

Udacity Machine Learning Nanodegree 2021

Capstone Proposal

Predicting hospital bed demand in Australia with Machine Learning

Ryan Ashton

Domain Background

Australia has one of the lowest population densities per land area in the world. With most of the population residing towards the East in cities such as Sydney, Melbourne and Brisbane. Australia is seen as a great place to migrate to as it is a rich country with many opportunities. USNews.com suggests that Australia is the 4th best country to migrate to (USnews.com, 2017)

Because Australia is still rapidly developing, the population needs to “spread out” across the vast land to avoid issues that comes with high population densities. With Australia forecasted to have rapid population growth coupled with the growing trend of urbanisation, the outer suburban population is projected to double or even triple in Australia (Archer J, et al., 2019). Because of this, the healthcare sector needs to predict how many hospitals and hospital beds are required in new developing urban areas.

In 1990 a paper called “Models for forecasting hospital bed requirements in the acute sector” suggested that simple “trend fitting” is not sufficient to predict how many hospital beds are required, meaning, a more sophisticated approach is required (R D Farmer 1, J Emami, 1990).

Ravaghi, H., Alidoost, S., Mannion, R. et al in 2020 found that good indicators for determining the number of hospital beds required are demographic changes, average length of stay, admission rates, and bed occupancy rates. This project will focus primarily on the demographic changes.

A lot of the demographic data is publicly available in Australia but not all in the one place. A lot of data collection and pre-processing will be required to construct the main dataset used to build this machine learning model.

Problem Statement

Australia is growing in population and therefore needs to create areas for people to live. With new developing urban areas, comes a need to understand how much healthcare demand there will be - specifically how many hospital beds.

Datasets and Inputs

The dataset is aggregated by me via many joins on public data provided by different Australian Government agencies. The hospital related data is provided by the Australian Institute of Health and Welfare (AIHW), whereas the demographic data was sourced from the Australian Bureau of Statistics (ABS). I linked the various datasets together via a postcode and Statistical Area Level 2 (SA2) mapping file that I requested (and received) from the Australian Government.

Hospital bed data is provided at the individual hospital level, whereas the demographic data is provided at the SA2 level. This means that I had to aggregate the sum of the bed numbers to their respective SA2 level so I could match the SA2 demographic data.

Solution Statement

I propose that a modelling tool (web application) is required to assist the Australian Government forecasters in understanding how many hospital beds are needed for areas that will soon be populated. This tool will be able to provide users the ability to rapidly simulate different scenarios.

Benchmark Model

As we have existing data for demographics and hospital bed data, so we can assume that this dataset can serve as the benchmark model as Australia adjusts their current healthcare demands with their healthcare supply. Newly populated areas with a certain demographic can follow the current supply and demand model.

Evaluation Metrics

The machine learning model will need to achieve at least 80% accuracy for the final features selected. A test and train dataset will be used to determine the accuracy of the model with the mean squared error used to evaluate for a regression (continuous) output. This will likely to be performed by the SciKit Learn library.

Project Design

The project is designed to be in 5 stages.

Stage 1: Collecting the data

This is finding as many data sources as possible related to Australian hospitals and demographics and joining these datasets together.

Stage 2: Exploratory Data Analysis

As there will be a lot of useful data collected here, it is worth exploring the data which could also be used for the web application.

Stage 3: Pre-Processing / Feature Engineering and Selection

This stage will be for preparing the final inputs (training and test data) prior to training the Machine Learning model

Stage 4: Selecting a Machine Learning model and Hyperparameter Tuning

Many regression models will be tested with the data to find the most suitable algorithm (without overfitting).

Stage 5: Deploying the model / Web Application

The model will be deployed as a Dash and Plotly App with Heroku. This will allow the user to provide inputs so they can simulate how many hospital beds are required in a new area.

References

Archer J., Houghton K., and Vonthethoff B. (2019). Regional Population Growth – Are We Ready? The economics of alternative Australian settlement patterns, Regional Australia Institute: Canberra.

Ravaghi, H., Alidoost, S., Mannion, R. et al. Models and methods for determining the optimal number of beds in hospitals and regions: a systematic scoping review. BMC Health Serv Res 20, 186 (2020). <https://doi.org/10.1186/s12913-020-5023-z>

R D Farmer 1, J Emami, 1990, 'Models for forecasting hospital bed requirements in the acute sector'. J Epidemiol Community Health <<https://pubmed.ncbi.nlm.nih.gov/2277253/>>

US News 2017, *Best Countries for Immigrants*, viewed 28 March 2021, <<https://www.usnews.com/news/best-countries/best-immigrants>>