Supervised Learning.

By: Jacob Cyr.

Project goals:

- -the goal of this project was to explore the data patterns within the dataset and explore the possible solutions to choosing a model.
- -Visualizations such as the correlation matrix allow for us to look into the relationships between predictor variables and the target variable.
- -I wanted to provide insights into the factors that contribute to diabetes.

The effect of glucose levels

- The analysis shows higher correlation between glucose levels and the target variable.
- -This information allows us to monitor glucose levels and prevent diabetes or mismanagement of diabetes.

BMI as an average

- We found that the average BMI for individuals with diabetes is higher than those without diabetes. This observation aligns with known risk factors.
- BMI may potentially be used as a screening tool for identifying people at risk for diabetes.

Age

- The dataset was collected from individuals around the age of 33 years.
- Healthcare professionals may want to develop future models for a diverse age population.

Ensemble model performance.

- Using the ensemble model worked as well or slightly better than the logistic regression.
- Ensemble models are good for situations where there may be more than one predictor affecting the outcome variable, this is valuable in complex healthcare scenarios.

Practical Application.

- The model can be used in a practical application by simulating populations by inputting relevant data.
- This means broader applications in healthcare are possible once a more robust model is developed.

Limitations

- The projects conclusions also acknowledge the limitations and potential for areas for future improvement. Due to computational constraints, certain algorithms like SVM and DBSCAN could not be explored. This suggests avenues for further research to explore additional modelling techniques.

Summary

- In this project, I used logistic regression and random forest models to predict diabetes based on analyzing a dataset. I found that factors like glucose levels and BMI are important for diabetes prediction. The models performed well, with both logistic regression and random forest showing good results. I learned about handling missing values, scaling data, and creating new features. The models can be used to predict diabetes for new individuals and groups. Overall, this project demonstrates the potential of machine learning in healthcare and highlights the need for further research and ethical considerations.

Photo credits: shutterstock.com