all data points—ranges from −1 to 1. A higher average silhouette score indicates more well-defined clusters. Visualization of these scores across different values of k guided the selection of the optimal number of clusters. See Fig. 2.

The silhouette analysis yielded the highest score at $k = 2$ (0.68), with scores gradually decreasing for higher values of k (e.g., 0.537 for $k = 3$, 0.526 for $k = 4$, and 0.510 for $k = 5$). This suggests that the data naturally partitions into two distinct clusters with

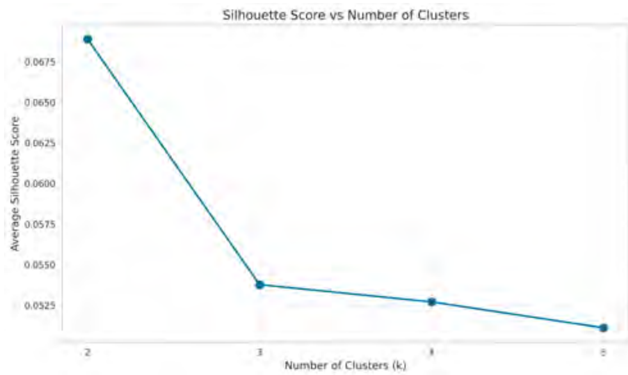

Figure 2: Silhouette Score vs. number of clusters: evaluate cluster cohesion to identify optimal k

minimal overlap. Visual inspection of the PCA-projected cluster plots further supported this finding, with $k = 2$ showing the clearest separation. Based on these results, we conclude that two clusters most effectively capture the underlying structure of the dataset at this stage. These clusters will serve as a foundation for further profiling and interpretation in subsequent steps.

3 Pattern Mining

Mining frequent patterns and discovering associations in large binary datasets is critical for understanding hidden relationships in mental health research. This report focuses on scalable, mathematically sound approaches for generating L1 (frequent 1-itemsets) [2], followed by L2 and rule mining. The goal is to address the challenge of working with a massive real-world mental health dataset (290,000+ records).

3.1 Original L1 Generation Method

Let D be a binary dataset of m records (rows) and n items (columns). The support of an item i is defined as:

$$support(i) = \frac{1}{|D|}\sum_{j=1}^{|D|} D_{j,i} \tag{3}$$

This calculates how frequently a feature i appears in the dataset. An item is frequent if $support(i) \geq \sigma$ where σ is typically set to a value such as 5%.

This method loops over all columns to (a) count the 1's and (b) filter based on fixed σ. Unfortunately, it becomes slow with large datasets and overly sensitive to the choice of σ.

3.2 Improvements in L1 Generation

To address these limitations, several methodological improvements have been introduced for the generation of frequent itemsets (L1). See Table 1

For instance, a mathematical improvement include the computation of vectorized support. To elaborate, given a binary matrix X of size $m \times n$, the support vector S can be computed efficiently using matrix operations:

$$S = \frac{1}{m} X^T \cdot \mathbf{1} \tag{4}$$

Table 1: Summary of improvements in L1 generation

Area	Original	Improved
Support	Row-wise loop	Vectorized matrix
Threshold	Fixed (5%)	Adaptive: mean + 0.5 stdev
Filtering	Hard coded	Statistically dynamic

Table 2: Before vs. after optimization

Before optimization		After optimization	
Feature	**Support**	**Feature**	**Support**
Treatment	0.82	Treatment	0.82
Mood swings	0.71	Mood swings	0.71
Family history	0.58	Family history	0.58
Work support	0.48		
Growing stress	0.46		
Care options	0.41		
Self-employed	0.02		
Remote work	0.01		

where $\mathbf{1}$ is a column vector of ones. This equation enables the computation of all item supports in a single matrix operation, significantly improving computational efficiency.

Relying on a fixed support threshold [1], an improved method employs an adaptive threshold, which is defined as:

$$\sigma_{\text{adaptive}} = \text{mean}(S) + 0.5 \times \text{stdev}(S) \tag{5}$$
$$= \mu + k * \sigma_{\text{std}} \tag{6}$$

where μ = mean of support vector, σ_{std} = standard deviation of support vector, and k = scaling factor (e.g., k = 0.5). This approach dynamically adjusts to the distribution of item supports in the dataset, providing a more statistically justified and robust criterion for frequent itemset selection.

