Prediction of Length of Stay for Elective Procedures in Irish Hospitals

MSc Research Project
MSc in Data Analytics

Ian Darbey
Student ID: X17168368

School of Computing
National College of Ireland

Supervisor: Dr Catherine Mulwa

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Ian Darbey
Student ID:	X17168368
Programme:	MSc in Data Analytics **Year:** 2019
Module:	Research Project
Supervisor:	Dr Catherine Mulwa
Submission Due Date:	12/08/2019
Project Title:	Prediction of Length of Stay for Elective Procedures in Irish Hospitals
Word Count:	**9919** **Page Count 25**

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: ………

Date: ………

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Prediction of Length of Stay for Elective Procedures in Irish Hospitals

Ian Darbey
X17168368

Abstract

Length of stay is a measure of hospital service delivery efficiency used throughout the world. The ability to predict a patient's length of stay has the potential to influence management decisions on when other patients may be scheduled for surgery and other elective procedures. Additionally, it would allow for the enhancement of resource allocation for nursing and support staff based on the expected patient dependency and for how long. This research project attempted to predict whether a patient will stay longer than the typical length of stay for a scheduled procedure. A variety of predictive models were developed with the final model, a gradient boosted decision tree, correctly identifying approximately 4 out of 5 long stay patients. The research output is developed to the standard of being suitable for deployment in one of Ireland's largest acute hospitals and the methodology such that it could be implemented in other Irish hospitals with appropriate datasets.

1 Introduction

An extended length of stay for patients has a significant impact on both the patient, as an individual, and the wider community of patients who are awaiting elective procedures. Extended length of stay negatively impacts bed availability for subsequent patients and contributes to lost operating theatre time and ultimately hinders a healthcare organisations ability to deliver a quality and cost-effective service. This project will use one hospital for research purposes with the methodology expected to be easily adapted to other acute hospitals in Ireland.

1.1 Motivation and Background

Length of stay (LOS) is a widely used metric in measuring the efficient and effective delivery of healthcare services throughout the world. Long lengths of stay are associated with higher per patient cost and reductions in overall capacity for elective procedures. This is especially prevalent in the delivery of planned elective procedures. Extended LOS for any given patient negatively impacts overall bed capacity and can result in planned procedures in subsequent days being cancelled due to the lack of an available bed for a patient post procedure. As such, being able to predict a patient's LOS is of potentially significant benefit both hospital managers, surgeons and other medical professionals delivering planned procedure services. It has the potential to facilitate enhanced booking practices based on enhanced predictability of bed availability. In the context of the hospital in focus for this project, Beaumont Hospital, it also has the potential to impact and inform the utilisation of the Framework for Safe Nurse Staffing and Skill-Mix (Department of Health, 2018).This is in the process of being rolled out throughout the hospital and seeks to assign nursing staff levels and expertise based on patient dependency. In addition, the hospital is currently in the initial scoping phase of a business case to expand its bed count. The use of predictive models for LOS derived from the hospital's own activity data presents an opportunity to construct more realistic simulations. This may help to define the potential increase in elective activity which the provision of additional bed capacity may facilitate.

These potential benefits serve a wider purpose in a social and health context. The hospital may be able to better understand the demands each patient may present following their elective procedure. This could facilitate better management of the overall patient cohort with enhanced service delivery and the potential to reduce day of procedure cancellations, which can have a negative impact on patients both physically and mentally.

Beaumont Hospital currently has nearly 7,000 patients, either classed as actively waiting (National Treatment Purchase Fund, 2019a), or with a provisional scheduled date set for their elective procedure (National Treatment Purchase Fund, 2019b).

Similar research has been performed throughout the world, which often focuses on very particular subsets of procedures or admissions. There are a large variety of delivery models in healthcare and due to the differences in community support levels outside of the hospital setting it is important to establish the results of this research for the Irish health service, in particular given Beaumont Hospital's status as one of the largest acute teaching hospitals incorporating a number of significant national and regional specialities.

1.2 Problem Statement

Current elective procedure scheduling practices are based on the experience of non-clinically trained staff and the approach is not based on any specific or well-defined practice methodology. The lack of a well-defined approach in addition to a lack of consideration of expected length of stay and resource / skill set requirements often results in unrealistic or impractical quantities or complexity of procedures being scheduled. This results in the cancellation of significant numbers of procedures on an annual basis. In order to enhance this process, it will be necessary to develop an understanding of the elements of a patient's history at the scheduling stage which can impact the length of stay. Such an understanding will assist in producing a predicted class of length of stay. To address the problem the following questions were investigated;

Research Question 1: Can patient length of stay be correctly predicted using information available prior to elective (planned and scheduled) admission?

Research Question 2: Can the factors which influence patient length of stay be identified and highlighted to the relevant clinical and management teams?

1.3 Research Objectives

The initial research will require an in-depth literature review of the current state of the field in order to establish prevailing methodologies, if any, and ensure that lessons from other published research can be identified and integrated into this research project, as well as identifying any significant research gaps. Table 1 shows the primary research objectives and the evaluation method to be used.

Table 2 shows the chapter structure for the remainder of this technical report.

