Forecasting Singapore Hospitals’ Bed Occupancy Rate with Machine Learning Models

Amanda Ho Shan Rui, Chen Tianying, Tiana, Chan Yew Chun Daniel, Chan Zhi Zheng Amos, Hemanshu Gandhi, Lee Wei Ren Matthew

National University of Singapore

Abstract

This study seeks to explore various machine learning models to forecast the Bed Occupancy Rates (BOR) of hospitals in Singapore. BOR forecasting is useful as part of efforts to mitigate the ongoing problem of insufficient beds and long wait time for hospital admissions, thereby facilitating bed capacity management. Given the data of the previous weeks, we found that our models can predict the following week’s BOR with reasonable accuracy.

Introduction*[[1]](#footnote-1)*

Singapore hospitals face high BOR and have the lowest hospital beds to population ratios among the developed world. The issue is of pressing concern as the growth of Singapore’s elderly population is expected to exacerbate overcrowding and overwhelming resources of medical services. In 2015, there were only 2.4 beds per 1000 people in Singapore (Singapore Business Review 2018). This ratio is estimated to only rise marginally to 2.6-2.8 over 2020 to 2030 and will still be below the Organisation for Economic Cooperation and Development (OECD) countries’ 2015 average of 4.7 (CIMB 2018).

To ameliorate the issue of hospital bed overcrowding in Singapore, we aim to build a Machine Learning (ML) model that can provide useful predictions of the general BOR in public hospitals. ML is particularly useful for prediction of BOR because it allows the users to predict weekly BOR without needing to fully understand all the complexities and factors that affect BOR. Predictive insights from our research will assist the Ministry of Health in making proactive, data-driven capacity management decisions. In addition, our predictions would assist in the effective management of hospital beds in public hospitals. By alleviating resource constraints in hospitals, we can achieve a better quality of life for patients and reduce strain on medical resources.

Proposed Model

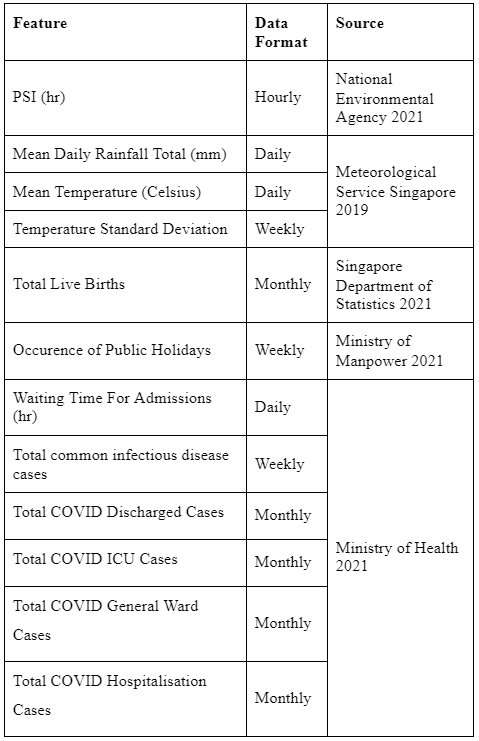
To select a predictive ML model for our application, we propose implementing various classification models for the forecasting of BOR in public hospitals and selecting the best performing one as our chosen model. Thus, five commonly used classification models were implemented for BOR prediction. The five models are Decision Tree (DT), k-Nearest Neighbours (KNN), Multi-layer Perceptron (MLP), Multinomial Naive Bayes Classifier (MNBC) and Support Vector Classifier (SVC).

Experimental Setup

Justification of Features

The features incorporated in our dataset are broadly categorized into two types, hospital-related features and non-hospital related features. Hospital-related features are features that directly influenced BOR while non-hospital related features are features commonly associated with hospital admissions in current literature (Tan et al. 2019). Additionally, BOR is influenced by anomalous events such as pandemics. Thus, we included COVID-related features to control for possible anomalous data in 2020.

