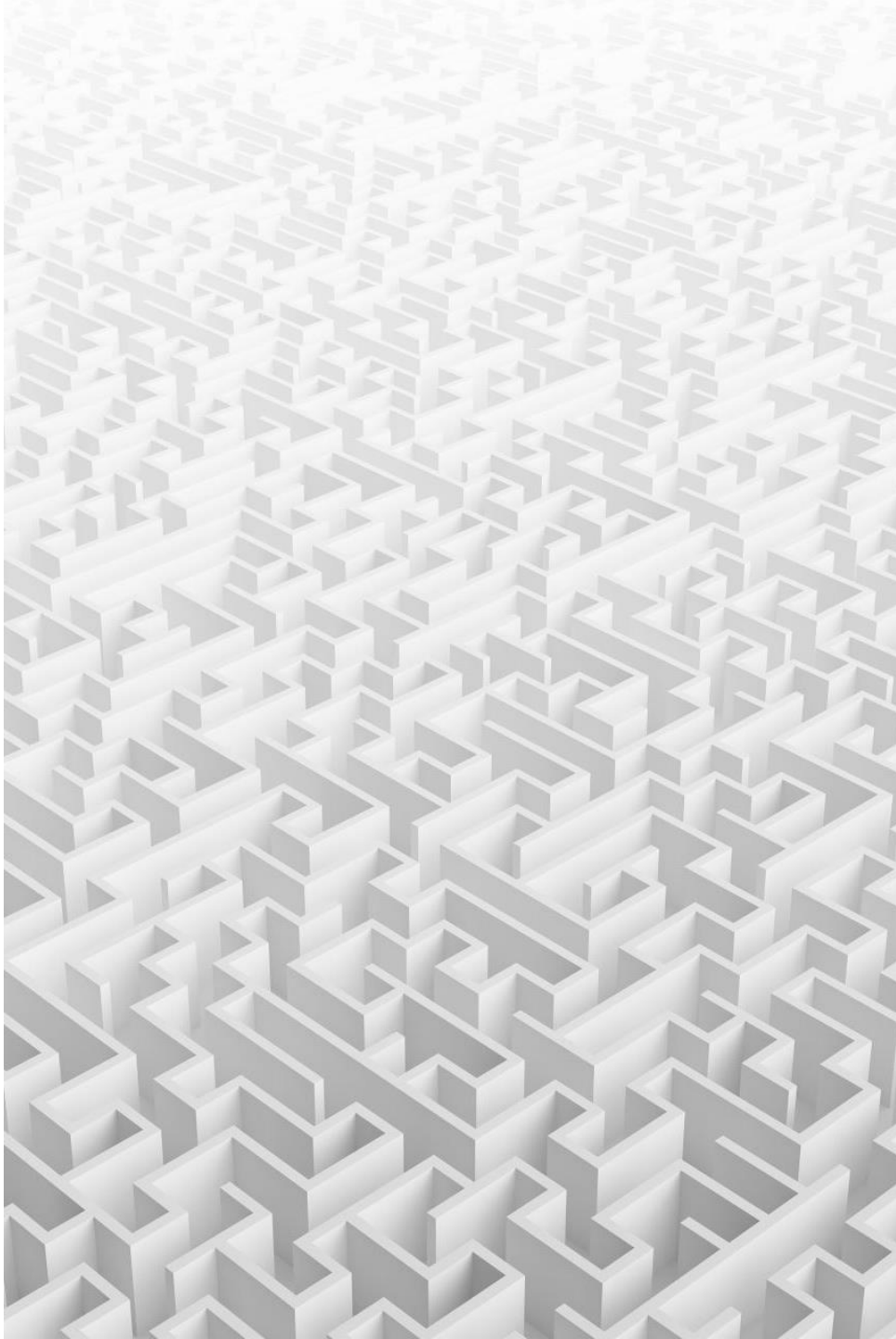
Substance Abuse Risk Analysis

- A Data-Driven Approach to Identifying Substance Abuse Risk
- Presented by Neal Iyer

