TASK STATUS PROGRESS UPDATE

Task 1: Diabetes Dataset Acquisition ✅ **(Completed).**

- ☑ Government-accredited open-source healthcare dataset website.

- ☑ Column header of dataset ***Demographics:*** (Age & Pregnancies), ***Clinical Measurements:*** (Glucose, Blood Pressure, Skin Thickness, Insulin & Body Mass Index), ***Genetic Risk:*** (Diabetes Pedigree Function) and ***Target***: Outcome.

- ☑ Columns: 9 & Rows: 768.

Task 2: Diabetes Data Cleaning & Preprocessing ✅ **(Completed).**

Objective: Ensure the dataset is clean and ready for EDA.
Key Actions:

- ☑ Remove duplicates.

- ☑ Remove Empty Cells.

- ☑ Handle missing values.

- ☑ Standardize formats.

Task 3: Data Transformation ✅ **(Completed).**

Task 4: Loading Cleaned & Processed Version of Breast Cancer Dataset ✅ **(Completed).**

Task 5: Exploratory Data Analysis ✅ **(Completed).**

**Objective:** Implementation of EDA before data modeling.
Key Actions:

- ☑ Renaming column header, getting summary information of the dataset.

- ☑ Checking & removing any duplicate values and empty cells.

- ☑ Addition of a new column from an existing column in the data set.

- ☑ Converting Categorical values into Numerical values.

- ☑ Basic Visualization using matplotlib & seaborn python library.

- ☑ Resetting dataset index using Python to save output in Excel & CSV format.

- ☑ Modeling.

Metric	📈 Logistic Regression	⚔️ Support Machine Vector (SMV)	🌲 Decision Tree	🌲🌲🌲 Random Forest
Accuracy	70.78%	70.78%	35.06%	✅71.43%
Precision	✅59.14%	56.92%	35.06%	57.35%
Recall	66.67%	68.52%	✅100%	72.22%
F1-Score	61.54%	62.18%	51.92%	✅63.93%

Task 6: Data Modeling ✅ **(Completed).**

Objective: Choosing the best-fit algorithm or model for the dataset.

Key Actions:

- ☑ **Evaluate different algorithms** (📈 Logistic Regression, ⚔️ Support Machine Vector (SVM), 🌲 Decision Tree & 🌲🌲🌲 Random Forest) — in testing phase.

- ☑ **Select and justify the best model** based on dataset characteristics.

- ☑ Model Training.

- ☑ **Data Engineering** & **Model Implementation** after selecting the best-fit model.

- ☑ Model Deployment Using Streamlit.

Task 7: Data Visualization & Communication ✅ **(Completed).**

- ☑ Power BI Visualization Report.
- ☑ PDF Visualization Report.
- ☑ Image Visualization Report.

Task 8: Project Review ✅ **(Completed).**

Task 9: Project Upload to GitHub ✅ **(Completed).**

Task 10: Project Task Report Submission ✅ **(Completed).**

Developer: Abduljabbar Nuhu

📞 **Phone:** 09036259681.

📧 Email: nuhuabduljabbar5@gmail.com

🎗 GitHub: *Completed* 💯