COMP 560

# Understanding How Health and Environment Influence GDP

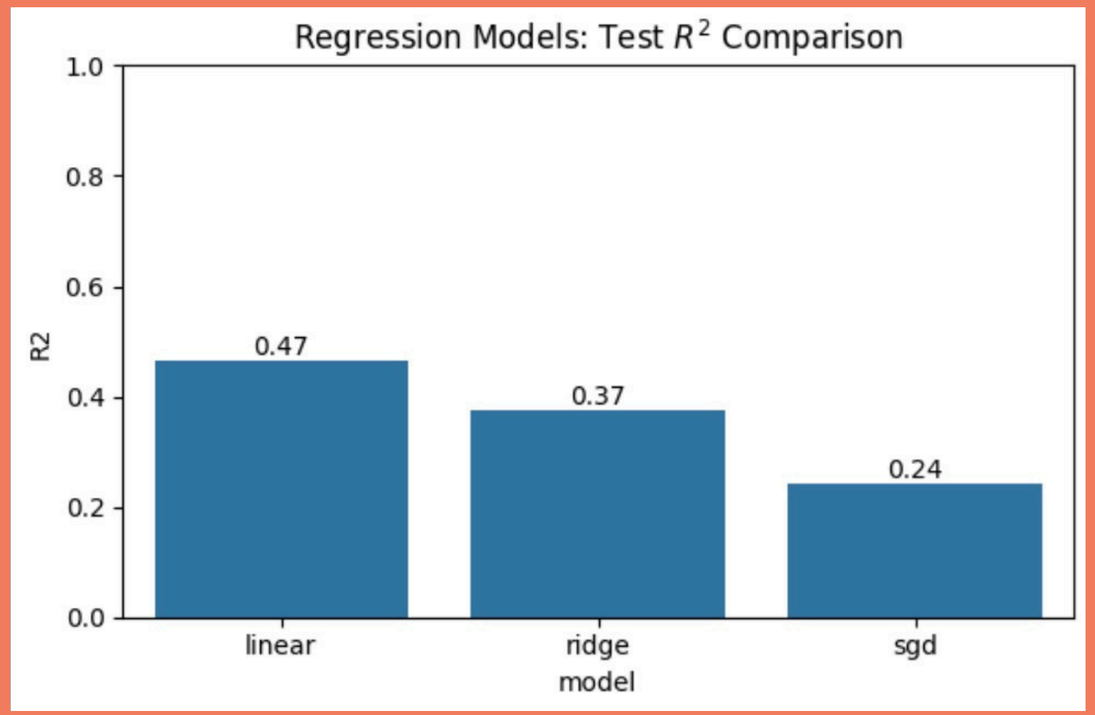
Siddhant Saxena, Justin Li,
Joanna Jipson, Nhu Le

COMP 560

# Our Focus In This Project

- We examine how environmental risk factors and clinical health statistics relate to a country’s economic performance (GDP)
- We apply multiple machine learning methods and use the Anthropic API to generate structured interpretations of patterns and correlations
- We analyzed a consolidated CSV using regression models and stochastic gradient descent based models

