

**EY Data
Challenge
Hackathon 2026**

Barbara Ortiz

**Innovating Mental Health
Risk Assessment: Predictive
ML & Employee Profiling**

**Workplace Mental Health
Risk Modeling**

Predicting Workplace Mental Health Outcomes

Objective:

- Predict current mental health condition.
- Predict likelihood of treatment seeking.
- Identify structural drivers.
- Segment employee profiles.



Mental health significantly impacts productivity and retention.

Approach:

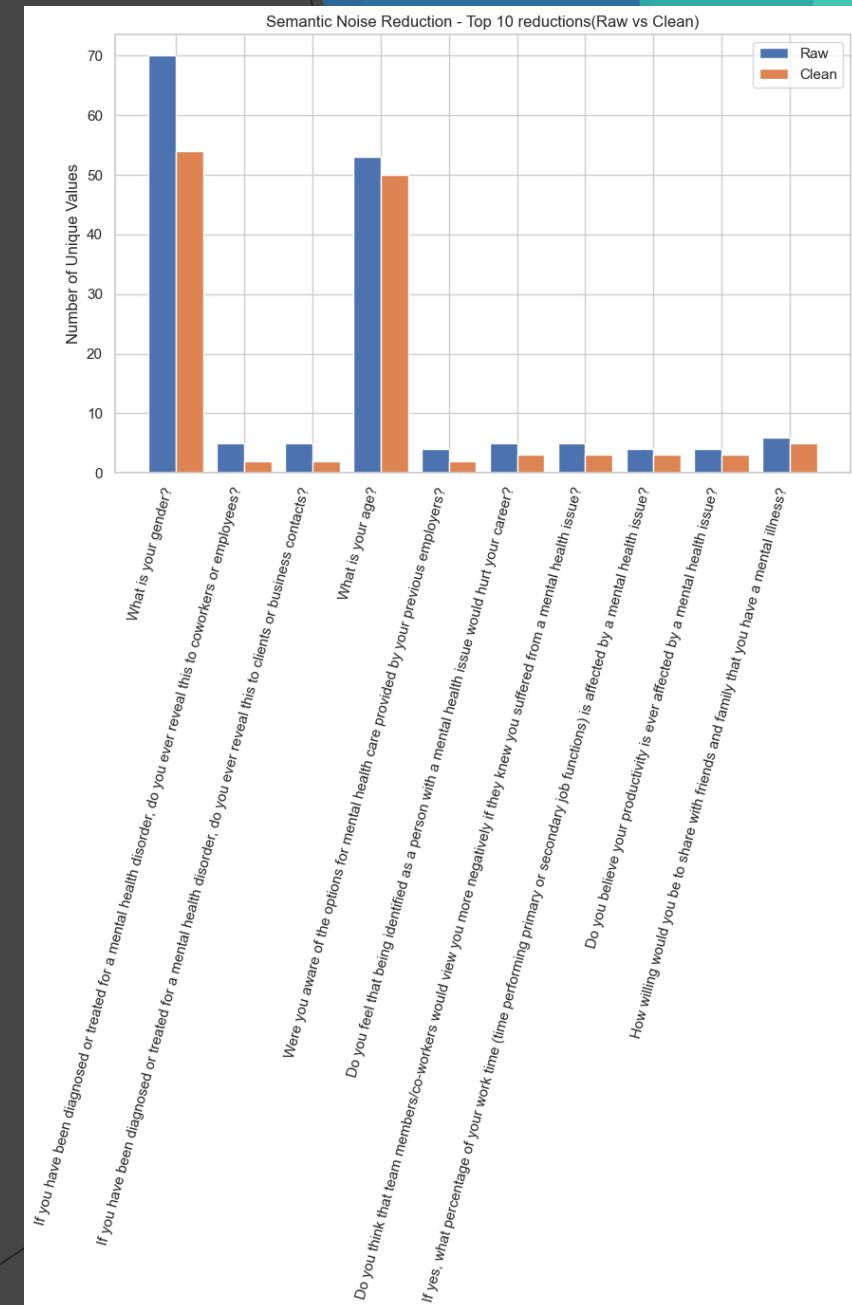
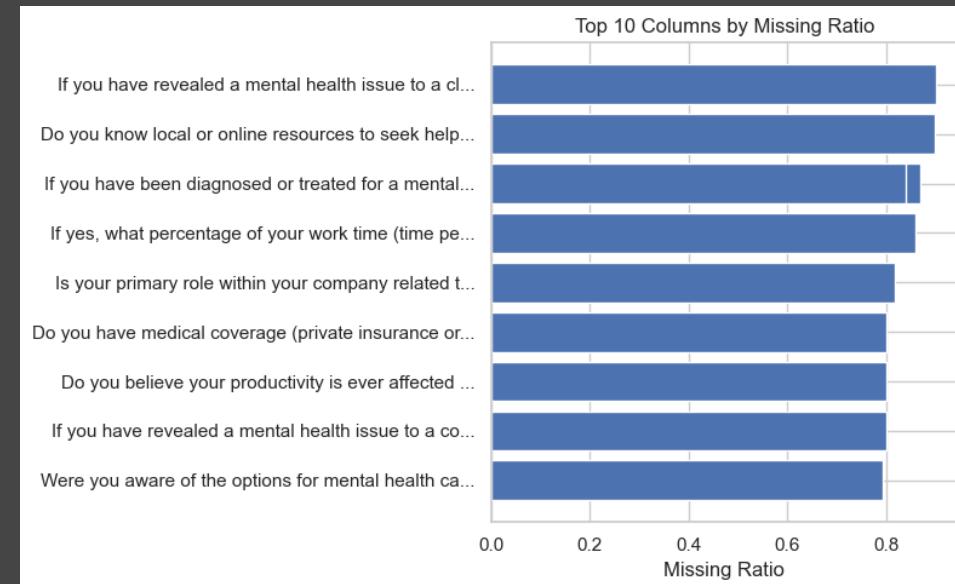
Data cleaning → Feature engineering → Predictive modeling → Clustering → Business insights