Table 1 – Research Objectives

Objective	Description	Evaluation Method
1	Review current state of the field	Literature Review
2	Identification of factors influencing length of stay	Correlation testing and importance factor rank
3	Implementation of Predictive Models for Length of Stay Classification	Specificity
3(a)	Implementation of Random Forest	Specificity
3(b)	Implementation of Naïve Bayes	Specificity
3(c)	Implementation of Artificial Neural Network	Specificity
3(d)	Implementation of Support Vector Machine	Specificity
3(e)	Implementation of Gradient Boosted Decision Tree	Specificity
3(f)	Implementation of Generalised Linear Model	Specificity
3(g)	Implementation of C5.0 Decision Tree	Specificity
3(h)	Implementation of Ensemble Model based on Gradient Boosted Decision Tree	Specificity
4	Evaluation and selection of predictive model (Objective 2) for deployment	Specificity

Table 2 – Chapter Structure

Chapter	Title	Description
2	Related Work	Provide an in-depth review of the current state of the field with a literature review
3	Methodology	Outline the research methods used and explain implementation choices
4	Implementation	Discuss the implementation methods used for each predictive model build
5	Evaluation and Results	Identify the best performing predictive model
6	Discussion	Reflect on the data pre-processing methods, the research, its contribution to the body of knowledge and its potential applications and the limitations of the research and associated data
7	Conclusion and Future Work	Summarise the findings of the research and identify areas of potential further or enhanced research areas which the work may inform or provide a methodology which could be suitably adapted for.
8	Acknowledgements	Acknowledge contributions of colleagues and others in completion of the project.
9	References	List of academic references

2 Related Work

2.1 Introduction

Related research in the area is abundant; many researchers hail from a clinical background while some researchers are from a data mining background and are finding success with the implementation of predictive models which is becoming a growing trend across the research as opposed to pure statistical analysis. With the continually expanding use of electronic health records in healthcare organisations, including in Ireland (St James's Hospital, 2018), the availability of data in a format that can be leveraged with the use of data analytics and machine learning methods is growing considerably.

Having a structured research approach informed by other research in similar areas is likely to be beneficial in the roll out and acceptance of predictive model usage based on health data in its various forms.

2.2 Review of Specific Procedure Speciality Groupings

2.2.1 General Surgery Procedures

The factors influencing length of stay were analysed and used in a data mining exercise to predict length of stay by Aghanjani and Kargani for general surgery patients (Aghajani and Kargari, 2016). The researchers used the data from 327 patient records and used Naïve Bayes, k-nearest neighbours and a decision tree, specifically C4.5 (Quinian and Ross, 1993) to predict length of stay and analyse the most influential factors. The decisions regarding what patient information to include in the research was informed by review of existing research and by a survey of 4 clinical experts to validate the suggested input data. The output for prediction was length of stay and the researchers treated this as a classification rather than a regression problem by dividing the length of stay in to 3 classes; 1-3 days, 4-5 days and 6+ days. The data was split 70/30 for training and test data and all 3 models had over 90% recall on the 1-3 day length of stay class. All three under 65% recall for the 4-5 day length of stay prediction and were over 85% for 6+ day length of stay prediction, with the decision tree performing best at 93% recall for this class. Overall the decision tree performed best with over 84% accuracy. Interestingly the models varied considerably regarding which features were the most influential with all 3 models in agreement on just one feature with just 3 other features being identified as influential across 2 models. In addition, there was some comparison of the features identified within the predictive models against the consensus from the clinical experts. In particular, average number of visits per day post operatively was deemed an influential factor by all 3 predictive models yet was only suggested to be an important factor by one of the clinical experts.

It was attempted to predict prolonged length of stay for surgery patients for both urgent and non-urgent operations (Chuang *et al.*, 2016) with urgent operations classified as patients who had their procedure within the first 2 days of admission. The length of stay was classed as prolonged if it exceeded the national average for that surgery. A decision tree, support vector machine and random forest were used with both operation classifications. All 3 models performed well with the non-urgent procedures with the random forest proving the most accurate model for both classes and having an area under the curve (AUC) of 0.938 for urgent operation length of stay prediction and 0.945 for non-urgent operations. The results while positive are perhaps reflective of the care model in use in the hospital used as the data source as patients are admitted 2 days prior to procedure in some cases for investigations. By comparison the target hospital for this study operates with a more structure preadmission pathway in which radiology and other investigations are complete, where possible, prior to admission and in many cases, admission is on the day of surgery.

2.2.2 Cardiology Procedures

A similar research project, more specifically targeted at cardiac patients (Hachesu *et al.*, 2013) employed a similar approach to the general surgery research, with the use of classification rather than regression, to assess the accuracy of the predictive models used. A radial-basis function artificial neural network, support vector machine and decision tree, specifically a C5.0 boosted decision tree, were used along with an ensemble model which combined the three models. While the authors also make use of post admission information, as opposed to the plan to use only information known prior to admission for this research, it ultimately provides some interesting results with artificial neural network struggling with the predictive task, with 54% overall accuracy, when compared to the other models. The support vector machine produced the highest overall accuracy figure of 96.4% while the ensemble model performed better with 98.2% sensitivity. While it did focus on information often updated or only available post admission, it highlighted some features such as blood pressure and cardiac history, which were shown to be important features in the analysis of the support vector machine, and would encourage thought on how to incorporate data about cardiac history in to the dataset used for developing the predictive models in future research.