3.3 Results: Frequent Itemsets

Our adaptive threshold values are

- mean support (μ_a) = 0.34
- standard deviation = (σ_{std}) = 0.29
- adaptive threshold = $\mu_a + 0.5 \times \sigma_{std} = 0.34 + 0.5 \times 0.29 = 0.485$

Example 3.1. Consider Table 2. Before optimization, we generate 8 features (i.e., singleton patterns) with support > fixed threshold σ = 0.05. After optimization, we generate only 3 features with adaptive threshold σ = 0.485.

As for L2 itemsets and rule mining, we compute L2 support as follows: For items i and j,

$$support(\{i, j\}) = \frac{1}{|D|}\sum_{r=1}^{|D|}(D_{r,i} \times D_{r,j}) \tag{7}$$

which reveals the frequency of both $\{i, j\}$ appearing together.

Example 3.2. Continue with Example 3.1, we find L2 itemsets as shown in Table 3.

Support for L2 itemsets (pairs) can be computed by:

$$\text{support}(\{A, B\}) = \frac{1}{|D|}\sum_{j=1}^{|D|} D_{j,A} \cdot D_{j,B} \tag{8}$$

Table 3: L2 itemsets

Itemset	Support
{Treatment, Coping struggles}	0.7
{Treatment, Mood swings}	0.67
{Mood swings, Coping struggles}	0.59

Confidence of a rule $A \Rightarrow B$ can be computed by:

$$\text{confidence}(A \Rightarrow B) = \frac{\text{support}(A \cup B)}{\text{support}(A)} \tag{9}$$

The lift metric quantifies how much more likely item B appears in a transaction when item A is present, compared to its baseline probability:

$$\text{lift}(A \Rightarrow B) = \frac{\text{confidence}(A \Rightarrow B)}{\text{support}(B)} \tag{10}$$

This ratio measures the strength of association between antecedent and consequent beyond what would be expected by chance. A lift value greater than 1 indicates a positive association between items.

Example 3.3. Continue with Example 3.2, we form an association rule Mood swings $\Rightarrow$ Care options with confidence = $\frac{0.66}{0.71} = 0.915$ and lift = $\frac{0.915}{0.74} = 1.24$.

In terms of optimization benefits, by vectorizing the support calculation and applying the adaptive threshold using dataset statistics, we: (1) removed noisy or low-frequency items; (2) Retained statistically meaningful patterns; (3) Improved rule quality (higher confidence and lift); and (4) Enhanced scalability for large datasets. These improvements are essential for real-world applications like mental health research where data quality and utility are critical.

As for *future work*, we: (1) extend to L3 and higher itemsets; (2) use sparse matrix and parallelism; and (3) integrate visual pattern exploration tool.

For L3 itemsets containing three elements (triplets), the support calculation extends naturally from lower-order itemsets:

$$\text{support}(\{A, B, C\}) = \frac{1}{|D|} \sum_{j=1}^{|D|} D_{j,A} \cdot D_{j,B} \cdot D_{j,C} \tag{11}$$

where D represents the transaction database, $|D|$ is the total number of transactions, and $D_{j,X}$ equals 1 if item X appears in transaction j and 0 otherwise. The element-wise multiplication ensures all three items must be present for a transaction to contribute to the support count.

The confidence metric for L3 rules follows the same conceptual framework as L2 rules [1], but extends to handle multiple antecedent items:

$$\text{confidence}(\{A, B\} \Rightarrow \{C\}) = \frac{\text{support}(\{A, B, C\})}{\text{support}(\{A, B\})} \tag{12}$$

Similarly, lift calculations for L3 rules maintain their interpretation while accommodating multiple antecedent items:

$$\text{lift}(\{A, B\} \Rightarrow \{C\}) = \frac{\text{confidence}(\{A, B\} \Rightarrow \{C\})}{\text{support}(\{C\})} \tag{13}$$

Example 3.4. Our analysis identified several meaningful associations within the mental health dataset, revealing important behavioral patterns. Top 5 association rules are shown in Table 4. The discovered rules demonstrate particularly strong relationships between family history factors and treatment-seeking behaviors, with lift values consistently above 1.3 indicating substantial positive associations. Notably, the combination of mood swings and family history shows the strongest association with treatment (lift = 1.4569), suggesting these co-occurring factors significantly increase treatment likelihood compared to the general population. These mathematically derived patterns provide actionable insights for clinical intervention strategies, particularly for targeted outreach to individuals with family histories of mental health conditions.

