Final Report: Machine Learning & AI Tasks

# Introduction

This report summarizes the key results, methods, and insights from a series of machine learning and AI tasks, including data analysis, prediction, and building a simple medical Q&A bot. Each task is briefly described below, highlighting the main findings and takeaways.

# Task 1: Iris Dataset EDA

* Goal: Explore the classic Iris dataset to understand feature distributions and relationships.
* What Was Done:
* - Loaded and summarized the data (150 samples, 4 features, 3 species).
* - Visualized feature distributions and relationships (scatter plots, histograms).
* Key Insights:
* - Petal length and width are the most useful features for distinguishing species, especially setosa.
* - Sepal measurements overlap more between species.
* - The dataset is clean and well-structured, making it ideal for introductory analysis.

# Task 2: Apple Stock Price Prediction

* Goal: Predict Apple's next-day closing stock price using historical data.
* What Was Done:
* - Downloaded daily AAPL stock prices (2022–2023).
* - Built and compared Linear Regression and Random Forest models.
* Key Insights:
* - Random Forest outperformed Linear Regression, achieving a lower Mean Squared Error (MSE ~8.15).
* - The model's predictions were reasonably close to actual prices, but further tuning could improve results.

# Task 3: Heart Disease Prediction

* Goal: Predict the presence of heart disease from patient data.
* What Was Done:
* - Used the UCI Heart Disease dataset (920 samples, 16 features).
* - Preprocessed data (handled missing values, encoded categories).
* - Built and evaluated classification models.
* Key Insights:
* - Models achieved reasonable accuracy (details not shown), with confusion matrices and classification reports used for evaluation.
* - Feature engineering and model selection are crucial for medical prediction tasks.

# Task 4: Medical Q&A Bot with Gemini API

* Goal: Build a simple medical question-answering assistant using the free Gemini API.
* What Was Done:
* - Set up the Gemini SDK and API key.
* - Created a function to send user health questions to the Gemini model, with a prompt for safe, friendly answers.
* - Tested the bot with sample questions (sore throat, paracetamol for children, antibiotics, insomnia).
* Key Insights:
* - The bot provides clear, safe, and friendly health information, always recommending seeing a doctor for serious issues.
* - Limitations include lack of personalized advice and the free model's capabilities.
* - Useful for general health awareness, but not a substitute for professional care.

# Task 6: Ames Housing Price Prediction

* Goal: Predict house sale prices using the Ames Housing dataset.
* What Was Done:
* - Cleaned and preprocessed the data (dropped columns with many missing values, one-hot encoding).
* - Compared Linear Regression and Gradient Boosting models.
* Key Insights:
* - Gradient Boosting performed much better than Linear Regression (MAE ~$17,500, RMSE ~$32,360).
* - Good data cleaning and model choice are essential for accurate predictions.