Business Understanding

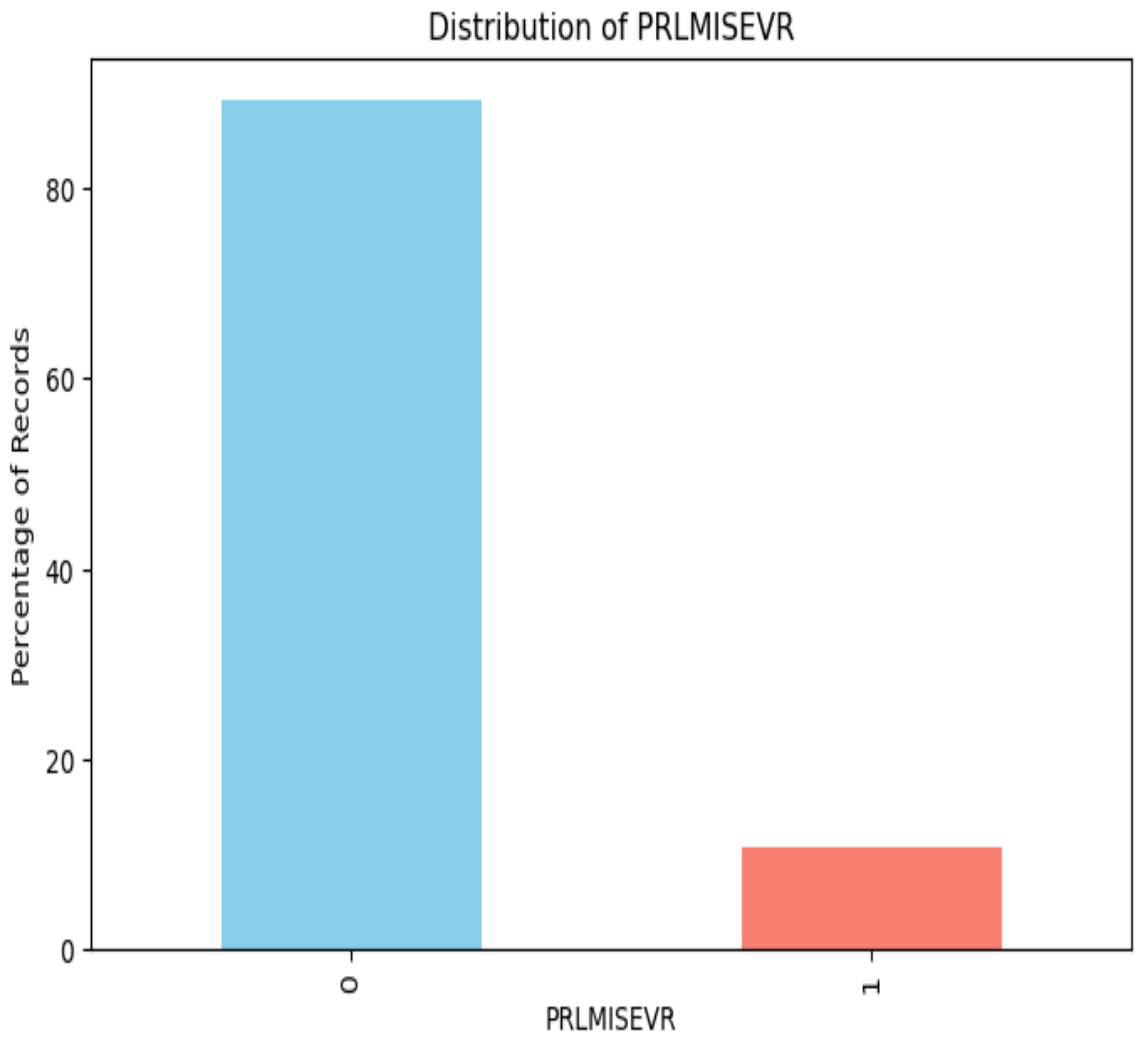
- Goal: Identify the more effective survey dataset for predicting substance abuse risk to assist healthcare professionals in early detection.

- Why It Matters: Early identification can guide resources toward individuals who are at higher risk, potentially reducing substance abuse cases.

