

# Mental Health Prediction using IBM Watson Auto AI

Leveraging AI for Early Detection and Awareness

Presented by: Hamza Haidar

Date: August 2025



# Project Objective: Proactive Mental Health Support

## Goal: Early Risk Prediction

Our primary objective is to identify individuals at potential risk of developing mental health issues by analyzing patterns in anonymous survey data. This proactive approach aims to facilitate timely interventions and support.

## Approach: Automated AI with IBM Watson AutoAI

We leverage IBM Watson AutoAI to streamline the entire machine learning lifecycle. This includes automated data preprocessing, intelligent model selection, efficient training, and robust evaluation, significantly accelerating our development.

# Core Technologies: IBM Watson AI Ecosystem

This project relies on the robust and integrated environment provided by IBM Watson Studio, specifically harnessing the power of its AutoAI capabilities.



## IBM Watson Studio (AutoAI)

Our central platform for end-to-end AI lifecycle management, enabling automated model building and deployment.



## Python 3.11 & Libraries

Utilizing a comprehensive stack including `ibm-watsonx-ai`, `autoai-libs`, `lale`, `scikit-learn`, `xgboost`, `lightgbm`, and `snappyml` for advanced machine learning operations.



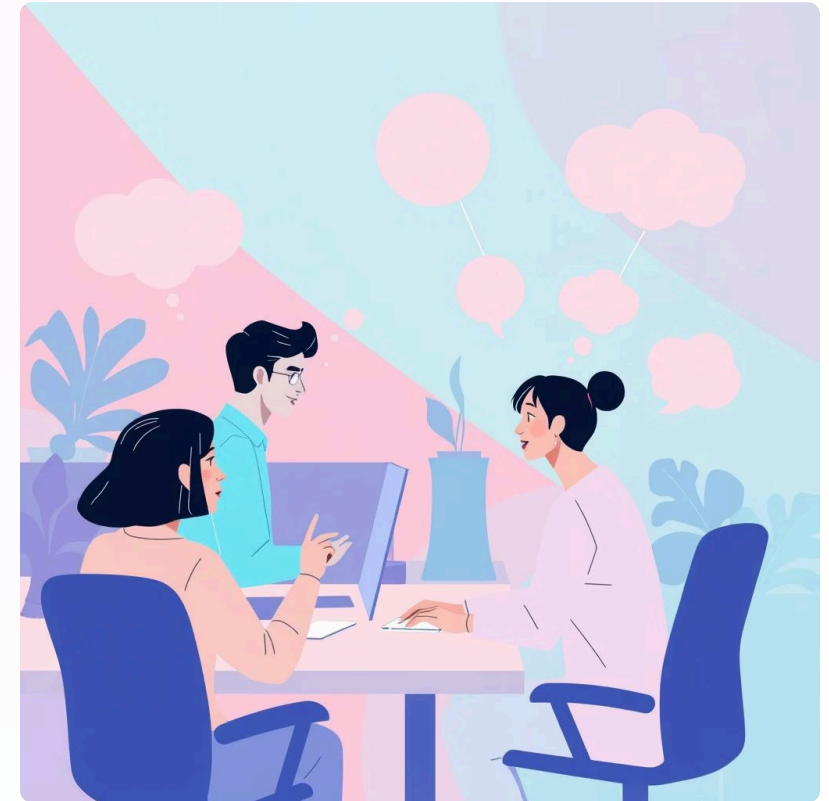
## watsonx.ai Runtime

The high-performance environment ensuring efficient execution and scalability for our AI models, crucial for handling complex data operations.

# Dataset Insights: Mental Health in Tech Industry

Our analysis is based on a **Mental Health in Tech Industry Survey**, capturing nuanced insights into the well-being of professionals in the technology sector. This dataset provides a rich foundation for predictive modeling.

- **Key Features:** Age, gender, family history of mental illness, impact of mental health on work performance (work interference), company size, and remote work status.
- **Target Variable:** A binary classification indicating whether the individual **sought professional treatment** for a mental health condition.
- **Prediction Goal:** To classify individuals into "Yes" (sought treatment) or "No" (did not seek treatment), enabling targeted support.



# AutoAI Experiment Setup: From Data to Model



The setup process within IBM Watson AutoAI is intuitive, designed to guide users through the crucial steps of preparing and configuring an automated machine learning experiment. This ensures a robust foundation for model training and optimization.

# Automated Model Training & Optimization

IBM Watson AutoAI's strength lies in its ability to automate the intricate processes of model training and optimization, significantly reducing manual effort and accelerating development cycles.

## Pipeline Generation

AutoAI intelligently generates and evaluates **multiple candidate pipelines**, each representing a unique combination of algorithms and preprocessing steps.

## Comprehensive Optimization

Each pipeline undergoes extensive optimization, including **automated data preprocessing**, sophisticated **feature engineering** to extract key insights, and fine-tuned **model hyperparameter tuning**.

