📋TASK STATUS PROGRESS UPDATE

Task 1: Cervical Cancer Dataset Acquisition ✅ (Completed).

- ☑ Government-accredited open-source healthcare dataset website.
- ☑ Column header of dataset ***Demographics:*** (Age, Number of Sexual Partners, First Sexual Intercourse, Num of Pregnancies), ***Lifestyle & Risk Factors:*** (Smokes, Smokes (years), Smokes (packs/year), Hormonal Contraceptives, Hormonal Contraceptives (years), IUD, IUD (years)), ***Sexually Transmitted Disease (STDs):*** (STDs, STDs (number), STDs: Condylomatosis, STDs: Vaginal Condylomatosis, STDs: Vulvo-Perineal Condylomatosis, STDs: Syphilis, STDs: Pelvic Inflammatory Disease, STDs: Genital Herpes, STDs: Molluscum Contagiosum, STDs: HIV, STDs: Hepatitis B, STDs: HPV, STDs: Number of Diagnosis, STDs: Time Since First Diagnosis, STDs: Time Since Last Diagnosis) and **Target:** Diagnosis/Outcome (Dx: Cancer, Dx: HPV, Dx, Hanselmann, Schiller, Cytology, Biopsy).
- ☑ Columns: 33 & Rows: 61.

Task 2: Cervical Cancer 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).

- ☑ Encoded categorical variables.
- ☑ Normalized/Standardized numerical features.
- ☑ Ensured dataset consistency.

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

- ☑ Final dataset saved in Excel & CSV format.
- ☑ Ready for EDA and modeling.

Task 5: Exploratory Data Analysis ⏳(In Progress).

Metric	📉 Logistic Regression	⚖️ Support Machine Vector (SMV)	🌲 Decision Tree	🌲🌲🌲 Random Forest
Accuracy	00.00%	00.00%	00.00%	00.00%
Precision	00.00%	00.00%	00.00%	00.00%
Recall	00.00%	00.00%	00.00%	00.00%
F1-Score	00.00%	00.00%	00.00%	00.00%

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.

Task 6: Data Modeling ❌ (Pending…)
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 ❌ (Pending…)

- ☑ Power BI Visualization Report.

- ☑ PDF Visualization Report.
- ☐ Image Visualization Report.

Task 8: Project Review ❌ (Pending…)

Task 9: Project Upload to GitHub ❌ (Pending…)

Task 10: Project Task Report Submission ❌ (Pending…)

Developer: Abduljabbar Nuhu

Phone: 09036259681.

Email: nuhuabduljabbar5@gmail.com

GitHub: ***Processing…***