2.3 Research Involving Specific Co-Morbidities

2.3.1 Diabetic Patients

A sample of patients from the Healthcare Cost and Utilisation Project (HCUP) Nationwide Inpatient Sample were used by Morton et al. to predict length of stay for diabetic patients (Morton *et al.*, 2014). The researches use random forest, support vector machine, support vector machine plus, multi task learning and multiple linear regression to determine if the patient is likely to fall in to the short-term or long length of stay category, with a threshold of 3 days beyond which the patient is classified as long stay. The researchers select 10,000 patient records from the overall dataset with class and gender balance considered. The models were built with 5-fold cross-validation and compared. The support vector machine plus model performed best with accuracy of 0.68 ± 0.01 across the 5 trials and an AUC of 0.76 ± 0.01. Of note, the multiple linear regression model performed poorly, likely due to the non-linear nature of the underlying problem. The study by Morton et al. contrasts well with a similar project specific to diabetic patients (Alahmar, Mohammed and Benlamri, 2018). The researchers again used 3 days as the division between long and short stay patients however this was derived based on the median length of stay within the dataset once all pre-processing steps had been completed. Additionally, the researchers used a different selection of machine learning techniques, specifically Naïve Bayes, generalized linear model, deep learning, distributed random forest, gradient boosting machine and a stacked ensemble of all the models. The researchers used an 80/20 training and test split and balanced the classes by oversampling the minority and under-sampling the majority class. The distributed random forest and gradient boosting machine both had an AUC of 0.8 with the stacked ensemble performing best with 0.81. The weakest performer was the Naïve Bayes model however excluding it from the stacked ensemble also reduced the ensemble model's effectiveness.

2.3.2 Patients with Congestive Heart Failure

Significant research on analysis and prediction of length of stay was conducted on patients with congestive heart failure (Turgeman, May and Sciulli, 2017) and cardiology patients (Tsai *et al.*, 2016). The study involving patients with congestive heart failure incorporated a total of 5 models including a cubist regression tree which uses a linear regression at the leaf nodes to minimise absolute error. Its use of a support vector machine to reduce error in cases where the patient had more previous admissions and more lengthy previous stays is an interesting approach and may prove insightful during this research project. This approach and the researchers use of multiple models contrasts with the research on cardiology patients (Tsai *et al.*, 2016) which focused on the deployment of an artificial neural network to predict length of stay and is only compared against the output of a linear regression model. The research ultimately struggles to produce an artificial neural network that markedly outperforms the linear regression model. The influential features identified in both research papers highlight the overall influence cardiac history and issues can have on length of stay. The paper focused on congestive heart failure produces some impressive AUC rates across the models with 0.80 produced ultimately which compares favourably against the overall accuracy of less than 70% produced by Tsai et al.

2.3.3 Stroke, Geriatric and Orthopaedic Patients

Length of stay for stroke patients (Al Taleb, Abul Hasanat and Khan, 2017) was predicted to fall within 4 classes (0-2, 3-7, 8-16 and >16 days) with the use of a Bayesian Network and a J48 decision tree (based on C4.5) compared following 10-fold cross validation. Both models performed quite well however the Bayesian Network proved more accurate overall with approx. 81%. Outside of stroke specific and post admission measures age was again an influential factor.

Research relating to a geriatric hospital (Liu *et al.*, 2006) presented a starkly contrasting definition of a long stay patient with 61+ days classed as long. The researchers used several Naïve Bayes models to impute missing features which resulted in an improvement in the classification performance above the initial approx. 55% overall accuracy, particularly in relation to the medium and long stay classes. Accuracy on the short stay class was the only time it was over 40% however, ranging 76.8%-84.3% across the Naïve Bayes and decision tree models. It may have helped overall performance if the researchers had reduced scope to predict 2 classes rather than 3 given the models difficulties with both higher length of stay classes.

Much of the research is specialty or even procedure specific with research in elective ankle surgery for end-stage ankle arthritis (Pakzad *et al.*, 2014), total hip arthroplasty (Elings *et al.*, 2015) and total and unicompartmental knee replacement (Ong and Pua, 2013) all under the collective of orthopaedics procedures. These research papers all identify age and comorbidities as contributing factors to length of stay. Given the differing approaches and datasets the results are not directly comparable, however, Pakzad et al. achieved AUC of 0.67 for preoperative known information and Ong and Pua produced a linear model with an R^2 value of 0.34 and accounted for 95% of testing data being classified within a confidence interval of ± 3 days.

2.4 A Critical Review of Specific Approaches, Techniques and Algorithms

A multi-tiered approach was taken to predict length of stay (Azari, Janeja and Mohseni, 2012) with K-means clustering used as a precursor to group similar patient profiles prior to the application of a wide variety of machine learning algorithms. The researchers developed an ideal value of k within the design process and then compare 3 k values, including the one which minimises the sum of squared errors, and no clustering as inputs to the variety of predictive models. The result is a mix between several models and the various clustered inputs in terms of what performs best. The rule based Jrip, a Bayesian network and support vector machine, proved to be the most consistent performers and as such the development of an ensemble model using these high performers may have proved more successful and generalisable.

Neural Networks were applied to the MIMIC III dataset (Gentimis *et al.*, 2017) to predict whether a patient would be a short or long stay (>5 days). The authors describe both a neural network and a random forest. However, they only apply the neural network which leaves the research with nothing but a linear model to be compared against for performance. The neural network was fit 100 times on random subsets of the data and produced overall accuracy between 0.75 and 0.823 and a mean of 0.78855. This compares well with a linear model, unspecified which variant, which produces 0.57. It's not clear why the authors chose to describe but not assess a random forest model within the research and while the approx. 80% accuracy versus the linear model is a significant increase it would have been beneficial to benchmark this performance against other non-linear models which may have been more effective addressing the research problem.