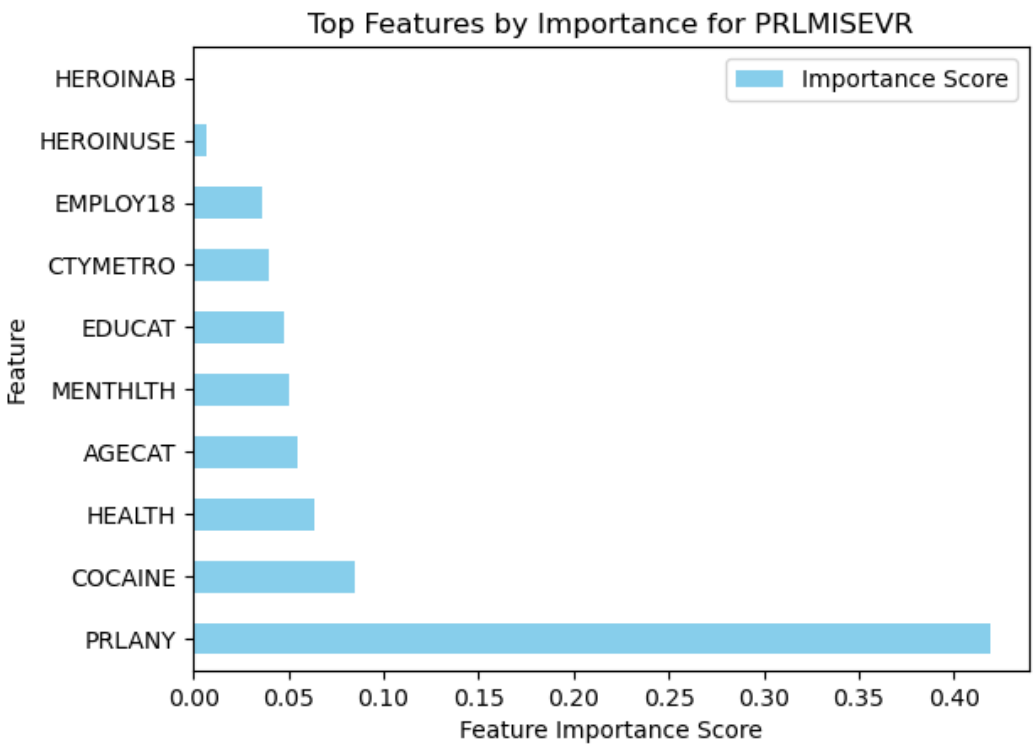


Data Understanding

- PRLMIS Data: General health survey covering physical health, demographics, and drug use.
- Student Addiction Data: Focused on younger individuals, with data on behavior, academic performance, and social influences.
- Main Question: Which dataset offers the strongest foundation for predicting substance abuse risk?



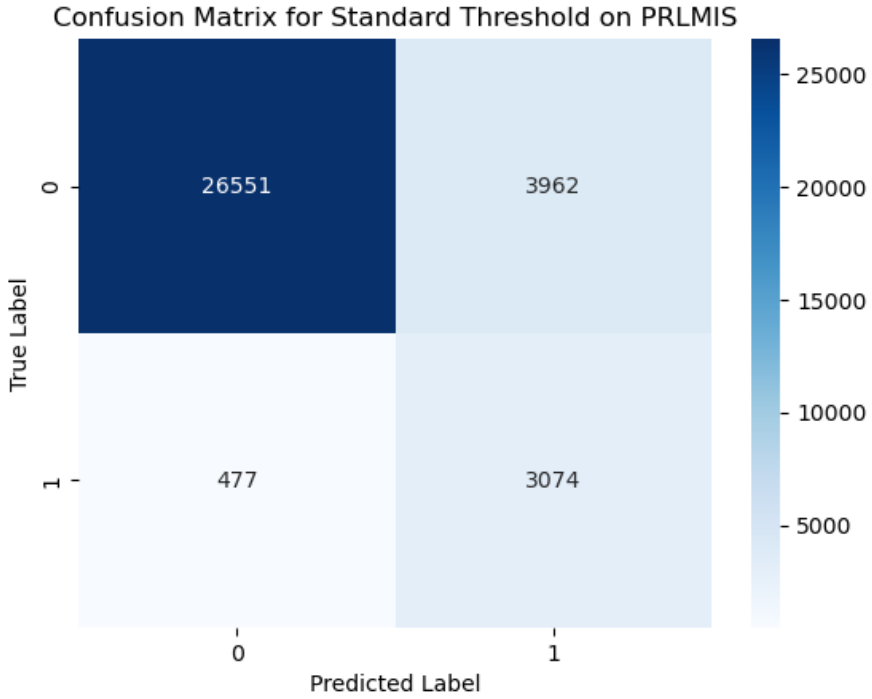
Data Preparation



- Chose the top 10 influential factors in each dataset to streamline analysis and focus on the most important indicators.
- Cleaned and prepared each dataset to ensure consistent, reliable results for our analysis.

