Augenblick 2025

ML1-CI: *Causal Inference Challenge*

Team: *PixelPair*

Overview

Understanding causality in healthcare is crucial for making informed medical decisions. This project implements a causal inference framework to analyze treatment effects using the **Infant Health and Development Program (IHDP)** dataset. We developed an **interactive dashboard** using **Streamlit**, integrating causal inference techniques to estimate **Average Treatment Effects (ATE)** and **Conditional Average Treatment Effects (CATE)**. The platform allows users to explore treatment outcomes, visualize confounder balance, and perform "what-if" simulations.

Day 1:

Methodology

1. Dataset & Preprocessing

- **Data Source:** IHDP dataset (synthetic fallback included).
- **Preprocessing:**
  - Ensured treatment variable is numeric.
  - Identified and managed confounders.
  - Generated counterfactual outcomes for evaluation.
  - Implemented **covariate balance checking** to assess confounder distribution across treatment groups.

2. Causal Inference Techniques

We implemented two methods to estimate treatment effects:

- **Propensity Score Matching (PSM):** Ensures comparable treatment/control groups.
- **Inverse Probability Weighting (IPW):** Adjusts for confounding bias using weighted estimates.
- **Double Machine Learning (DML):** Utilizes flexible machine learning models to estimate treatment effects while controlling for confounders.

Additionally, we trained **RandomForestRegressor models** separately for treated and control groups to predict factual and counterfactual outcomes.

3. Model Training & Validation

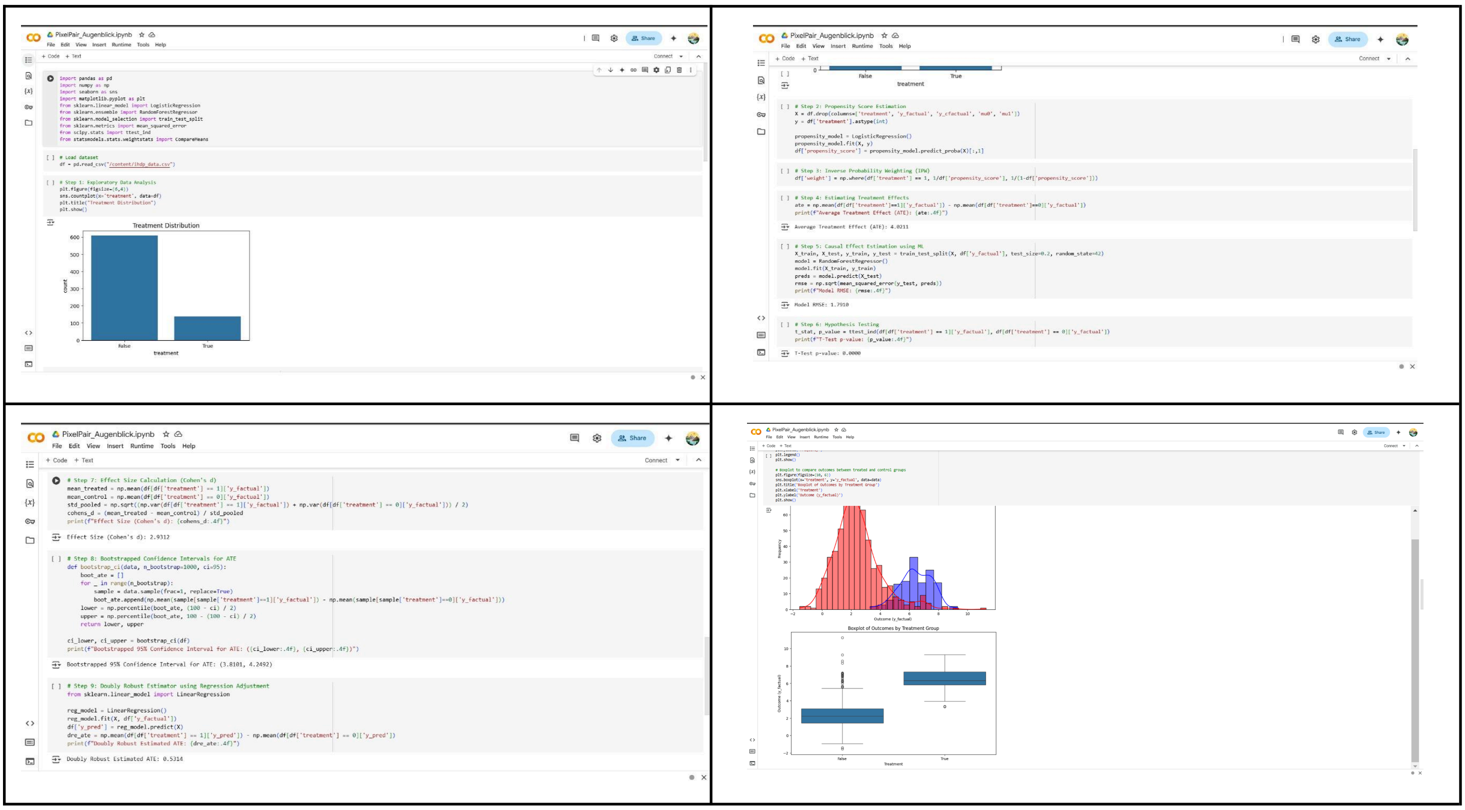
- **Feature Engineering:** Used **income, birth weight, parental education, health index, housing quality, and neighborhood safety** as key confounders.
- **Training Pipeline:**
    - Normalized features using **StandardScaler**.
    - Built separate models for **treated (T=1) and control (T=0)** groups.
    - Evaluated models using **R^2 score and Mean Squared Error (MSE)**.
- **Validation:** Verified against synthetic counterfactual outcomes to measure accuracy.