Factors which significantly influence length of stay and readmission were shown to be interrelated (Kelly *et al.*, 2012) and importantly, while the authors did not discuss a predictive model in detail, the analysis of the factors is excellently detailed. Median length of stay increases with age and does not show any significant gender bias, especially in elective admissions, while public patient median stay was longer and patients who do not discharge to their own home also had an increased length of stay along with patients with a defined comorbidity. The odds ratio for each of this reflective this and, interestingly, the readmission odds ratio for each factor was often within the range of the odds ratio for the same factor for elective and emergency patients.

There has been research into methods to optimise surgical planning based on the use of data mining output to understand underlying factors which cause variance. Goal programming is used (Li *et al.*, 2017) to optimise between the bed capacity of the hospital in focus along with waiting list and theatre operating room scheduling with the variance of length of stay contributing to a much broader surgical schedule. A similar goal is research in relation to optimising operating room time when capacity is shared between elective and emergency procedures (Lamiri, Grimaud and Xie, 2009). Both of these papers frame the bigger strategic picture, which this research may form part of in the future, in particular with simulation modelling now possible directly using R with the use of the simmer package (Ucar, Smeets and Azcorra, 2017).

The increasing use of electronic systems to manage patient information has led to the application of machine learning to the information within to provide insights. Length of stay is predicted using a combination of clustering and predictive modelling (Kumar and Anjomshoa, 2018) with CART, random forest and k-NN used for the predictive elements. The authors favoured the patient groupings identified by the CART model due the increased interpretability of the results based on patient information. Many of the models perform well and ultimately 53% of the length of stay variance was definable. This contrasts with the research using electronic health records (Baek *et al.*, 2018) which uses a multiple linear regression model and accounts for approx. 26% of the variability within their test data. This perhaps reinforces some of the other work which has shown linear regression to struggle with the prediction of length of stay.

A number of models were used to test and develop a generalisable predictive model for length of stay (Steele and Thompson, 2019) using a HCUP Florida SID dataset. Using 10-fold cross validation the researchers achieve AUC of 0.899 for a Naïve Bayes model and 0.900 for a Bayesian network. They then tested the models on the next quarterly batch of patient data and achieved 0.904 and 0.901 respectively suggesting the models were generalisable. Given the significant differences in care delivery in the United States and Ireland it will be interesting to see if this level of performance can be achieved with the dataset used for this research project.

2.5 Conclusion and Identified Gaps

The related work in this area will help guide the project where applicable, particularly in relation to feature selection. This is not just an interesting data problem to research, active surgeons in Ireland are working on improving length of stay (Solon, Coffey and McNamara, 2013) and the combination of data mining research and clinical input in to this project is expected to lead to more actionable insights in addition to a predictive model or models. Given the target organisation does not currently have a fully electronic patient record in place, it is hoped that, the combination of this research project and the previous research involving electronic health records may provide a structured approach for similar and related projects in the future. The worldwide nature of the literature is reflective of the interest in the field, however the lack of specific research in Ireland, or even Europe, in this topic serves to highlight a significant research gap which this research project is positioned to begin filling. This chapter addresses research objective 1, to review the current state of the field, from Table 1 – Research Objectives.

3 Methodology, Design and Data Preparation

3.1 Introduction

Understanding length of stay is a core element of operational planning for scheduled procedures in an acute hospital setting. Knowing and understanding how long specific cohorts of patients will take to recover and complete their acute post procedure stay is an important aspect to ensure optimum resource utilisation. Given the expected needs and likely length of stay for a given set of patients appropriate nursing care, both in terms of resource level and appropriate skills, can be planned accordingly. Additionally, it would help inform simulation modelling for long term planning and allow the hospital to assess desired elective throughput against the bed count and staffing requirements to support same. Being able to model such a simulation based on samples from actual patients as opposed to broad stroke averages would lead to a more refined and likely more realistic simulation.

