CONCLUSION:

In this project, addressed the challenges posed by class imbalance using the Synthetic Minority Over-sampling Technique (SMOTE). Given that the dataset exhibited a highly skewed class distribution, traditional training methods could lead to biased predictions, favoring the majority class. SMOTE was applied to generate synthetic samples for the minority class, thereby improving the model's ability to learn from underrepresented instances.

Since accuracy alone is not a reliable metric in imbalanced datasets, we also considered macro-average metrics, which provide a more balanced evaluation by giving equal weight to each class. After applying SMOTE, we observed a marginal decrease of 1% in accuracy for the CatBoost model compared to the version without SMOTE. However, there was a noticeable improvement in the macro-average metrics, ensuring fair and effective classification across all classes. Based on these findings, we conclude that CatBoost with SMOTE offers the best overall performance in handling class imbalance.

To further enhance model efficiency, feature selection was performed using Information Gain (IG), derived from Mutual Information (MI). IG evaluates the contribution of each feature towards reducing uncertainty in the target variable, helping identify the most relevant and informative predictors. The top 15 features were selected based on IG scores, ensuring that only the most meaningful attributes were retained.

By incorporating both SMOTE for class balancing and IG for feature selection, we achieved an optimal trade-off between accuracy and computational efficiency. The final CatBoost model, trained with the selected features and tuned hyperparameters (200 iterations, learning rate = 0.01), attained an accuracy of 84%. Although the macro-average score slightly decreased, the reduced feature set led to a lighter and computationally efficient model while maintaining strong predictive performance.

To enhance model interpretability, we applied SHAP (SHapley Additive Explanations) and LIME (Local Interpretable Model-agnostic Explanations):

- The SHAP waterfall plot illustrated how features influenced specific predictions. Features like BMI (-0.14), HighChol (-0.14), Smoker (-0.06), Veggies (-0.04), GenHlth (-0.04), and PhysActivity (-0.04) negatively impacted predictions, while DiffWalk (+0.05) had a slight positive effect.

- The SHAP summary bar chart highlighted global feature importance, identifying GenHlth (+0.08), HighChol (+0.07), and BMI (+0.07) as the most influential predictors, followed by HighBP (+0.06), Age (+0.05), and Smoker (+0.03).

- LIME was used for class-wise explainability, offering localized insights into individual predictions for different classes. LIME generated human-interpretable explanations by perturbing input data and observing its effect on model outputs. This helped examine

why certain minority class samples were misclassified and improved transparency in the model's decision-making process.

In summary, this project highlights the effectiveness of combining SMOTE (for handling imbalance) and Information Gain (for feature selection) to build a robust and efficient classification model. The final model strikes a balance between predictive performance and computational feasibility, making it a suitable choice for real-world applications where class imbalance is a major concern.

FUTURE SCOPE:

To further improve the model, the following advancements can be explored:

1. Advanced Models for Better Minority Class Representation

   o Exploring Deep Learning models with attention mechanisms for improved classification fairness.

2. More Robust Feature Selection Techniques

   o Applying Recursive Feature Elimination (RFE) to iteratively remove less relevant features.

   o Combining Mutual Information, L1-regularization, and Wrapper Methods for improved feature selection.

3. Improved Explainability with LIME for Class-Wise Interpretability

   o Using LIME to analyse misclassifications, helping to understand why certain minority class samples were incorrectly predicted.

   o Enhancing LIME visualizations to provide per-class interpretability, allowing domain experts to compare predictions across different groups.

4. Domain-Specific Feature Engineering

   o Incorporating domain knowledge to create new predictive features that improve classification accuracy.

   o Leveraging unsupervised learning techniques (e.g., clustering) to detect hidden patterns in minority class samples.

By adopting these improvements, the model can achieve higher fairness, better class-wise understanding, and improved overall prediction reliability.