Data Preparation & Normalization

I prioritized interpretability over aggressive imputation to avoid artificial signal distortion.

Decisions:

- Conditional logic applied to dependent survey questions
- Ordinal encoding for ordered scales
- Binary mapping for yes/no variables
- Standardization applied only for clustering



Feature Engineering & Index Construction

Composite indicators were constructed to quantify latent organizational constructs such as perceived support and stigma.

Mental Health Support Index:

$$Support = \frac{Benefits + KnowledgeOfOptions + FormalCommunication + LearningResources + AnonymityProtection + ParityWithPhysicalHealth}{6}$$

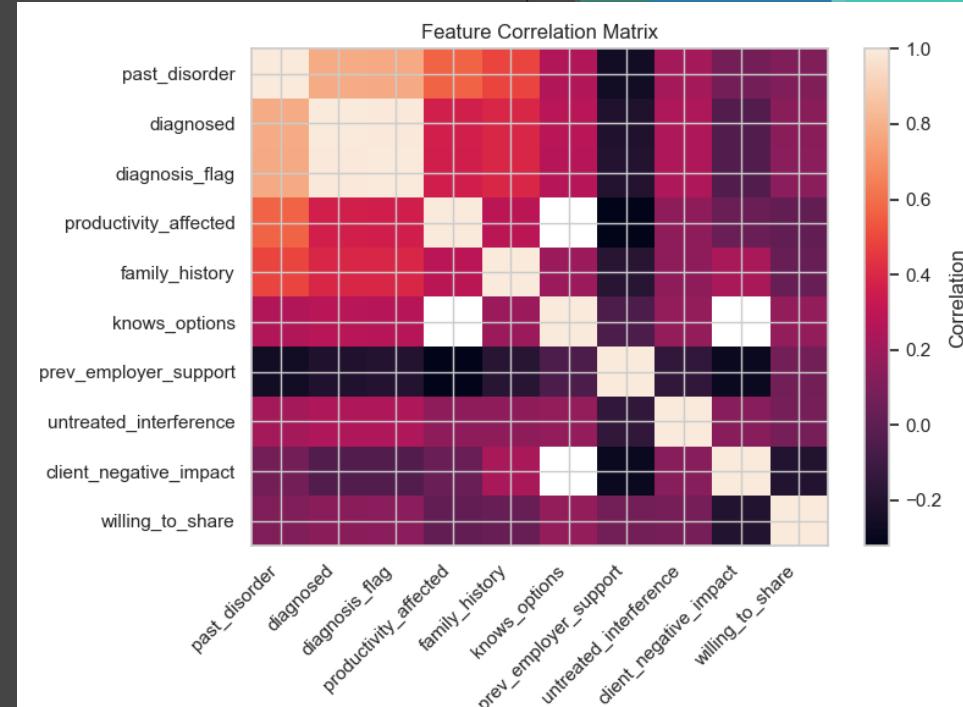
Workplace Stigma Index:

$$Stigma = \frac{CareerImpact + EmployerReaction + ObservedDiscrimination + (1 - WillingnessToShare)}{4}$$

Organizational Openness Score:

$$Openness = \frac{WillingnessToShare + KnowledgeOfOptions + EmployerSeriousness}{3}$$

Indexes reduce dimensionality while preserving structural meaning.

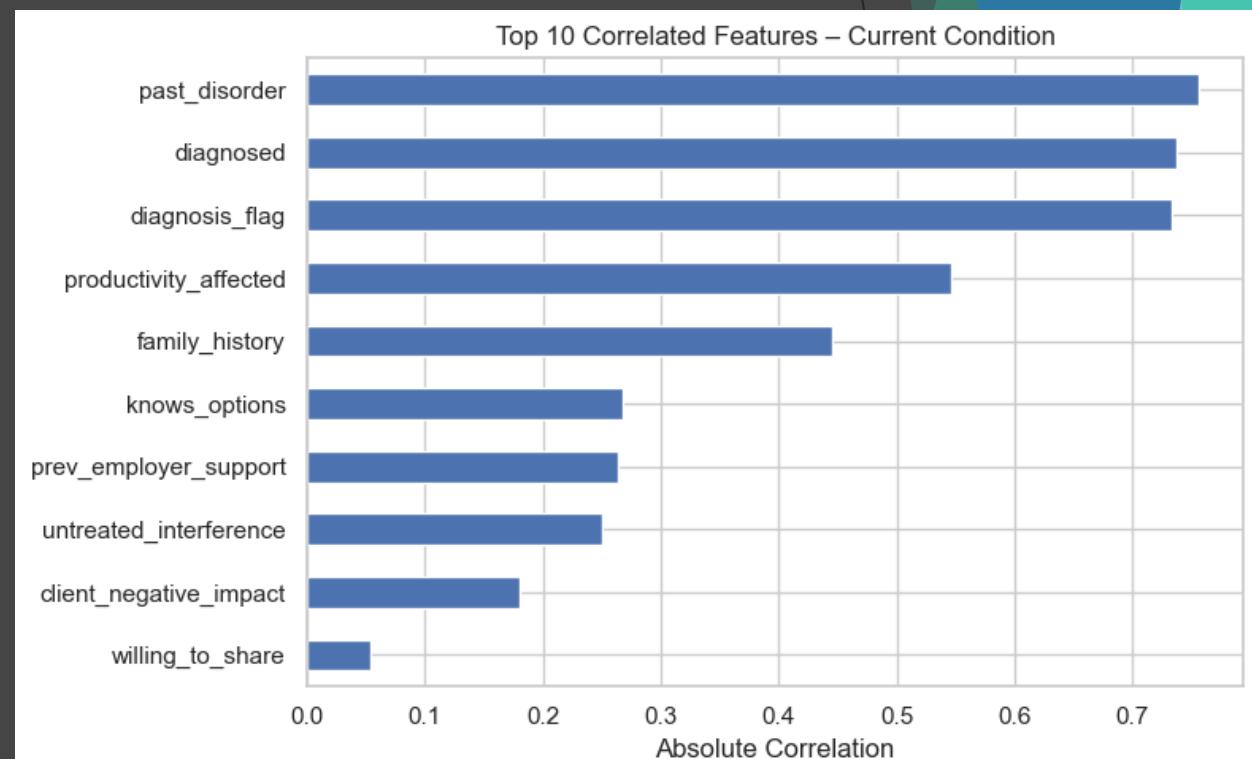
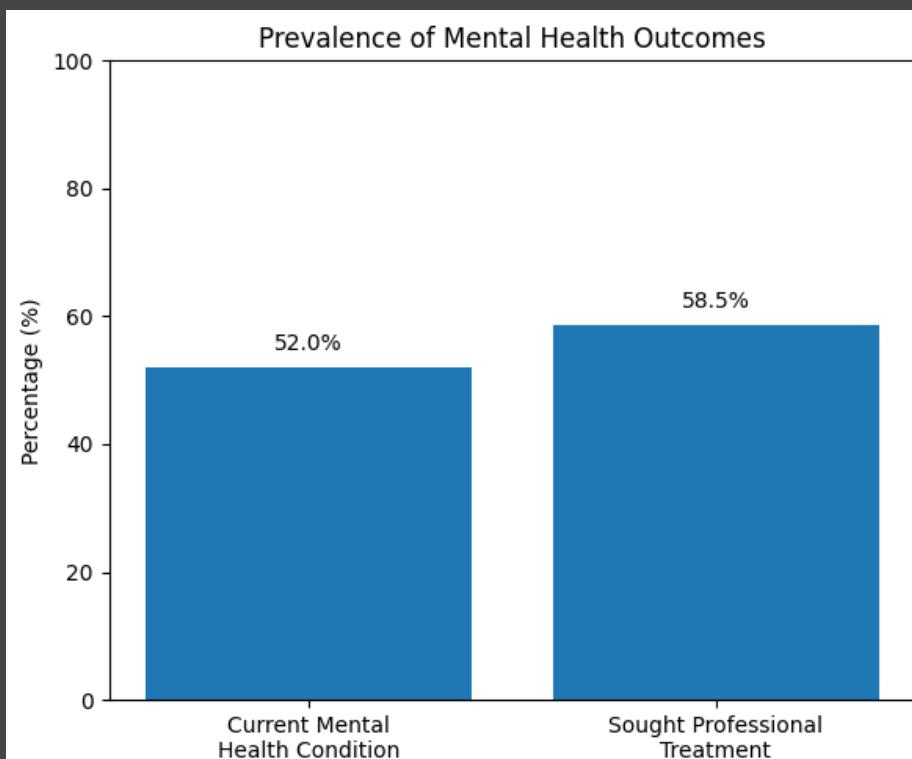