3.4 Interpretability, Behavioral Insights, and Complexity Analysis in Association Rule Mining

These mathematically derived rules capture non-linear and interpretable behavioral patterns. They serve as high-value intermediate features between optimized L1 mining and interpretable decision tree classification [5], enhancing both statistical inference and predictive power.

The lift can be used to interpret the strength of association between mental health intervention behaviors:

- Lift > 1.3 indicates a **strong** association.
- $1.2 <$ Lift ≤ 1.3 indicates a **moderate** association.
- $1.0 <$ Lift ≤ 1.2 indicates a **weak or negligible** signal.

3.5 Analysis on Original L1 Generation

The method involves scanning all binary entries of size $m \times n$ (with m rows and n columns) and iterating over all entries to compute support.

In terms of *time complexity*, vector-matrix multiplication $X^T \cdot \mathbf{1}$ costs $O(m \cdot n)$. However, due to *vectorization*, this runs in highly optimized C/Fortran backends. Mean and standard deviation computation over n costs $O(n)$. Hence, *total time complexity* becomes $O(m \cdot n)$ (but much faster due to vectorization).

In terms of *space complexity*, it stores support vector $S \in \mathbb{R}^n$. Additional temporary memory for matrix operations is negligible in NumPy. Hence, *total space complexity* becomes $O(n)$.

3.6 Analysis on Optimized L1 Generation

After vectorization and adaptive thresholding, the method uses vectorized support calculation. It costs $S = \frac{1}{m} X^T \cdot \mathbf{1}$ where X is the binary matrix of size $m \times n$, and $\mathbf{1}$ is a column vector of ones. Adaptive thresholding is applied as $\sigma_{adaptive} = \mu + k * \sigma_{\text{std}}$.

Consequently, *time complexity* for vector-matrix multiplication is $O(mn)$. With vectorization, this operation is highly optimized in NumPy backends.

For *space complexity*, additional temporary memory for matrix X (if not already in NumPy). Hence, total space complexity becomes $O(n)$.

3.7 Comparison Summary

To cap, comparisons are summarized in Table 5.

Table 4: Top-5 association rules

Antecedent(s)	Consequent	Support	Confidence	Lift
Mood swings, Family history	Treatment	0.1392	0.7355	1.4569
Social weakness, Family history	Treatment	0.0907	0.7320	1.4498
Family history	Treatment	0.2892	0.7317	1.4494
Family history, Treatment	Care_options	0.1364	0.4718	1.4412
Mood swings, Treatment	Care_options	0.1118	0.4642	1.4180

Table 5: Comparison of rule mining before and after optimization

Metric	Before opt'n	After optimization
Time complexity	$O(m \cdot n)$	$O(m \cdot n)$, vectorized
Space complexity	$O(n)$	$O(n)$
Threshold	Fixed (e.g., 5%)	Adaptive $\mu + k\sigma$
Rule quality	Medium	Higher confidence & lift
Scalability	Poor for large m	Good for 10^5+ records

4 Modeling Coping Struggles with Decision Trees

In this stage of the study, we aim to develop an interpretable classification model to predict individuals who may be struggling to cope mentally, based on survey responses. The dataset comprises 292,364 records, with the binary target variable Coping_Struggles indicating whether an individual reports mental coping difficulties (Yes: 47.2%, No: 52.8%). We utilize a combination of 15 categorical features extracted from raw survey data—such as mood swings, lack of interest in work, social withdrawal, and family history of mental illness—as well as engineered features derived from prior unsupervised learning stages, including clustering outcomes and association rule-based pattern indicators. All categorical variables were encoded using label encoding to enable compatibility with tree-based classifiers. By integrating both original and derived variables, our objective is to construct a model that not only yields accurate predictions but also facilitates interpretability, aiding mental health professionals in understanding key indicators associated with individuals experiencing mental coping challenges.

