



8/8/2025

Diabetes Risk Prediction

Evaluation Report



Kungulio, Seif H.
DATA 650: CAPSTONE PROJECT

Evaluating a Diabetes Risk Prediction Model for Ethical Decision-Making

Introduction

For public health agencies, understanding why a predictive model flags an individual as high risk for diabetes is just as important as how well it predicts. Interpretability builds trust with stakeholders like the Texas Department of State Health Services and supports transparent, equitable outreach strategies. Ethical evaluation ensures that predictive tools do not unintentionally introduce bias or exacerbate healthcare disparities. Together, these practices align machine learning with public health goals—improving early detection, reducing long-term costs, and advancing health equity.

Explain My Model Using Interpretation Tools

After establishing a baseline logistic regression model (accuracy = 0.82, recall = 0.56), a Random Forest classifier was selected for deployment due to its superior performance (accuracy = 0.85, recall = 0.66). Random Forest's ability to capture non-linear interactions among variables like BMI, general health, and chronic comorbidities provided a meaningful lift in sensitivity—critical for proactive outreach.

To interpret the model, SHAP (SHapley Additive Explanations) was applied. SHAP quantified the contribution of each feature to individual predictions, allowing both global and case-level insight.

Top features identified by SHAP:

- BMI – Higher BMI significantly increased predicted risk.
- General Health (GenHlth) – Poorer self-rated health strongly correlated with risk.
- Chronic Risk Load – Aggregated count of cardiovascular/metabolic conditions.
- Age group – Older categories carried higher risk.
- Healthcare Barrier Index – Access limitations were linked to higher predicted risk.

These drivers matched public health knowledge while also highlighting the importance of social determinants (e.g., healthcare access) alongside clinical risk.

Interpret the Results

SHAP analysis confirmed that BMI, general health, and chronic disease load dominate predictions, aligning with epidemiological literature. The inclusion of Healthcare Barrier Index as a top driver reinforced the link between access limitations and undiagnosed/untreated conditions.

An unexpected finding was the lower-than-anticipated influence of physical activity, despite its known protective role. This prompted consideration that self-reported exercise data may suffer from bias or over-reporting, diluting its predictive weight.

The model's emphasis on measurable clinical and access-related indicators, rather than solely behavioral variables, provides actionable levers for targeted interventions.

Evaluate Ethical Implications

A fairness check revealed potential concerns:

- Age Bias – Model recall was slightly lower in the youngest age bracket (18–34), suggesting possible under-identification of early-onset diabetes cases.
- Access Proxy Risk – The Healthcare Barrier Index may serve as a proxy for socioeconomic status. Automated use of this variable in outreach prioritization could unintentionally stigmatize or deprioritize certain groups.
- BMI Sensitivity – Heavy reliance on BMI could disadvantage individuals with atypical body composition (e.g., high muscle mass), though this is less prevalent at the population level.

Without mitigation, these biases could influence resource allocation in ways that reinforce existing health disparities.

Explore Ethical Guidelines and Standards

Evaluation followed the ISO/IEC 23894 AI Risk Management framework (ISO/IEC, 2023), ensuring structured consideration of transparency, fairness, and continuous monitoring. Additional reference points included:

- ACM Code of Ethics (ACM, 2018) – Mandates algorithmic accountability and harm minimization.
- Microsoft Responsible AI Principles (Microsoft, 2023) – Emphasizes stakeholder impact and bias mitigation.

From these, we implemented:

- Quarterly subgroup audits to monitor recall across demographic groups.
- Flagging of proxy variables (e.g., Healthcare Barrier Index) for careful review before operational use.
- Transparent documentation of model logic and top features for public health decision-makers.

Use AI to Review My Work

An AI review (ChatGPT) validated the model's public health alignment and raised several points:

- Reiterated recall priority in outreach contexts.
- Recommended exploring boosting algorithms (e.g., XGBoost, LightGBM) for possible gains without large interpretability loss.
- Suggested monitoring class imbalance effects and testing synthetic oversampling (SMOTE) in future iterations.
- Noted that monthly charges-style economic variables in other domains can mask geographic/socioeconomic bias—relevant here for healthcare access variables.

Summarize the AI Feedback

Key takeaways from AI review:

- Justify recall priority in stakeholder communication.
- Test advanced tree-based models for further performance lift.
- Monitor and adjust for class imbalance before major retraining cycles.
- Simplify visual explanations of SHAP results for non-technical audiences.

These insights have been incorporated into next-phase planning, with emphasis on balancing performance improvement with stakeholder trust.

Conclusion

The Random Forest model outperformed the logistic regression baseline in identifying Texans at high risk for diabetes, delivering improved recall without sacrificing generalization. SHAP-based interpretation clarified which features drive predictions, confirming domain expectations while revealing underweight factors like physical activity.

Ethical evaluation flagged potential age- and access-related biases, guiding mitigation steps such as subgroup monitoring and careful handling of proxy variables. AI-assisted review further strengthened the interpretability plan and identified future modeling opportunities.

The result is a more transparent, equitable, and deployment-ready model—capable of informing targeted, fair, and effective public health interventions.

References

- ACM. (2018). ACM Code of Ethics and Professional Conduct. <https://www.acm.org/code-of-ethics>
- ISO/IEC. (2023). ISO/IEC 23894: Artificial Intelligence—Guidance on Risk Management. International Organization for Standardization.
- Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30, 4765–4774. [https://proceedings.neurips.cc/paper_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf)
- Microsoft. (2023). Responsible AI Standard: Enabling Responsible AI Practices Across Microsoft. <https://www.microsoft.com/en-us/ai/responsible-ai>
- American Diabetes Association. (2023). The Economic Costs of Diabetes in the U.S. <https://www.diabetes.org/resources/statistics/cost-diabetes>
- Centers for Disease Control and Prevention. (2024). Behavioral Risk Factor Surveillance System Survey Data. <https://www.cdc.gov/brfss/index.html>