4. Estimation of Treatment Effects

- **ATE Calculation:** Difference in mean outcomes between treatment and control groups.
- **CATE Calculation:** Estimates variation in treatment effects across subgroups.
- **Individual Treatment Effect (ITE):** Provides personalized estimates of treatment benefits for each observation.

Google Collab Link :
https://colab.research.google.com/drive/1Lv5zv8IyswpBgpLSqSGFMfj8bbJ_BH0b?usp=sharing

https://colab.research.google.com/drive/1VEEtOli-_Xr-rVq1tmGTaoazB_UIVogG?usp=sharing



Day 2:

Interactive Dashboard Features

1. Input & Results Page

- **User Input Sliders** for modifying confounder values.
- **Real-time predictions** of factual and counterfactual outcomes.
- **Visualization of estimated individual treatment effects.**

2. Dataset Overview

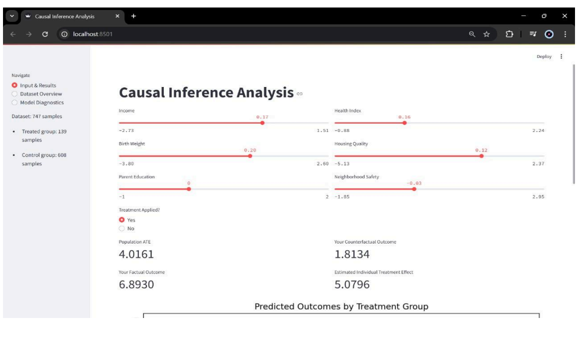
- **Summary statistics and sample data preview.**
- **Confounder balance visualization** between treated and control groups.
- **Histograms showing confounder distributions by treatment.**

