Hi, this is my presentation for the NORP Structured Project: Impact of Healthcare Policies on California Counties.

First, I will introduce the Nonprofit Organization Research Panel as a project designed to collect and visualize nonprofit datasets, promoting nonprofit research.

This platform provides a convenient space for sharing datasets and research in a collaborative environment. Moreover, this project is planned to eventually be moved into the research database as another worthwhile research project.

Regardless, the project’s focus is on the reality that healthcare is a significant concern for Americans. Thus, research into exploring the effects of various healthcare policies is essential to both identify and enact beneficial healthcare laws to restore the public’s faith in the American healthcare system.

Likewise, California has passed several sweeping laws to transform the healthcare industry in California, shifting its residents from being uninsurance to its Medicaid program. For example, the law analyzed in this project is a Medi-Cal expansion for all income-eligible young adults, regardless of immigration status, enacted in 2019.

Moreover, the motivation for the project is that there is a significant increase in innovations in the artificial intelligence field, which introduces new possibilities into policy analysis.

For example, models like an Autoregressive integrated moving average or ARIMA model can be used to perform an interrupted time series analysis to determine said impact, where an interrupted time series analysis is a study on the difference between the data trends before and after an intervention date. Thus, this new technique can capture trends that traditional statistical analysis methods have not been able to identify. Where said information could have a significant impact on the healthcare industry, through rapid policy changes.

Moreover, related research papers have explored the use of AI models to determine the effect of certain events. For example, there was a research paper titled “Current Use And Evaluation Of Artificial Intelligence And Predictive Models In US Hospitals” that shows the implementation of AI models in the prediction of patient outcomes as a result of the deterioration of their health, and the use of AI models to suggest treatment options to ensure the best outcome. Thus, this demonstrates how AI models are already being utilized to make predictions in the healthcare field.

Nevertheless, the general overview of the procedure taken in creating the datasets and frontend interactive diagram includes first gathering the relevant datasets from organizations like the California Health and Human Services through their data portal. These datasets primarily contained high-quality information; however, given the specific needs of this project, a thorough cleaning of the datasets was necessary. Consequently, the much smaller and easier-to-manage datasets were then manipulated into a specific table format, indicating their value, the current year, whether the intervention occurred, and how long the intervention lasted. This table was passed through an ARIMA model to obtain an estimate of the predicted values, and these values were then passed through a Gaussian Processes model to estimate values for the future. The data frame was then converted to CSV files and subsequently to JSON files. Then I created the frontend application using the Next.js framework with React components and React-based packages. Finally, I created unit tests using Jest and end-to-end tests using Playwright to ensure basic functionality of the web application. Then, I ran the original and cleaned datasets through a data visualization notebook to visualize the differences in size and number of data values between the datasets.

Then, for data cleaning, I would use the Pandas library to replace hospital values with county values, creating relationships between the county and the statistics in question. The next step would be to split the values with multiple counties in the same data frame into separate rows. Then, I would group all relevant identifying information, such as year and county, to sum all statistics, and then convert the data frame to cleaned CSV files.

Furthermore, the implementation would be based on using the ARIMA model from the statsmodels library to predict the impact of policies and create a line chart to indicate the model prediction values. Then use the Gaussian Process model to estimate the future ten years using the sklearn package.

Moreover, an example of the resulting dataset is the ARIMA and Gaussian Process models, which created a 96.44% decrease in the number of deaths from 2018 to 2019. The models helped demonstrate how the training and testing of the various datasets contributed to the creation of accurate datasets. The testing and training ratio was selected to be 0.5, as the limited amount of data necessitated a decrease from the default value of 0.7 to 0.5. Then, I created 100 predictions for the prior samples using the Gaussian Process, which was a good fit for the y1 line, leading to the conclusion that the prior is roughly credible, and the rest of the model’s execution can proceed. Thus, I also created 10 predictions using the Gaussian Process model.

In addition, data visualization was contained in a notebook that utilized the Pandas library to remove all missing values, allowing for the creation of useful tables and charts. Likewise, it was also used to determine the total file size and total number of data values for each dataset, allowing for comparison between the unclean and clean datasets to illustrate the effects of data cleaning. Likewise, the resulting difference is displayed in several bar charts to facilitate a more straightforward comparison.

Likewise, the skills learned include using QGIS software to edit GeoJSON files for integration into the Next.js web app. Then, learning the Redux JavaScript library to create a global store for the Next.js web app. Then, learning the Jest and Playwright testing frameworks to test the next.js based web app.

Similarly, I learned about the Flexibility of Ideas, where the switch from the Retrieval Augmented Generation model to a Time Series Model with Linear Regression, and then to an Autoregressive Integrated Moving Average model with a Gaussian Processes model, demonstrates the importance of improving existing ideas. I learned about Application Pull by researching the increasing use of artificial intelligence to make strategic decisions, as the leaders wanted more accurate models to become better at decision-making. I learned about Technology Push as the introduction of generative artificial intelligence in the healthcare industry was led by developers pushing for the advancement of technology and encouraging hospitals to adopt it.

Moreover, these are the deliverables I need to complete for the project, which include gathering datasets and learning about AI models. Cleaning datasets for use in AI models and creating Figma designs for website layout. Creating an AI Model implementation, finishing said AI Model implementation, and adding instructions for its use to the README file.

Then, I needed to code the front end of the Next.js website’s interactive diagram and integrate the AI model with the website diagram.

Then, I needed to create unit tests to test the model and end-to-end integration tests to ensure that the integration between the datasets and the frontend map was working. Then I needed to create a video presentation of the project and complete final end-to-end integration testing.

Thanks for watching my presentation.

References

Nong, Paige, et al. “Current use and evaluation of artificial intelligence and predictive models in US hospitals.” *Health Affairs*, vol. 44, no. 1, 1 Jan. 2025, pp. 90–98, https://doi.org/10.1377/hlthaff.2024.00842.

Arruda, Flavio Regis de. “Interrupted Time Series (ITS) in Python.” *Xboard.Dev*, 1 Jan. 2022, www.xboard.dev/posts/2020\_01\_01\_interrupted-time-series-python-part-I/index.html.

“Pydata Berlin 2019: Gaussian Processes for Time Series Forecasting (Scikit-Learn).” *Dr. Juan Camilo Orduz*, 10 Oct. 2019, juanitorduz.github.io/gaussian\_process\_time\_series/.

Santamato, Vito, et al. “Exploring the impact of artificial intelligence on Healthcare Management: A combined systematic review and machine-learning approach.” *Applied Sciences*, vol. 14, no. 22, 6 Nov. 2024, p. 10144, https://doi.org/10.3390/app142210144.

Chen, Yan, and Pouyan Esmaeilzadeh. *Generative AI in Medical Practice: In-Depth Exploration of Privacy and Security Challenges (Preprint)*, 22 Sept. 2023, https://doi.org/10.2196/preprints.53008.