3.2 Modified Methodology Design

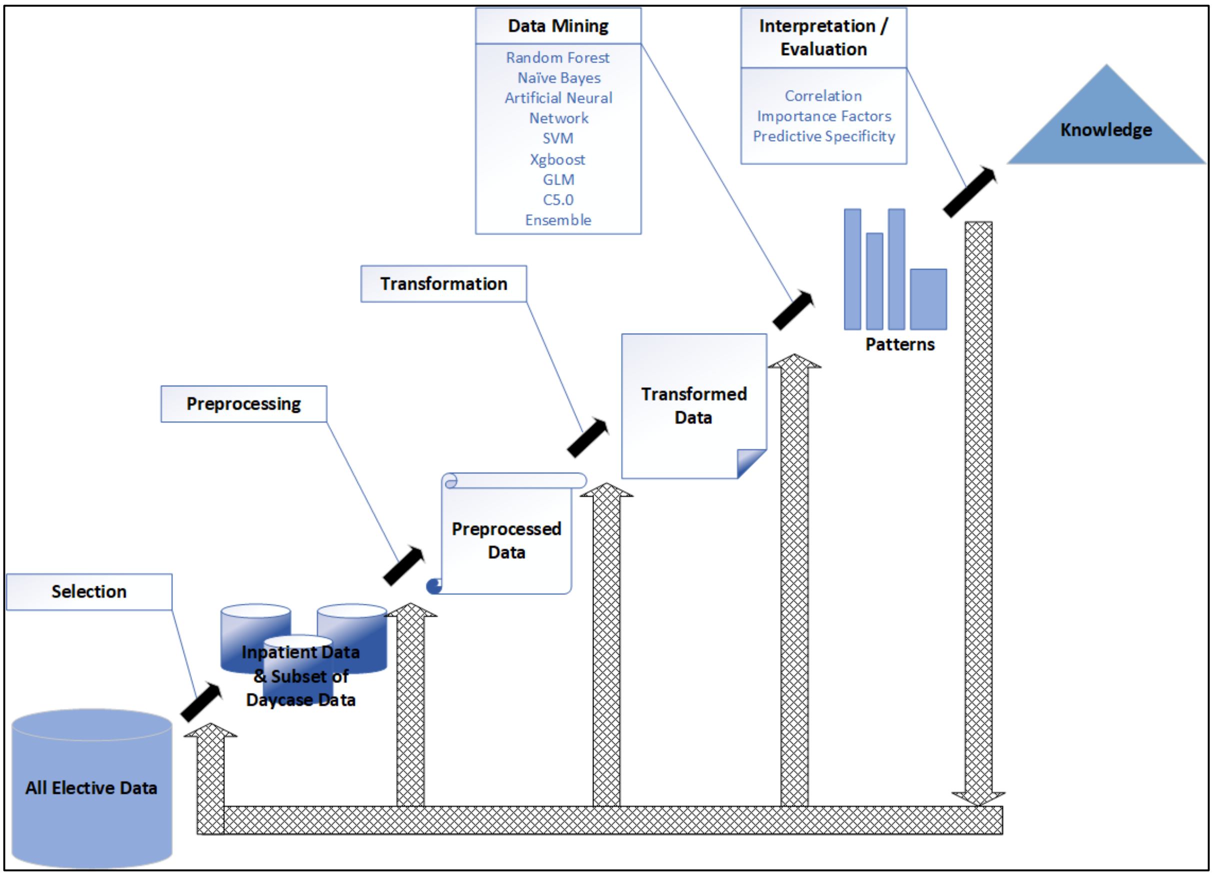


Figure 1 – Patients Length of Stay Methodology

The research methodology is intended to follow a modified approach based on the methodology outlined in the KDD process (Fayyad, Piatetsky-Shapiro and Smyth, 1996).
It is intended that in addition to the predictive models the knowledge gained will also be an outline of the biggest influential factors for each procedure and patient cohort to better inform patient management and supports, as appropriate.

In order to establish a generalisable model or models it is quite likely, given the wide variety of procedures, that a preliminary clustering of procedures will be required as part of the pre-processing stage to ensure the predictive modelling does not need to cover an inappropriate amount of variance from procedure to procedure and each model would be trained on each cluster separately. This is expected to help to ensure that simpler procedures are clustered together, and that significantly varying procedures are analysed and understood separately, for example, a simple hernia repair and a bowel resection done to treat a cancer patient are very different procedures with different speciality and recovery timeframes. This subdivision of the primary dataset is expected to ensure enhanced insights and feature analysis in tandem with the granularity to ensure that, if appropriate, each subgroup has the most appropriate predictive modelling method applied to it based on the models expected accuracy and sensitivity from training and testing. While clustering is the expected approach, it's possible the insights gained at the data mining and evaluation stages, or any of the previous stages, may ultimately alter this approach in favour of a different approach which may result in a single model choice. The selection, pre-processing and transformation elements of the process are discussed in section 3.3, the data mining / pattern discovery is addressed in section 3.4 and in chapter 4, while the interpretation and evaluation is addressed in chapters 5 and 6.