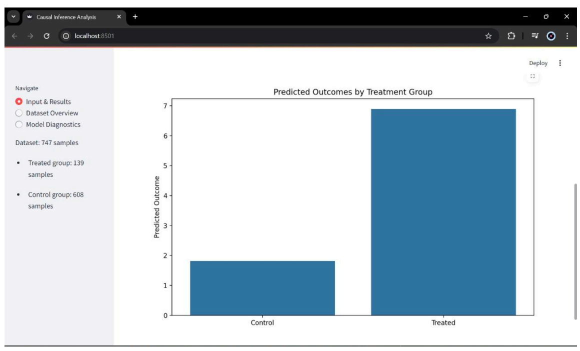
3. Model Diagnostics

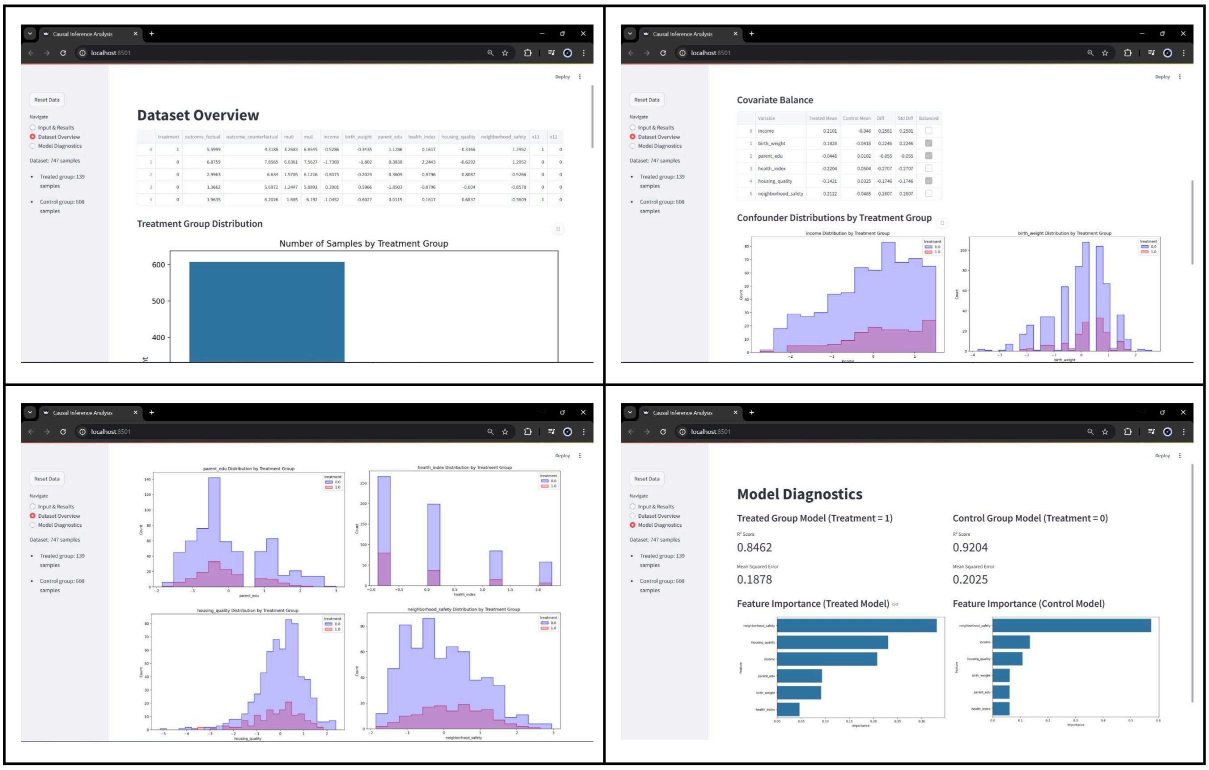
- **Performance metrics (R^2, MSE) for treated and control models.**
- **Feature importance visualizations** for both models.
- **Comparison of predicted outcomes across treatment groups.**

Key Findings & Insights

- **Treatment Effects:** Estimated **ATE suggests a positive impact of specialized child care on cognitive test scores.**
- **Confounder Balancing:** Some confounders, such as **parental education and income**, show imbalance, requiring advanced adjustments.
- **Model Performance:** RandomForest models provide reliable predictions, with feature importance aligning with expected causal relationships.
- **Heterogeneous Effects:** CATE analysis reveals that treatment effects vary across different subgroups, emphasizing the need for personalized intervention strategies.







Conclusion

This project successfully implements a **robust causal inference pipeline** with an **interactive application** for analyzing healthcare interventions. The dashboard provides an intuitive way to explore treatment effects, confounder balance, and perform scenario-based simulations. Our approach ensures **statistical rigor, interpretability, and practical usability** in making causal claims from observational data.

Deliverables

- **Streamlit Dashboard:** Interactive platform for exploring causal relationships.
- **Colab Notebook :** Detailed documentation of data processing and model implementation.
- **Final Presentation:** Summary of key insights and methodology.

This work demonstrates **advanced causal inference techniques**, model robustness, and an engaging **user-friendly interface**—making it a valuable tool for healthcare analysis.