GTC ML Project 2 – Diabetes Prediction Model

Diabetes is a significant global health challenge where early detection can dramatically improve patient outcomes. Your task is to **build a predictive model** that can classify individuals as **diabetic** or **non-diabetic** based on diagnostic measurements. This project will take you from a pre-cleaned dataset through to a functioning predictive system, solidifying your understanding of the core machine learning workflow.

Phase 1: Become a Data Explorer!

Dive into the dataset and uncover the story within. Explore key questions:

- How many patients have diabetes versus those who don't?
- What's the relationship between glucose levels and the outcome?
- Does BMI play a significant role?

Use graphs, charts and summary statistics to uncover patterns and insights. This is your chance to be a data detective! We encourage you to search for resources on **"EDA for Classification"** to discover creative and effective visualization techniques.

Phase 2: Prep Your Data for Prime Time

Great models require great data. Prepare your dataset by:

- Standardizing your features to ensure all variables are on the same scale.
  Understanding why standardization matters—search for **"Why standardize data for ML?"** to learn more.
- Splitting your data into training and testing sets to ensure your model generalizes well to new, unseen patients.

Phase 3: Build, Train and Compete!

This is where the real fun begins! Choose your algorithms—will you use:

- Straightforward **Logistic Regression**?
- Powerful **Support Vector Machine (SVM)**?
- Robust **Random Forest**?

We challenge you to implement **at least two different models**. Go beyond the basics: search for **"Hyperparameter tuning with GridSearchCV"** to supercharge your model's performance. Train, tune and compare your models to see which one comes out on top!

Phase 4: Launch Your Prediction Engine!

Bring your model to life by building a prediction function that:

- Takes new patient data as input.
- Returns an instant prediction—**Diabetic** or **Non-Diabetic**.

This is your chance to create a real-world tool that demonstrates the power of ML in healthcare.

Good luck, and happy modeling!

We can't wait to see what you build.

GTC Team