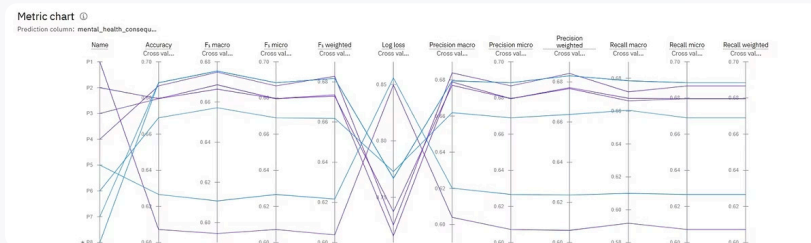
## Optimal Pipeline Selection

The system then ranks these pipelines based on a specified metric (e.g., accuracy), prioritizing the **best-performing yet simplest pipeline** for deployment.

This automation ensures that the final model is not only accurate but also robust and efficient, ready for real-world application.

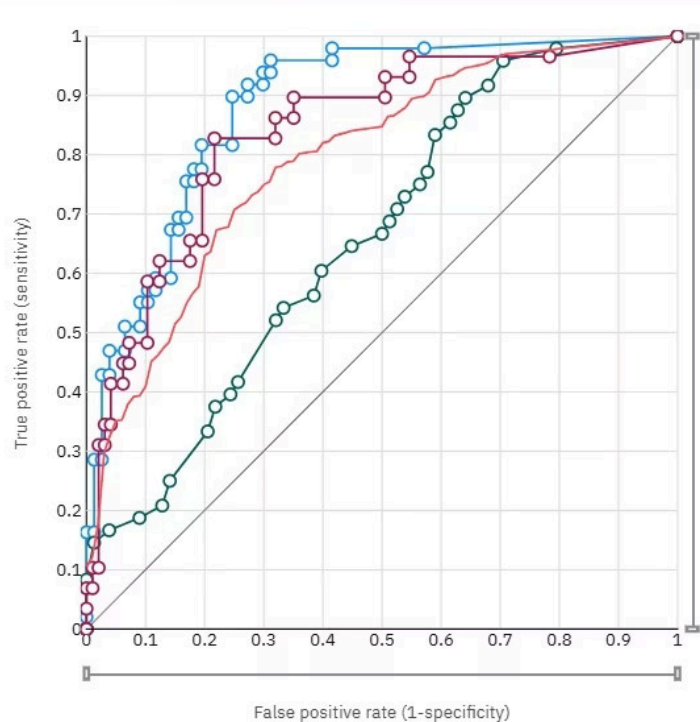
# Visualizing & Comparing Model Performance

AutoAI provides intuitive visualization tools to compare the performance of different generated pipelines, ensuring transparency and informed decision-making.



- **Top Pipeline Visualization:** We can visually inspect the performance of the top-performing models, analyzing key metrics and their behavior across different thresholds.
- **Diagnostic Metrics:** Detailed ROC curves and confusion matrices provide a comprehensive understanding of model sensitivity, specificity, and classification accuracy.
- **Export & Integration:** The best-performing pipeline is easily exportable as a scikit-learn compatible model, allowing for seamless integration into existing workflows or custom applications.

# Key Findings and Predictive Power



Our AutoAI experiment yielded promising results, demonstrating the potential for early identification of individuals at risk of mental health challenges.

- **Best Accuracy: 68.8%** - This indicates the model's ability to correctly predict whether an individual sought treatment based on their survey responses.
- **ROC-AUC Score: 0.61** - A respectable score for binary classification in sensitive domains, showing the model's capacity to distinguish between positive and negative classes.

## Top Contributing Features:

- **Mental Health Consequence:** The perceived impact of mental health on daily life was the strongest predictor.
- **Family History:** A history of mental illness in the family significantly influenced the prediction.
- **Work Interference:** The extent to which mental health issues affected work performance was also a critical factor.



# Transformative Business Impact

This AutoAI-driven solution offers significant advantages for organizations committed to employee well-being.

## ✓ Automated Risk Detection

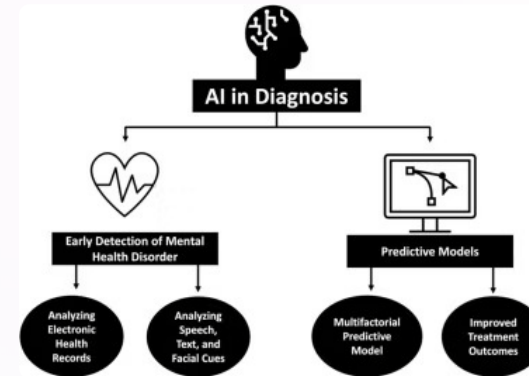
Provides a scalable and consistent method for identifying individuals who might benefit from early intervention, reducing reliance on subjective assessments.

## i Empowering HR & Wellness Teams

Offers data-driven insights, allowing HR and wellness departments to allocate resources effectively and develop targeted support programs.

## 📄 Low-Cost & Interpretable AI

Leveraging AutoAI minimizes development costs and time, while the transparency of the identified top features ensures interpretability, building trust in the predictions.



# Future Directions & Expansion

Our project lays a strong foundation for future advancements in mental health prediction and support.



## Real-Time Integration

Integrate the trained model with a real-time survey application to provide immediate feedback and insights to individuals and wellness providers.



## Continuous Monitoring

Implement robust monitoring for model drift, ensuring the predictive accuracy remains high over time as new data emerges or trends shift.



## NLP Analysis Expansion

Extend the analysis to incorporate Natural Language Processing (NLP) of open-ended survey responses, uncovering deeper qualitative insights into mental well-being.

These steps will enable a more dynamic, comprehensive, and impactful mental health support system.