3.3 Data Selection, Pre-processing and Transformation

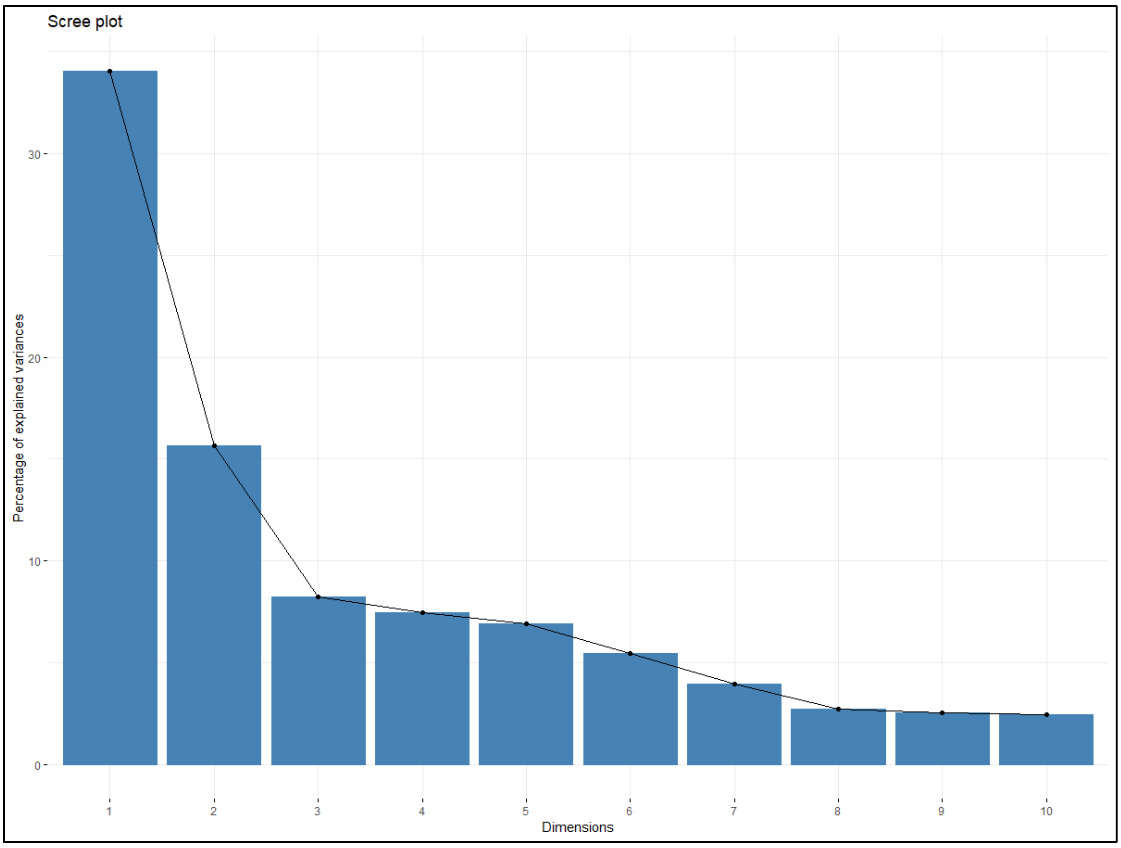
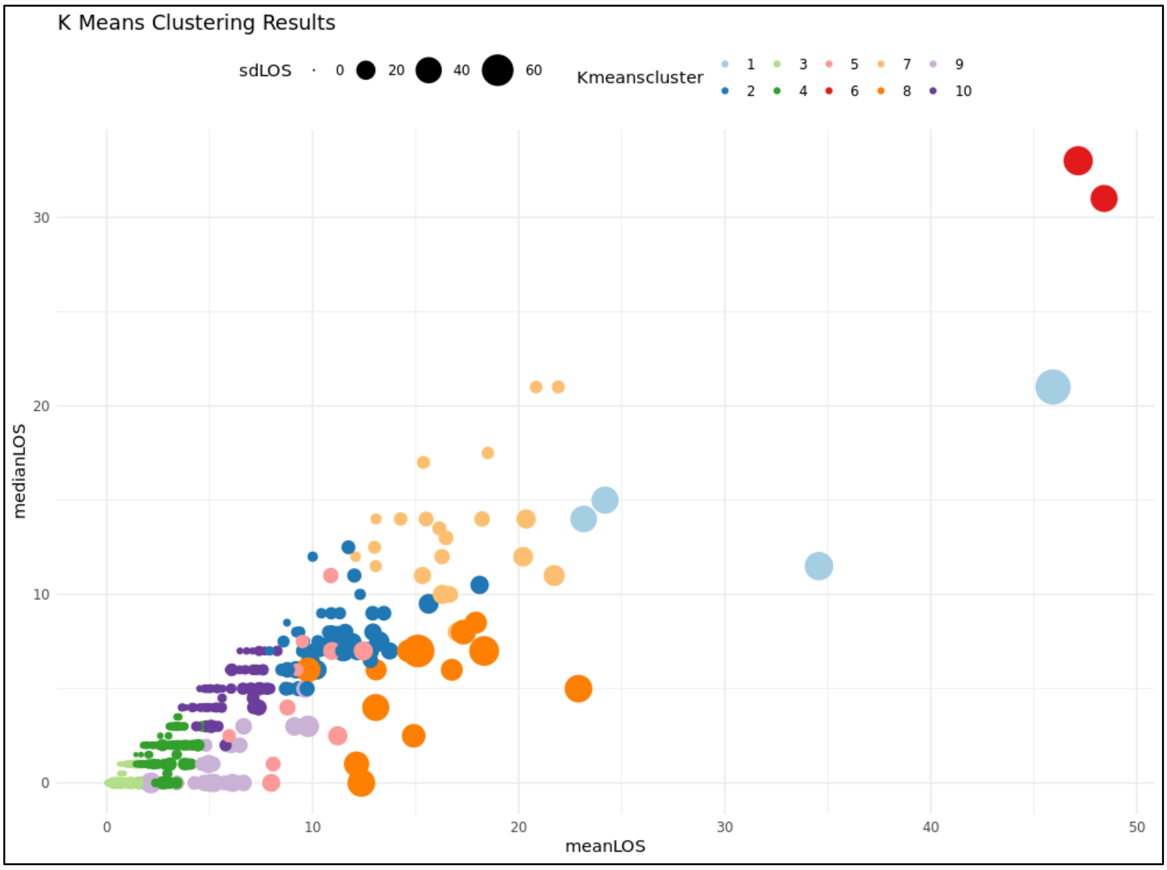


Figure 2 - Scree Plot for Inpatient History

The dataset request to the hospital has been heavily informed and influenced by the literature review and the clinical expert consultation, Appendix 1 – Expert Interview Notes.

The nature of the dataset was such that selection, pre-processing and transformation were iterative processes which informed each other as well as being influenced by the exploratory data analysis and data mining process.
Given the dataset was initially split over several tables, several joins were required to achieve the intended dataset. The history number is unique to each patient and was used to facilitate joins between the relevant tables. Importantly the joins needed to be date sensitive by filtering immediately post join. This helped to ensure the project goal of predicting at scheduling stage is respected and that for each elective admission only the information which was available prior to that admission is included. At the end of the data pre-processing stage the history number will be removed in order to maximise the anonymity of the dataset.
The number of columns which record the outpatient and inpatient attendance history is quite large, over 30 for each dataset. As such it is necessary to reduce these to a more manageable set of dimensions. Principal components analysis identified that a significant portion of the variability can be reduced to 2 components in the case of the OPD history and 3 in the case of the inpatient history, as shown in Figure 2 - Scree Plot for Inpatient History.