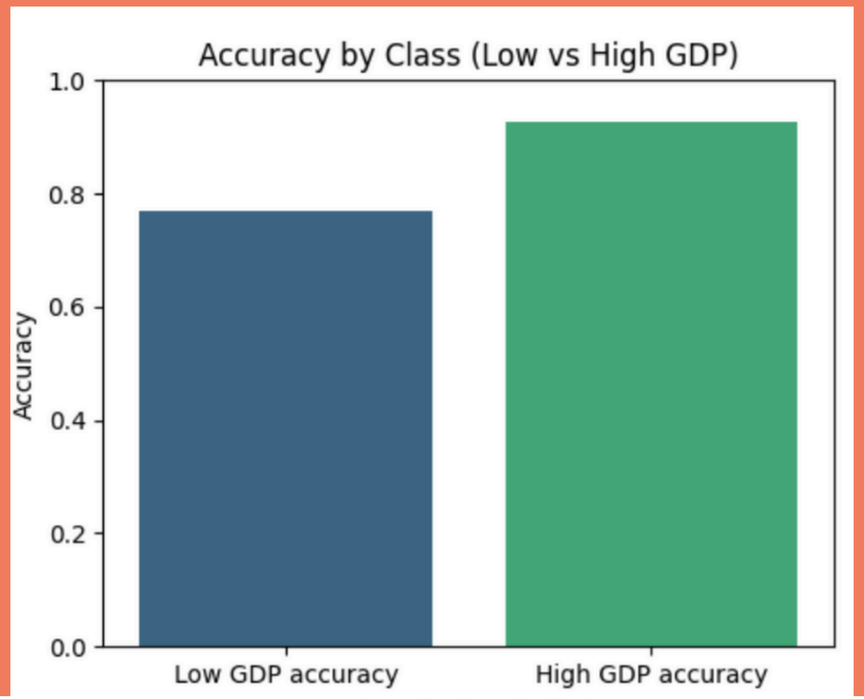
COMP 560

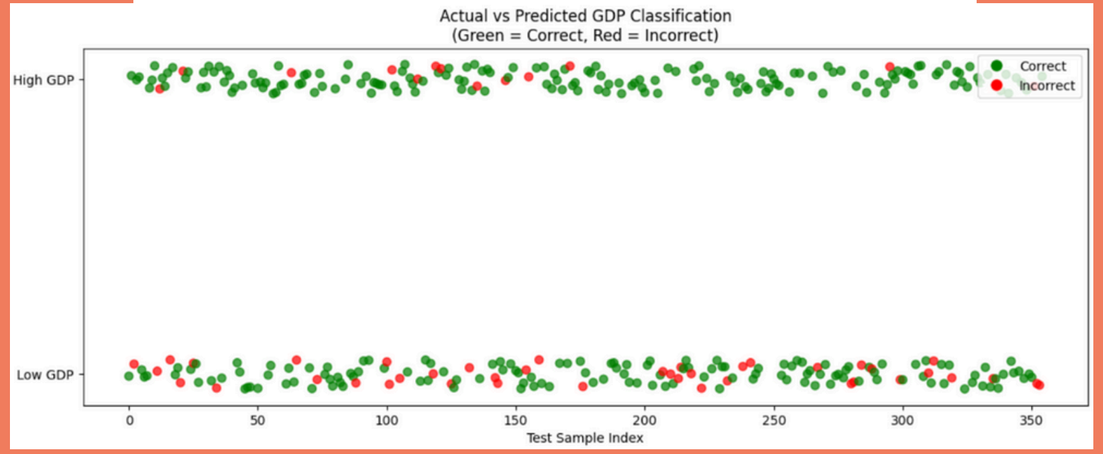



# Data & Methods

- Combined data from 3 sources, including a Kaggle dataset from WHO records on cause-specific death counts (1990-2017)
- Implemented three models to estimate GDP from environmental and clinical risk factors
  - Linear regression optimized using stochastic gradient descent (SGD)
  - Logistic regression optimized using logistic loss & gradient descent
  - Ridge regression model was also used
- Anthropic Claude API for interpretation
  - Used prompt engineering to generate summaries of model results
  - Submitted model outputs (JSON) to Claude to produce explanations aligned with our analysis goals

COMP 560




# Results

- Regression Models:
  - Linear, ridge, and SGD regressions showed moderate performance, with linear regression achieving the highest $R^2$
- Classification Models:
  - Logistic regression (liblinear & GD solvers) achieved high accuracy and strong F1-scores, & can distinguish high vs. low GDP countries
- Models that used both environmental and clinical features together performed better than those using either group alone, suggesting both factor types are important for GDP prediction
- We found that ML models capture relationships between risk factors and GDP, but more advanced methods are needed to dig deeper

COMP 560

# Thank You!