Exploratory Analysis

The data reveals clear and stable mental health patterns. Past conditions, professional diagnosis, and workplace productivity impact consistently align with current mental health status.

These findings suggest that mental health risk is structured and persistent, not random.

Overlapping diagnostic indicators were reviewed to ensure robust and unbiased modeling.

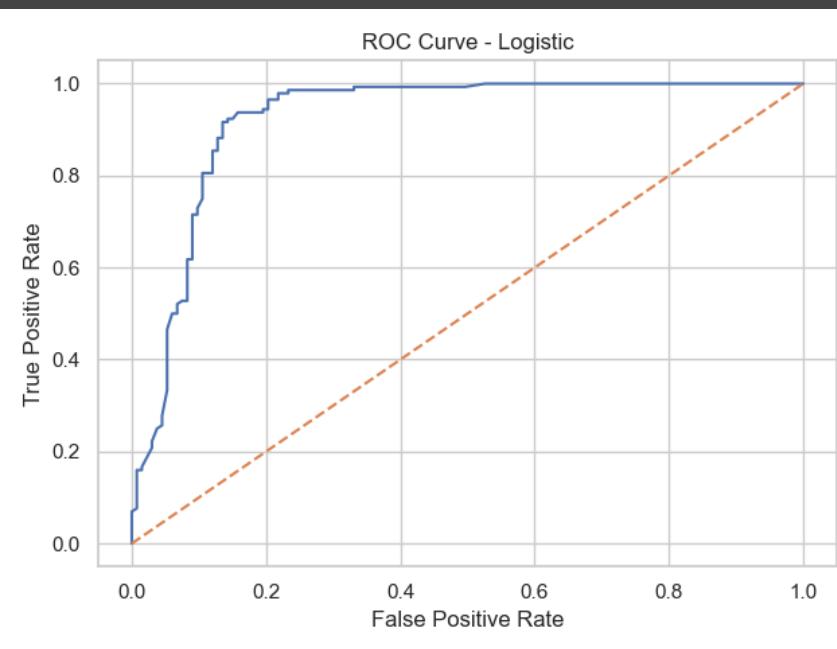


Predictive Model: Current Condition

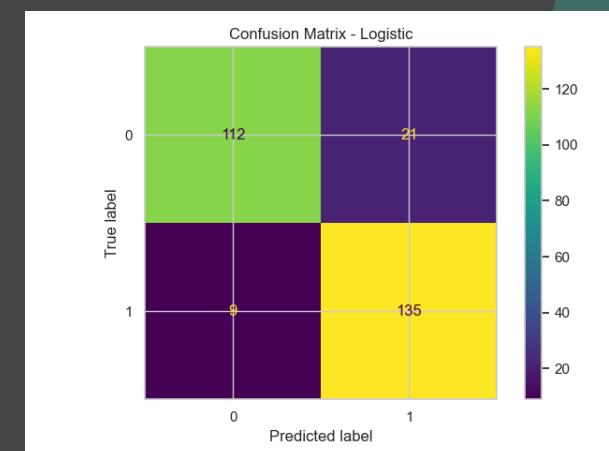
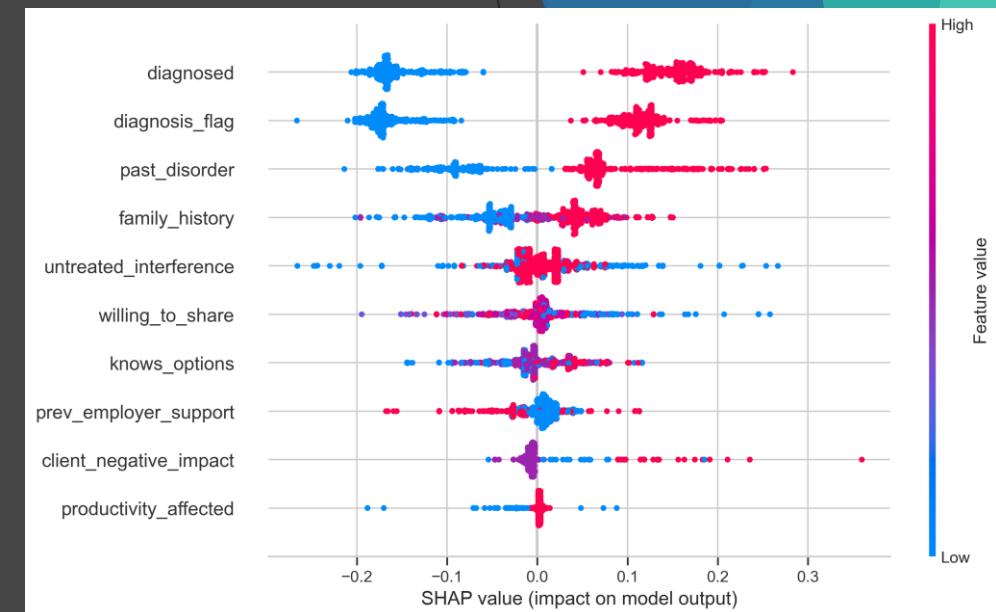
Logistic Regression outperformed Random Forest, suggesting predominantly linear and structured relationships.

Top predictors:

- Past disorder
- Formal diagnosis
- Productivity interference



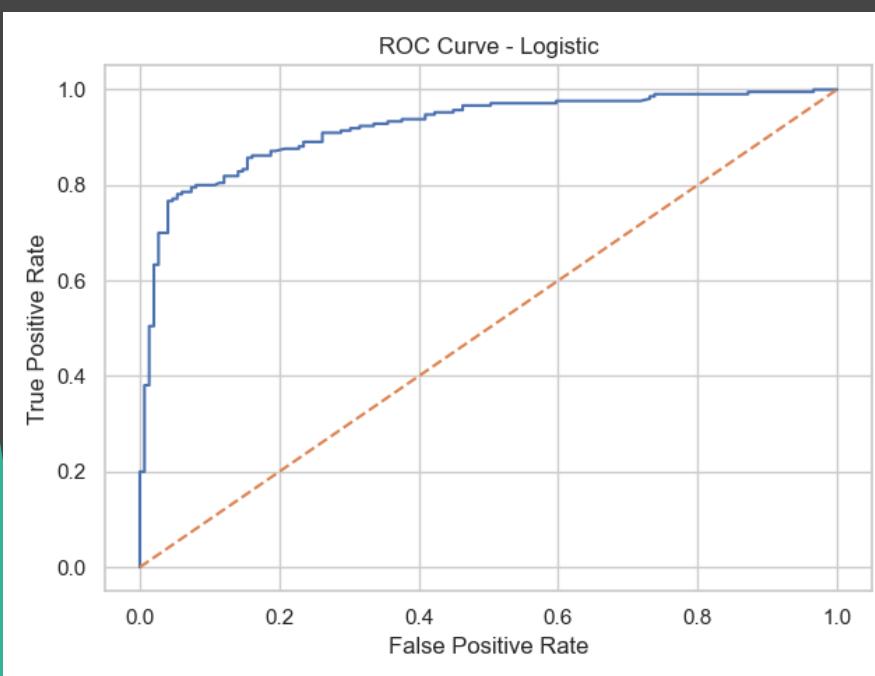
The model correctly identifies 94% of employees experiencing a mental health condition, with an overall accuracy of 89%.



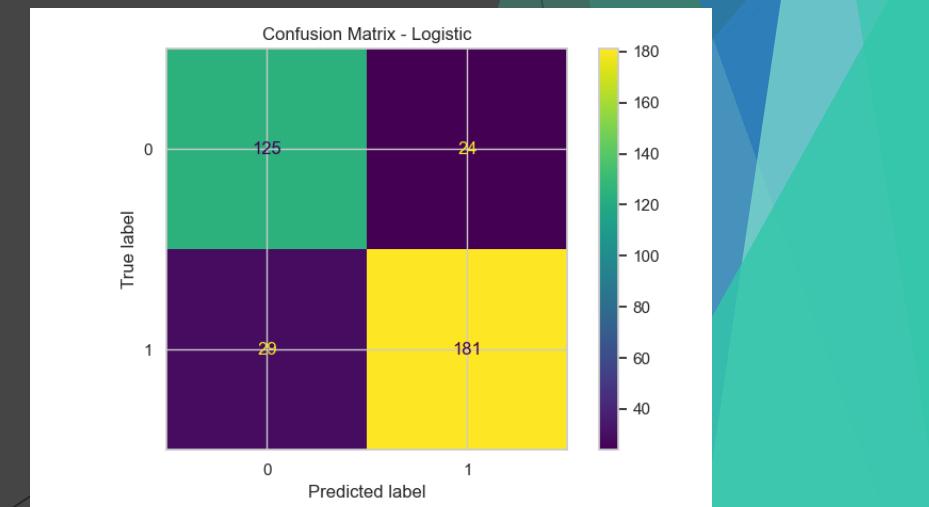
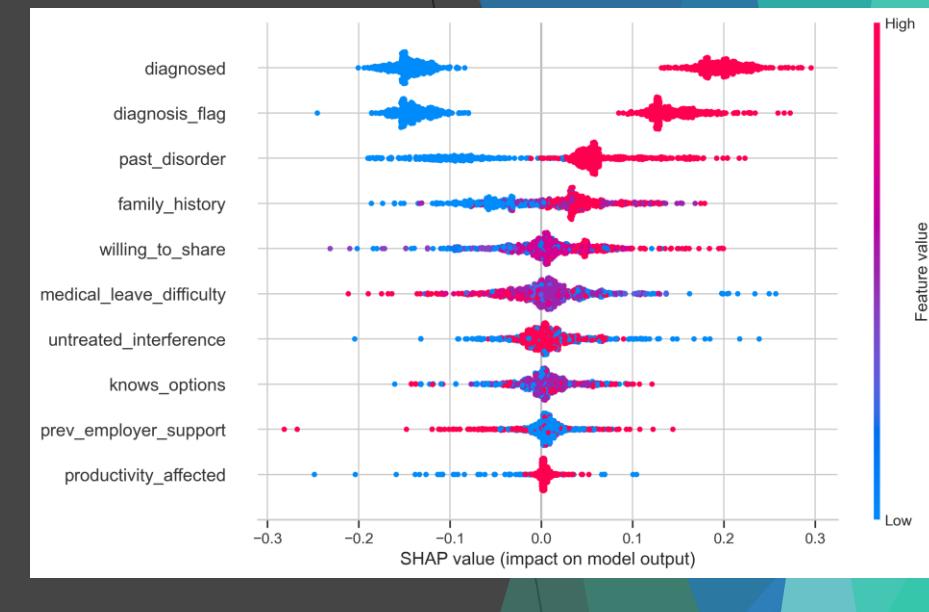
Predictive Model: Treatment Seeking