Dataset

A list of features used for testing and training of our predictive model is shown in Table 1. All features were formatted in weekly data periods to facilitate building a predictive model that forecasts weekly BOR. Features not available in weekly data periods were formatted into weeks based on data format. For instance, features obtained in hourly format were formatted by averaging each 168 hour window. Features obtained in daily format were formatted by averaging each 7 day window. Features obtained in monthly format were formatted by dividing the total with the number of days in the month and aggregating each 7 day window.

Note that Occurrence of Public Holidays refers to the occurrence of at least one official public holiday in Singapore. A value of 1 indicates occurrence and 0 indicates non-occurrence for each week.

Waiting Time for Admissions measures the period between “time of decision to admit patient by doctor” to “patient enters inpatient ward”, and Total common infectious disease cases refers to infectious diseases where there are at least 5 weeks of non-zero cases across all weeks of the dataset. A total of 29 common infectious diseases were used as 29 separate features.

In addition to features shown in Table 1, discretized weekly BOR was included as class labels. Raw BOR data was originally collected in daily format from 8 Singapore public hospitals (Alexandra Hospital, Changi General Hospital, Khoo Teck Puat Hospital, Ng Teng Fong General Hospital, National University Hospital (Adult Ward), Singapore General Hospital, Sengkang General Hospital, Tan Tock Seng Hospital) (Ministry of Health 2021). As features used for prediction were national statistics, we calculated an estimated national average BOR as

The number of beds in each of the 8 public hospitals were obtained from official hospital websites (CGH 2021; KTPH 2021; NTFGH 2021; NUH 2021; SGH 2019; SKH 2018; NHG 2019). Weekly BOR was obtained by taking the mean of the aggregated estimated national average daily BOR for each 7 day window. Weekly BOR were then discretized by using strata of 5 percent (0-5, 5-10 … 95-100) and then used as class labels.

To build a weekly predictive BOR model, a sliding window containing 4 weeks of feature data was used for prediction of the following week’s BOR class label. All features were normalized. In total, 114 weeks of feature and BOR data between September 2018 to January 2021 were used. The dataset was randomly split into training and testing sets using a 80:20 split via train\_test\_split provided by scikit-learn.

Implementation and Optimization of Models

Five classification models were trained on previously split training set using the scikit-learn library available in Python. To optimise each of the models implemented, multi-grid search using GridSearchCV with 5-fold cross validation was used to identify hyperparameters that gave the highest accuracies. Models were tested on the previously split testing set using identified hyperparameters.

Table 1 List of features used for input training dataset.

Online Resources, Tools & Code Implementations Used

Data manipulation, model fitting, optimisation of hyperparameters and evaluation of model efficiency were carried out on Python 3.7.10 using the Google Colab platform. Model fitting and evaluation of model performance were carried out using scikit-learn 0.22.2 on Python.

Feasibility and Success of Model

Evaluation of Model Performance

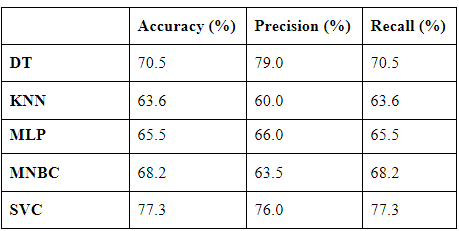
To identify the top performing model, we compared the five models based on their accuracy, precision and recall. The scores of DT, KNN, MLP, MNBC and SVC were obtained by performing the split, train and test procedure mentioned above 10 times and calculating the mean score; the mean scores are tabulated in Table 2.

Table 2: Accuracy, precision, and recall scores of DT, KNN, MLP, MNBC and SVC models

Out of our five implemented models, SVC was found to generally outperform the rest for accuracy and recall. This indicates that SVC may be the most suitable classification model for predicting weekly BOR.

Analysis of Models

The following section will analyse our model evaluation results. We consider the advantages and disadvantages of the five implemented models in relation to our obtained dataset and the BOR prediction problem.

Decision Tree (DT)

The DT model is a popular classification model because of its white box implementation, providing interpretability for its classification results. It is also robust to noise, which is especially useful for our problem because BOR is influenced by manual data collection and anomaly events such as pandemics.

K-Nearest Neighbours (KNN)

KNN does not require prior assumptions on the distribution of data and is suitable for low-dimensional datasets. Our results showed that KNN performed the poorest compared to the remaining models. This may be caused by KNN’s sensitivity to outliers which we posit is a common occurrence in our dataset.

