Capstone Project Proposal

Predicting mortality in the Emergency Department utilizing Deep Learning Methods

Student: Porfirio Basaldúa González

Domain background

The time is ripe for a technological revolution in healthcare. According to Mckinsey, over the last decade pharmaceutical companies have been compiling research and development data into datasets, while “the US federal government and other public stakeholders have been opening their vast stores of health-care knowledge, including data from clinical trials and information on patients covered under public insurance programs” (Mckinsey & Company, 2013)
All this new influx of information has already been applied by the different stakeholders to create savings and improve processes, for example “Kaiser Permanente has fully implemented a new computer system [...] and achieved an estimated $1 billion in savings from reduced office visits and lab tests” (Mckinsey & Company, 2013). The same stakeholders acknowledge that there are still many opportunities for innovation and improvement within the healthcare domain, Mckinsey & Company estimate that “since 2010, more than 200 new businesses have developed innovative health-care applications. About 40 percent of these were aimed at direct health interventions or predictive capabilities.”

Of these possible innovations, predictive capabilities are of special interest to health care professionals, since it will allow them to “interrupt the patient’s [projected] trajectory and set him on the proper course” (Adamson, D., 2015, p. 30) this capability is known as Prescriptive Analytics, and is considered a “Holy Grail” (Adamson, D., 2015, p. 30) within the healthcare domain.

This project will aim to create such a predictive solution, focusing on anticipating the mortality in Emergency patients, taking in consideration the following restrictions:
- Since one of the main impediments to use advanced technologies (i.e. Machine Learning) is the lack of expertise (Adamson, D., 2015, p. 10), the proposed solution has to be easy to use.
- The solution has to demonstrate sufficient capabilities to handle large amounts of data.
- The resulting predictions have to be available in real time, as they will aid assessments in emergencies, where time is vital.

Problem statement

Can patient healthcare records predict the likelihood of mortality in an emergency room visit?

Solution statement

Deep Learning models combined with large amounts of data are well suited for prediction problems, and cloud infrastructure will ensure the system has enough resources to handle big-data concerns effectively. With this, the proposed solution is as follows:

Using cloud infrastructure, the researcher will create a data pipeline that will expose an API through the internet which will allow the healthcare professionals to access real time predictions from a machine learning model. This predictions will assess the patient's probability of dying while in emergency care, giving health care professionals a tool for decision making, shortening the time needed for the patient to receive adequate care.

Datasets

This project will use the Medical Information Mart for Intensive Care (MIMIC) dataset. According to Johnson, A. et al., this dataset "is a large, freely-available database comprising de-identified health-related data from patients who were admitted to the critical care units of the Beth Israel Deaconess Medical Center." It contains data from 2008 to 2019 collected from Metavision bedside monitors and is arranged in different relational tables.

Even though the data is de-identified, access to the data is conditioned to the following requisites:
1. Completing a training course in human subjects research.
2. Signing the data use agreement. Adherence to the terms of this agreement is paramount.
3. Accessing the data directly in the cloud.

Benchmark model

Li C. et al. (2021) have proposed a LightGBM model with an accuracy of 93.6% for early mortality predictions in emergency departments, this will be used as a benchmark.

Evaluation metrics

The evaluation metric will be the same used in the benchmark model, which is accuracy.

Project design

The project will be segmented in the following steps:

1. Complete the mandatory training to access and explore the dataset.
2. Set an ELT workload in a multicloud environment. Since the data is hosted in GCP, the researcher will set up a multicloud environment to extract the data from GCP into AWS, where the processing of the information will take place.
3. Hyperparameter tuning.
4. Multi instance training of the model.
5. Endpoint provisioning and API exposure.
6. Evaluate the results.
7. Improvements and future work. Create a UI that will enable end users to apply the predictions in real situations, Preferably, this UI will be directly integrated into existing systems in the ED departments of the hospitals. Alternatively, a new standalone application can be built and leased for use in ED departments across hospitals.

References

Adamson, D. (2015, August 13). Big Data in Healthcare Made Simple: Where It Stands Today and Where It's Going. Retrieved March 12, 2023, from https://www.slideshare.net/healthcatalyst1/big-data-in-healthcare-made-simple-where-it-stands-today-and-where-its-going

Johnson, A., Bulgarelli, L., Pollard, T., Horng, S., Celi, L. A., & Mark, R. (2021). MIMIC-IV (version 1.0). PhysioNet. https://doi.org/10.13026/s6n6-xd98.

Kayyali, B., Knott, D., & Kuiken, S. V. (2013, April 1). *The Big-Data Revolution in US health care: Accelerating value and innovation*. McKinsey & Company. Retrieved March 12, 2023, from https://www.mckinsey.com/industries/healthcare/our-insights/the-big-data-revolution-in-us-health-care#/

Li, C., Zhang, Z., Ren, Y., Nie, H., Lei, Y., Qiu, H., Xu, Z., & Pu, X. (2021). Machine learning based early mortality prediction in the emergency department. *International Journal of Medical Informatics*, *155*, 104570. https://doi.org/10.1016/j.ijmedinf.2021.104570