Knowledge of support resources and openness variables were stronger predictors of treatment than stigma alone.

Access awareness > stigma perception in influencing treatment behavior.



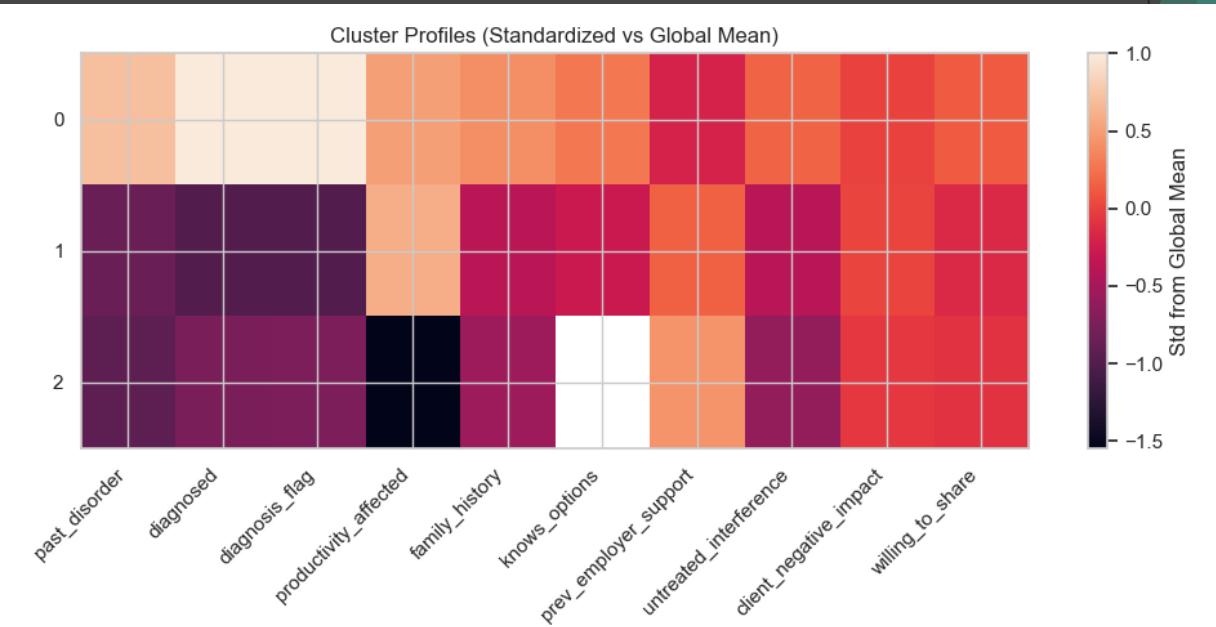
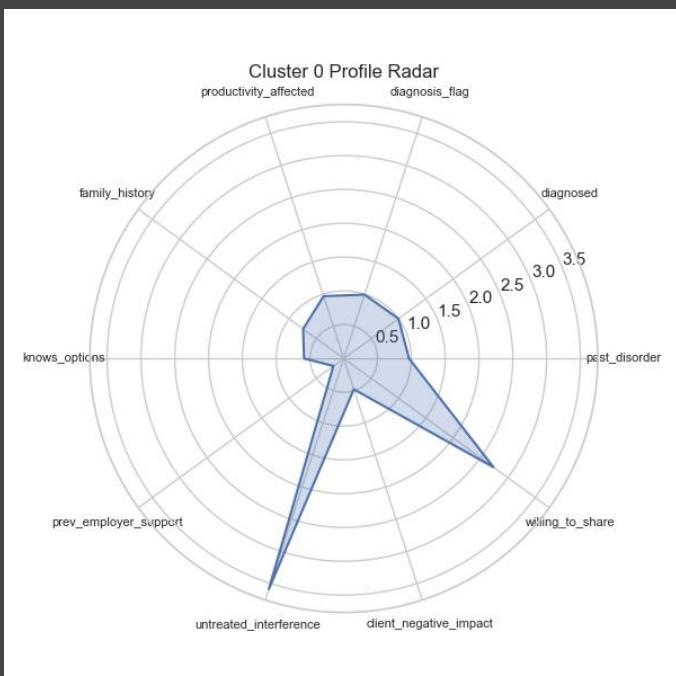
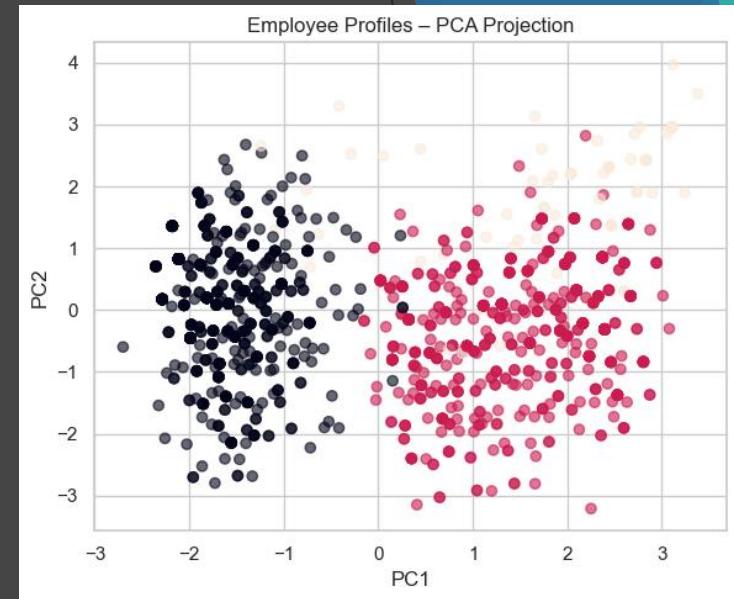
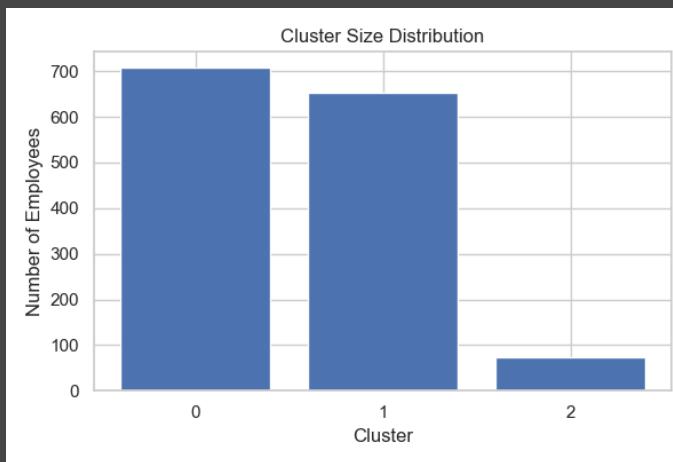
The model correctly identifies 86% of employees experiencing a mental health condition, with an overall accuracy of 85%.



Clustering: Employee Profiles

Segmentation reveals actionable groups beyond binary prediction.

1. Supported & Open
2. Affected but Undiagnosed
3. Low Support & High Stigma



Business Recommendations & Strategic Impact

① Early Risk Detection

Implement voluntary confidential screenings.

② Visibility of Resources

Quarterly awareness campaigns.
Internal mental health dashboards.



③ Target Hidden Risk Segment

Employees with productivity impact but no diagnosis.

④ Manager Psychological Safety Training

Increase Organizational Openness Score.

Final Reflection

Workplace mental health is predictable, measurable, and structurally influenceable.

The strong predictive signal ($\text{ROC-AUC} > 0.92$) combined with actionable segmentation demonstrates that data-driven strategies can meaningfully reduce untreated cases and productivity loss.

Data-driven policies transform support from reactive to preventive. Mental health strategy is not only ethical, it is economically strategic.