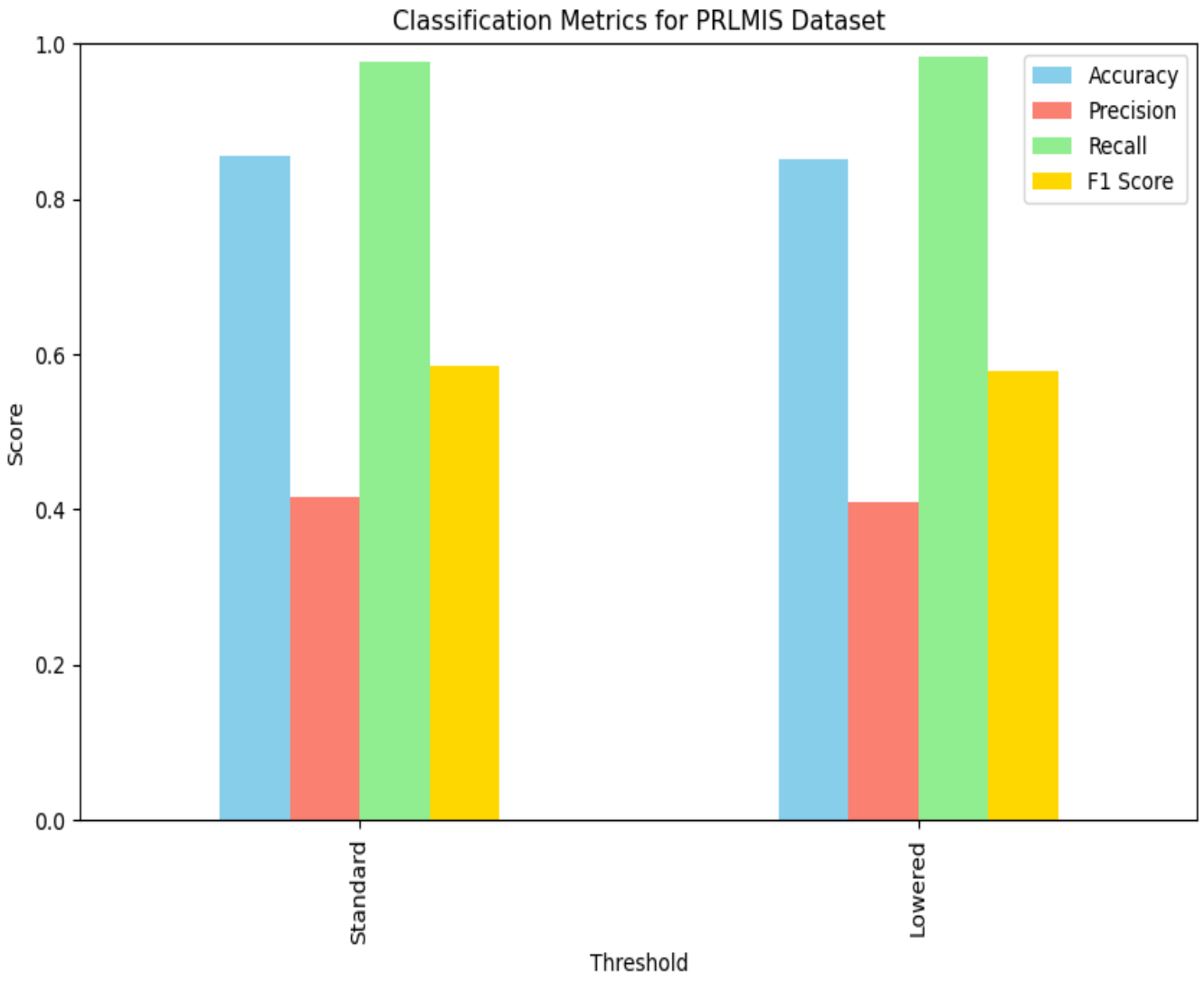
Modeling

- Use multiple methods to identify high-risk individuals effectively.
- Emphasize methods that can capture subtle patterns in data, making the identification of at-risk individuals more accurate and thorough.



Evaluation

- Recall: Measures how well high-risk individuals are identified.
- Precision: Indicates how often the "at-risk" label is accurate.
- Accuracy: The overall success rate of predictions.



Conclusion & Next Steps