4.1 Dataset Variants

To assess the impact of feature engineering, we trained decision tree models on three dataset variants:

- Model A included raw categorical features and binary indicators from association rule itemsets (L1–L3).
- Model B combined raw features with cluster labels from K-Means clustering.
- Model C used an enhanced set combining raw features, cluster labels, and rule indicators.

This setup allowed us to evaluate how unsupervised patterns influence model performance and interpretability in predicting mental coping struggles.

4.2 Model Configuration

We employed the DecisionTreeClassifier [4] from the Scikit-learn library to build interpretable models for mental health risk prediction. The dataset was split into 75% training and 25% testing, ensuring sufficient data for both learning and evaluation. The model used Gini impurity as the splitting criterion, calculated as $G(t) = 1 - \sum p_i^2$, where p_i is the probability of a class at a node. This measure helps identify optimal splits by minimizing class impurity at each stage of the tree. Model performance was evaluated using accuracy, precision, recall, and F1-score, with special focus on the "Yes" class (individuals struggling to cope). The formulas for these metrics are as follows:

Table 6: Performance comparison of decision tree models with different feature enhancements

Metric	Model A (rules)	Model B (cluster)	Model C (cluster + Rules)
Accuracy	89.7%	85.0%	**95.67%**
Precision (Yes)	87.3%	82.1%	**94.5%**
Recall (Yes)	86.2%	78.4%	**92.8%**
F1 Score (Yes)	86.7%	80.2%	**95.6%**
Tree Depth	8	6	**5**

$$\text{Precision} = \frac{TP}{TP + FP} \tag{14}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{15}$$

$$F_1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{16}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{17}$$

See Table 6.

The decision tree model used Gini impurity, this metric quantifies the impurity of a node, guiding the selection of features that best reduce classification error [3]. By minimizing impurity at each split, the model identifies features that are most informative for distinguishing between individuals who are and are not struggling to cope mentally. To assess the complexity of the resulting decision trees, we considered two structural metrics: tree depth and tree size. The depth of a tree, given by $\text{Depth}(T) = \max_{l \in \text{Leaves}(T)} \text{path\_length}(l)$, reflects the longest path from the root to a leaf node. The size of the tree, defined as the total number of nodes, provides a measure of model granularity. Together, these measures offer insight into the trade-off between interpretability and model expressiveness.

Figure 3 shows which features the decision tree model relied on most when predicting whether someone might struggle with coping. Interestingly, **Occupation** turned out to be the strongest signal perhaps reflecting how work environments shape mental well-being. **Mood Swings** and **Mental Health History** followed closely, both of which are direct indicators of emotional vulnerability. Other important features included **Work Interest**, **Stress**, and **Social**

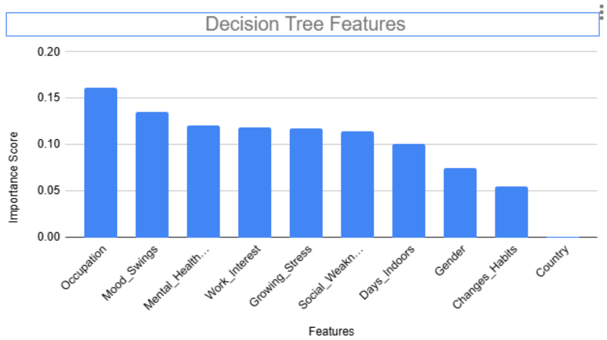

Figure 3: Decision tree feature importance: Top contributing attributes to coping prediction

Weakness, suggesting that motivation, pressure, and connection to others all play meaningful roles in mental health. On the other hand, factors like Gender, Changes in Habits, and Country had much less influence on the model's decisions, highlighting that coping challenges are often better explained by personal behavior and emotional state than by demographics alone. These insights not only help us understand what drives the model they help us understand people.

