Capstone 2 Project Documentation

Problem Statement

The United States is currently undergoing a drug shortage, causing delays and discontinuations of treatment [1]. The U.S. healthcare system is experiencing a strain on resources due to an aging population, and increased drug usage correlates with age [2, 3]. A model to predict drug demand based on population can forecast demand for future generations to mitigate supply chain bottlenecks.

Results

Machine learning model was built using Bayesian Ridge Regression with an accuracy of 3.18%. Using untested data from the same source, the model predicts 5940 prescriptions dispensed in 2023.

Constraints

Model predicts drug demand based off population statistics without consideration of external factors related to drug production, such as physical capacity of manufacturing plants and economic viability.
Additional data is advised to be collected due to shifts in data collection methods and exclusion of data to align with the scale of other data sources limiting the data used to train the model.

Data Sources

- Prescriptions dispensed 2009-2022 https://datasetsearch.research.google.com/search?src=0&query=prescription%20data&docid=L2cvMTFwd2JjdjlmYw%3D%3D
- Prescription usage with age https://www.statista.com/statistics/184442/us-population-with-usage-of-prescription-drugs-by-age/
- Population data
- https://www.statista.com/statistics/241488/population-of-the-us-by-sex-and-age/

References

---

1. https://www.ama-assn.org/delivering-care/public-health/drug-shortages#
2. https://www.statista.com/statistics/184442/us-population-with-usage-of-prescription-drugs-by-age/
3. https://www.census.gov/library/stories/2023/05/2020-census-united-states-older-population-grew.html