**Multilayer Perceptron (MLP)**

MLP is a fully connected neural network and requires large amounts of information. Given the small dataset used, MLP could have performed poorly because of limitations with our data size. Additionally, MLP may be less suitable because it has a black box implementation and lacks interpretability for users to explain BOR prediction results. However, MLP may be a more expressive predictive model compared to other classification models and can be considered given a larger dataset.

Multinomial Naïve Bayes Classifier (MNBC)

MNBC is a popular classification model because it typically outperforms other models for small datasets such as the one used in our project. Additionally, MNBC is simple to implement and offers interpretability of its prediction results. However, it’s main assumption of conditional independence between features is commonly violated for real-world data. For example, features used for our BOR prediction such as rainfall and temperature features might not be conditionally independent. Thus, feature dependency should be considered when implementing MNBC for BOR prediction.

Support Vector Classifier (SVC)

SVC typically performs well when the dataset is small and when the decision boundary is irregular. It also has a white box implementation that provides interpretability for its classification results. The model also converges to a global minimum, generally resulting in better accuracy. This is also consistent with our results that showed SVC generally outperformed all remaining implemented models. However, the performance of SVC for prediction is known to be negatively impacted by noise.

Conclusion

SVC is shown to perform the best out of the 5 implemented models. Given the constraints of our research, and the data set obtained, we recommend using SVC as the ML model for predicting BOR of hospitals in Singapore. This study will aid in the efficient allocation of resources by hospitals, alleviating hospital bed overcrowding and achieve better quality of care for patients.

Future Work

Additional predictive features can be used for prediction; some predictive features could not be obtained because of patient confidentiality or non-publicly available data (e.g. length of hospital stay etc). Furthermore, obtaining more granular data such as hospital-specific data will allow medical professionals to make more actionable hospital-specific capacity management decisions rather than on nation-wide predictions. Obtaining daily data and building predictive models that can forecast further into the future will also facilitate a more dynamic capacity management.

In our study, we were limited by the amount of data we could access; 114 weeks of data was obtained for this study. More data would allow us to improve our existing models and leverage upon more expressive neural-network models such as Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM). RNN and LSTM could be better suited for our application since they are designed to work with sequence prediction problems but were not explored due to the limited data we have. We can also adopt ensemble learning to combine multiple models to get an aggregate result for better accuracy. More data would also allow us to create models that predict future BORs that are more precise, as our current model discretizes the weekly BOR by a strata of 5%.

Implications of Study

Ethical and Legal Implications

All raw features were obtained from public sources in the form of aggregates. For instance, the feature ‘Total COVID Discharged Cases’ did not disclose the individual details of COVID patients. However, should we choose to use non-publicly available data such as length of hospital stay for improved prediction, patient confidentiality should be maintained.

Novelty & Impact on Quality of Life

Despite the persistent problem of high BOR in Singapore, little data-driven studies focused on predictive models have been done to alleviate this issue in Singapore.

Through predicting BORs with the models we have evaluated, medical stakeholders are able to gain an approximation of future BORs. In doing so, they could make more informed decisions in bed management. This ensures that the issue of insufficient beds could be mitigated. Moreover, it is also possible to use the model to approximate how much each feature affects the BOR. This is because with the model, we are, in theory, able to extract the weights of each perceptron and work backwards to estimate how much each feature affects the BOR; again, this would help provide more insights to stakeholders.

Reflections

Crafting our idea and hypothesis was initially challenging as we had to formulate a problem that could be approached through machine learning, and also consider if relevant data was accessible.

Moreover, certain features of interest were either not available, or were not granular enough (ie not weekly data) to be useful. Furthermore, data from official data sources (Singstat, data.gov.sg) was insufficient for our purposes. A significant amount of data had to be manually scraped from government websites with a self-implemented python script.

Finally, selecting the appropriate ML model and optimizing its hyperparameters was difficult due to multiple factors. For instance, if our gathered features and the time step to aggregate them by (e.g. daily, weekly) were sufficiently expressive, and how we should feed the features into the model (e.g. window-based, rolling average). We had to also consider how we could standardize the implementation for different ML models for benchmarking, and tweak their hyperparameters with grid search.