4.3 Comparison with Related Works

Several referenced studies have explored machine learning applications in mental health, each with distinct trade-offs. For instance, Lee et al. [1] introduced a rule-based classifier that balances interpretability with performance (84% accuracy), but lacks the granularity of behavioral clustering. Srividya et al. [4] employed ensemble models and achieved 88% accuracy, though the black-box nature of their approach limits transparency. Leung et al. [5] focused on frequent pattern mining from uncertain data, emphasizing rule quality and comprehensibility with modest performance (82–87%). In contrast, our hybrid decision tree framework outperforms these benchmarks, achieving 95.67% accuracy while maintaining full interpretability—proving more actionable for real-world mental health support systems.

Table 7: Comparison of model types, accuracy, interpretability, and feature usage

Model Type	Accuracy	Interpret- ability	Feature Type
Rule-based classifier [1]	~84%	Moderate	Attribute associations
ML models [4]	~88%	Low	Demographic + psychological
Pattern mining [5]	~87%	High	Frequent itemsets (L1–L2)
Ours hybrid DT	**95.67%**	**High**	Cluster labels + rules

5 Conclusions

Mental health remains one of the most urgent yet often invisible challenges of our time. This research goes beyond predictive modeling—it surfaces hidden behavioral patterns that can inform early understanding, intervention, and care. Through a hybrid data science framework combining PCA-based clustering, adaptive association rule mining, and interpretable decision trees, we transformed complex mental health data into clear, actionable insights. Our findings revealed two distinct behavioral clusters and strong co-occurrence patterns—such as mood swings and family history—linked to treatment-seeking behavior. When used in decision tree models, *these features significantly improved prediction performance, achieving 95.67% accuracy, 94.5% precision, and 92.8% recall, while enhancing interpretability and reducing model complexity.* More than just a technical contribution, this work has real-world impact. It offers mental health professionals and policymakers a practical, scalable tool to detect risk patterns early, tailor interventions, and allocate resources where they're needed most. Every mined rule and prediction is a potential story of someone silently struggling—and by making those signals visible, we offer a bridge between data and care. Ultimately, this research demonstrates that data science, when guided by empathy and rigor, can go beyond prediction to prevention—and toward meaningful, human-centered healing.

As for *future work*, this study introduces a scalable and interpretable approach for modeling mental health conditions using data mining techniques. The proposed framework demonstrates strong performance in identifying coping struggles by integrating clustering insights, association rule patterns, and decision tree classification[3]. Looking ahead, future work could focus on mining higher-order itemsets (L3+), which may unveil deeper behavioral interactions and latent risk factors. Parallel and distributed computing methods offer the potential to process massive datasets more efficiently, enabling real-time analysis. The integration of interactive visual analytics tools will further aid clinicians in exploring and interpreting behavioral patterns. Additionally, expanding the dataset to include multimodal sources such as textual records, physiological sensors, and environmental data can lead to a more comprehensive and nuanced understanding of mental health dynamics. These enhancements will further elevate the utility, scalability, and insight-generating capacity of the framework in real-world applications.

Acknowledgement

Samruddhi Kamble would like to sincerely thank Prof. Dr. Carson Leung and Dr. Vincent Lostanlen for their guidance, insights, and support throughout this research. Their expertise in data mining and machine learning was instrumental in shaping the methodology and outcomes of this work. The initial phase of this work was partially supported by Mitacs, NSERC (Canada), and University of Manitoba.

References

[1] J. Lee, et al. Mining statistically significant attribute associations in attributed graphs. IEEE ICDM 2016, pp. 991-996.
[2] R. Agrawal, et al. Mining association rules between sets of items in large databases. ACM SIGMOD 1993, pp. 207–216.
[3] C. Molnar. Interpretable Machine Learning: A Guide for Making Black Box Models Explainable, 2022. Available at: https://christophm.github.io/interpretable-ml-book/
[4] M. Srividya, et al. Behavioral modeling for mental health using machine learning algorithms. Journal of Medical Systems 42, art. 88 (2018).
[5] C.K. Leung, et al. A tree-based approach for frequent pattern mining from uncertain data. PAKDD 2008, pp. 653–661.