SIADS 699 Team 16
Numan Suri, Hamza Asad, Asad Kamal

Team Gemini

GitHub:
https://github.com/numansuri/Capstone-MADS

Project Overview

Goal
Evaluate whether large language models (LLMs) like Gemini can assist with feature selection and model building — and how they compare to traditional statistical methods — using heart disease prediction as a case study.

Data Source
National Health Interview Survey (NHIS) from the CDC, including health, demographic, and behavioral data from a representative sample of U.S. adults.

Target Variable
heart_disease (binary) — derived from self-reported:
- Congestive heart failure (CHDEV_A)
- Coronary heart disease (ANGEV_A)
- Myocardial infarction (MIEV_A)

Why This Matters
LLMs are increasingly part of the data science toolkit — but should they be? This project investigates whether incorporating LLMs improves or hinders model development compared to a traditional, methodical approach.

Machine Learning Pipeline

Data Cleaning & Mapping
Categorical codes were translated into meaningful labels (e.g., education, race, insurance coverage).

Feature Engineering
Created a binary heart_disease target variable based on three cardiovascular conditions: CHDEV_A, ANGIV_A, MIEV_A.

One-Hot Encoding
Transformed all categorical features into binary indicators using ColumnTransformer.

Feature Selection
Applied chi-square tests and t-tests; retained only features with $p < 0.05$.

Train-Test Split
Split the processed data into training and test sets with stratification on the target.

Models
Trained and evaluated:
- Logistic Regression
- Random Forest
- XGBoost
- CatBoost

Results

Traditional

Logistic Regression

	precision	recall	f1-score
0	0.97	0.74	0.84
1	0.21	0.77	0.33
accuracy			0.74
macro avg	0.59	0.76	0.59
weighted avg	0.91	0.74	0.80

Random Forest

	precision	recall	f1-score
0	0.96	0.79	0.87
1	0.22	0.67	0.34
accuracy			0.78
macro avg	0.59	0.73	0.60
weighted avg	0.90	0.78	0.83

XGB Classification

	precision	recall	f1-score
0	0.97	0.77	0.86
1	0.22	0.72	0.34
accuracy			0.77
macro avg	0.59	0.75	0.60
weighted avg	0.91	0.77	0.82

Gemini 2.5

Logistic Regression

	precision	recall	f1-score
0	0.977	0.747	0.847
1	0.221	0.803	0.346
accuracy	0.752	0.752	0.752
macro avg	0.599	0.775	0.597
weighted avg	0.915	0.752	0.806

Random Forest

	precision	recall	f1-score
0	0.950	0.911	0.930
1	0.314	0.460	0.373
accuracy	0.874	0.874	0.874
macro avg	0.632	0.685	0.651
weighted avg	0.898	0.874	0.884

XGB Classification

	precision	recall	f1-score
0	0.972	0.773	0.862
1	0.228	0.754	0.351
accuracy	0.772	0.772	0.772
macro avg	0.600	0.763	0.606
weighted avg	0.912	0.772	0.820