Group Member Contributions

All members contributed equally to the ideation, data collection, dataset crafting, implementation of models and the report. We each implemented different models: Amanda & Tiana worked on KNN, Amos on MLP, Daniel on DT, Hemanshu on SVC and Matthew on Naïve Bayes.

References

CIMB 2018. Navigating Singapore. Retrieved from https://s3-ap-southeast-1.amazonaws.com/investingnote-production-webbucket/attachments/9aec67bbf6e85d01e926.

Changi General Hospital 2021. About CGH Home - Changi General Hospital. Retrieved from http://www.cgh.com.sg/about/Pages/default.aspx.

Edarabia 2021. Singapore Public & Private Holidays in 2021 (Full List). Retrieved from https://www.edarabia.com/singapore/public-holidays/.

Khoo Teck Puat Hospital 2021. Corporate Profile - Khoo Teck Puat Hospital. Retrieved from https://www.ktph.com.sg/about-us/corporate-profile.

Meteorological Service Singapore 2021. Historical Daily Records | - Weather. Retrieved from http://www.weather.gov.sg/climate-historical-daily/.

Ministry of Health 2021. Beds Occupancy Rate (BOR) - MOH.. Retrieved from https://www.moh.gov.sg/resources-statistics/healthcare-institution-statistics/beds-occupancy-rate-(bor).

Ministry of Health 2021. Situation Report - MOH. Retrieved from https://www.moh.gov.sg/covid-19/situation-report.

Ministry of Health 2021. Waiting Time for Admission to Ward - MOH. Retrieved from https://www.moh.gov.sg/resources-statistics/healthcare-institution-statistics/waiting-time-for-admission-to-ward.

Ministry of Health 2021. Weekly Infectious Diseases Bulletin - MOH. Retrieved from https://www.moh.gov.sg/resources-statistics/infectious-disease-statistics/2018/weekly-infectious-diseases-bulletin.

National Environment Agency 2021. Resources. Retrieved from https://www.haze.gov.sg/resources/historical-readings.

National Healthcare Group 2019. Tan Tock Seng Hospital - National Healthcare Group. Retrieved from http://corp.nhg.com.sg/TTSH/Pages/default.aspx.

National University Hospital 2021. Overview - NUH | National University Hospital. Retrieved from http://www.nuh.com.sg/About-NUH/Pages/overview.aspx.

Ng Teng Fong General Hospital 2021. Overview - NTFGH | Ng Teng Fong General Hospital. Retrieved from https://www.ntfgh.com.sg/About-NTFGH/Pages/Overview.aspx.

Sengkang General Hospital 2018. Sengkang General and Community Hospitals open their doors to serve the Northeast community. Retrieved from https://www.skh.com.sg/news/announcements/sengkang-general-and-community-hospitals-open-their-doors-to-serve-the-northeast-community.

Singapore Business Review 2018. Chart of the Day: Singapore's hospital beds inched up by a measly 2.3% in the last decade. Retrieved from https://sg.finance.yahoo.com/news/chart-day-singapores-hospital-beds-013500403.html.

Singapore Department of Statistics 2021. SingStat Website. Retrieved from https://www.singstat.gov.sg/.

Singapore General Hospital 2019. Hospital Overview - Singapore General Hospital. Retrieved from https://www.sgh.com.sg/about-us/corporate-profile/hospital-overview-singapore-general-hospital.

Tan, K. W.; Ng, Q. Y.; Nguyen, F. N. H. L.; Lam, S. S. W. 2019. Data-driven decision-support improvement through predictions of bed occupancy rates. Retrieved from https://ink.library.smu.edu.sg/cgi/viewcontent.cgi?article=5687&context=sis\_research.

Yi, S. B. 2018. Alexandra Hospital to have new facilities, more beds by 2020. Retrieved from https://www.straitstimes.com/singapore/health/alexandra-hospital-to-have-new-facilities-and-around-300-beds-by-2020.

1. Copyright © 2021, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved. [↑](#footnote-ref-1)