.

Figure 3 - K-means Clustering Results

As part of the preparatory analysis the median, mean and standard deviation of length of stay for each procedure was calculated. Additionally, the interquartile range was calculated and used to calculate the trim point for long length of stay as per the National Quality Assurance and Improvement System (Royal College of Surgeons Ireland, 2017). This is calculated as the 75% quartile value plus 3 times the interquartile range. Given the volume of procedures carried out it was necessary to perform some form of grouping of procedures, using the above, into a manageable cohort. A balance between actual procedure variation and a manageable number of clusters led to the choice of 10 clusters. The choice of 10 clusters was considered to be an appropriate compromise between the statistical elements and the practicality of explaining the groupings to the relevant stakeholders especially those who have a specific interest in a small subset of the overall procedures. Following this K-means clustering was performed based on a k value of 10 and 10,000 random starts, this resulted in groupings as per Figure 3.

The lengths of stay where then classified as "Good" or "Long Stay" with the mode being the division point. Lengths of stay shorter or equal to the mode were classed as "Good" and anything stay longer than the mode was classed as "Long Stay". Initially this produced a reasonable approach with the use of the mean, median and clustering, with the median as the classification split point, however, further consideration led to some experimentation and the use of the mode. Ultimately the use of the mode enhanced the predictive model performance for both classes and as a result the clustering provided little predictive value.

The reason for using the mode as opposed to the median or mean is both operational and statistical. The median is midpoint for each distribution so is affectively similar but to a lesser extent to the mean due to extreme values, the domain in question also meant that the variation was predominantly upwards which shifted the median of the distribution to higher values. Using the mode as the division point ensures that any management or interventions (e.g. applying for home care supports pre-admission) are appropriately targeted whereas the use of the mean, by comparison, may have led to unnecessary management effort. The dataset was then filtered to procedures classed as inpatient (which included day-cases which stayed overnight due to the nature of retrospective classification of patient episodes) which led to an imbalance towards Long Stay which was countered by sampling the day cases to approximately balance the dataset.

3.4 Exploratory Data Analysis

The final dataset for predictive modelling contains 27,213 elective admissions with the relevant principal components analysis output of outpatient and inpatient background history. The split on gender and LOSclass, is relatively even, as shown in Figure 4.

It is quite notable that over half the patients in the final dataset exceed the mode length of stay for their procedure. The histograms of patient age at admission and mean length of stay for procedure, Figure 5, both show some underlying trends for their respective data. Both have been binned in groups of 5 and it is notable that the higher age ranges have increased patient counts and conversely that mean length of stay is distributed so significantly to lower values (plots of median and standard deviation show a similar pattern) which, when viewed with the length of stay classification split in mind, presents an interpretation that it is not necessarily procedure complexity which is most impactful on length of stay classification but is likely something particular to the patient or their management which has a greater impact.

The correlation matrix of the final dataset, as shown in Figure 6, illustrates some interesting characteristics. The mean, median and standard deviation of length of stay are shown to be correlated and this is somewhat expected however, it is perhaps surprising that standard deviation is not as strongly correlated as mean and median which is perhaps suggestive that the variance within length of stay is not as significantly associated with procedure type as some others.

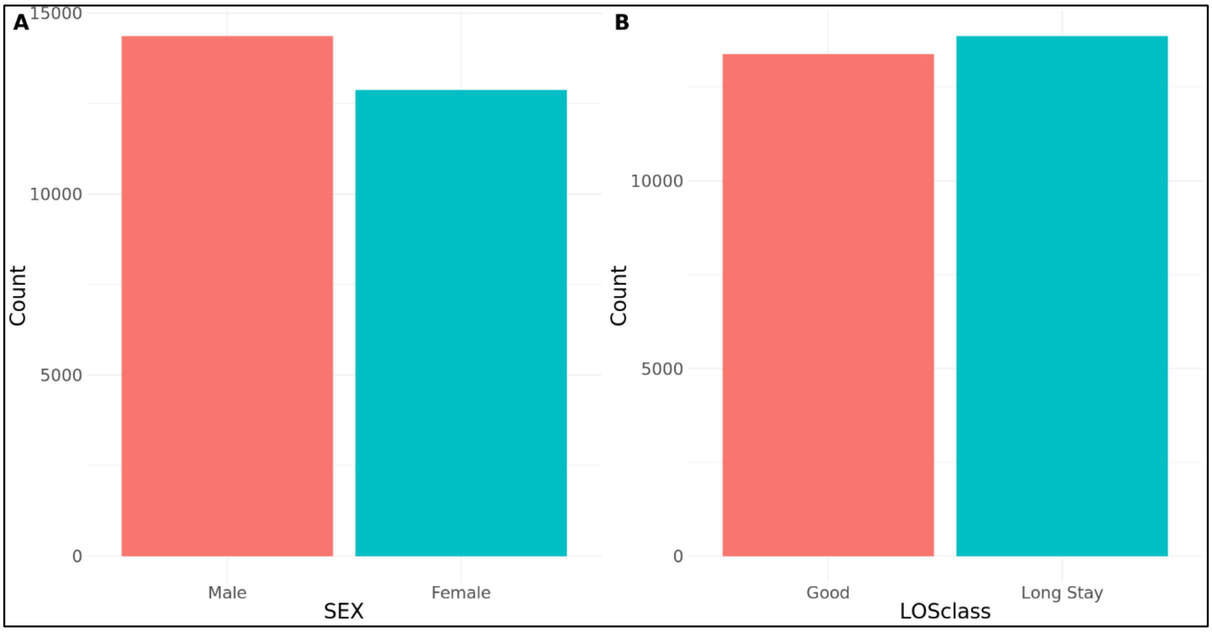


Figure 4 - A Gender Split **B - Length of Stay Classification**

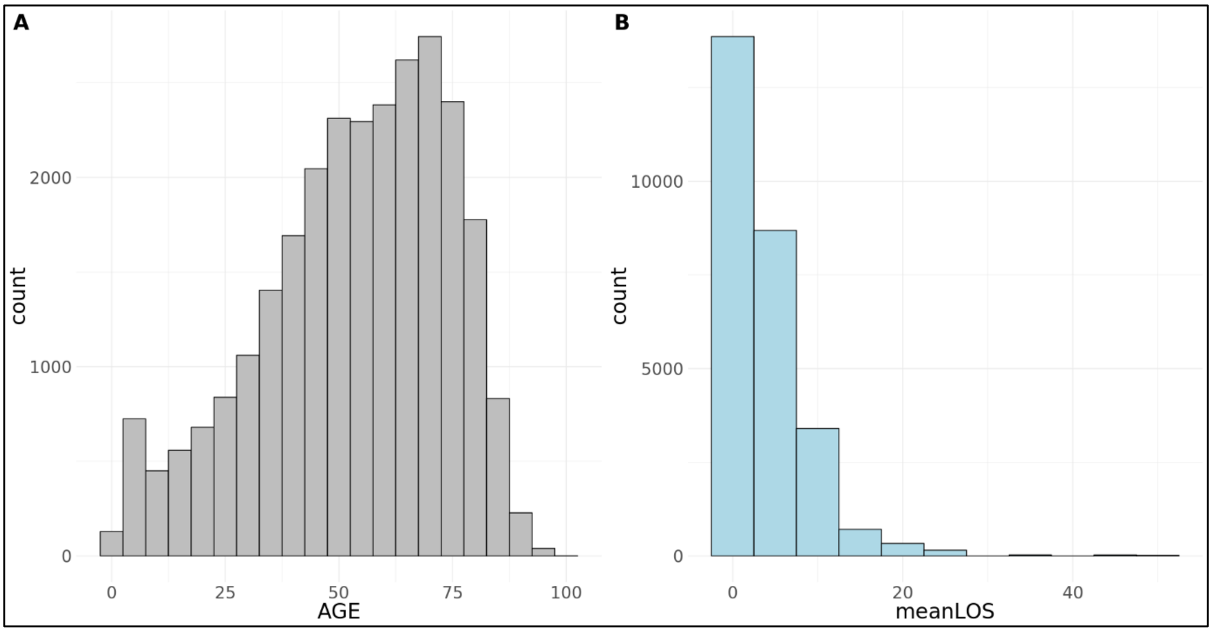


Figure 5 – A- Histogram of Patient Age **B- Histogram of mean Length of Stay**

Aside from the expected correlations between the length of stay summary statistic variables it is also notable to see a strong correlation between the first principal component of the outpatient background history and the inpatient background history. Logically this may be somewhat expected as patients who are frequently inpatients would tend to have outpatient follow ups post discharge, and at least logically, it may suggest they have more comorbidities for which more regular attendance at both the accident and emergency department and the outpatient services provided by the hospital is required.

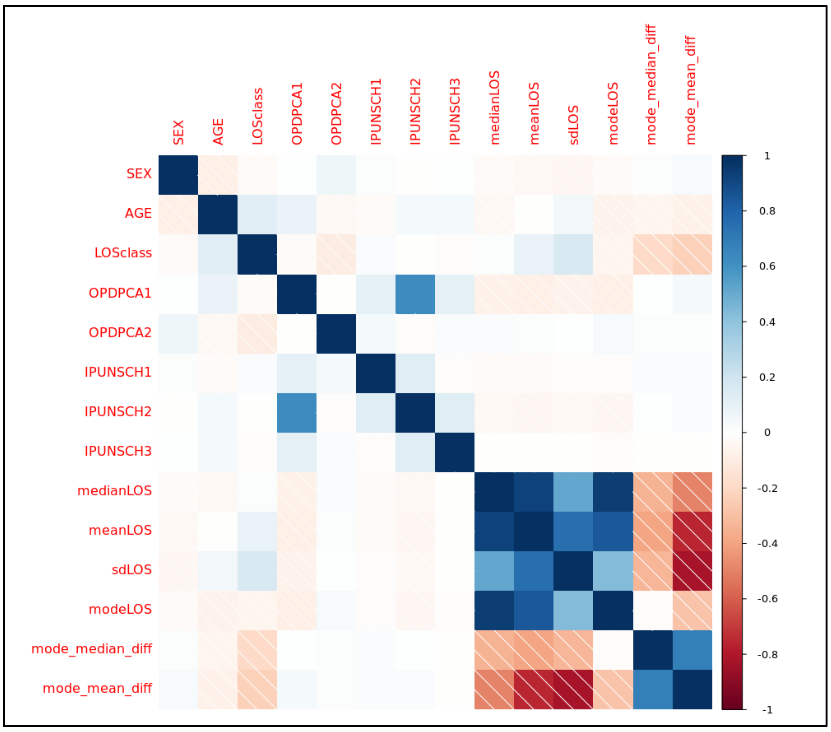


Figure 6 – Correlation Matrix for Final Dataset

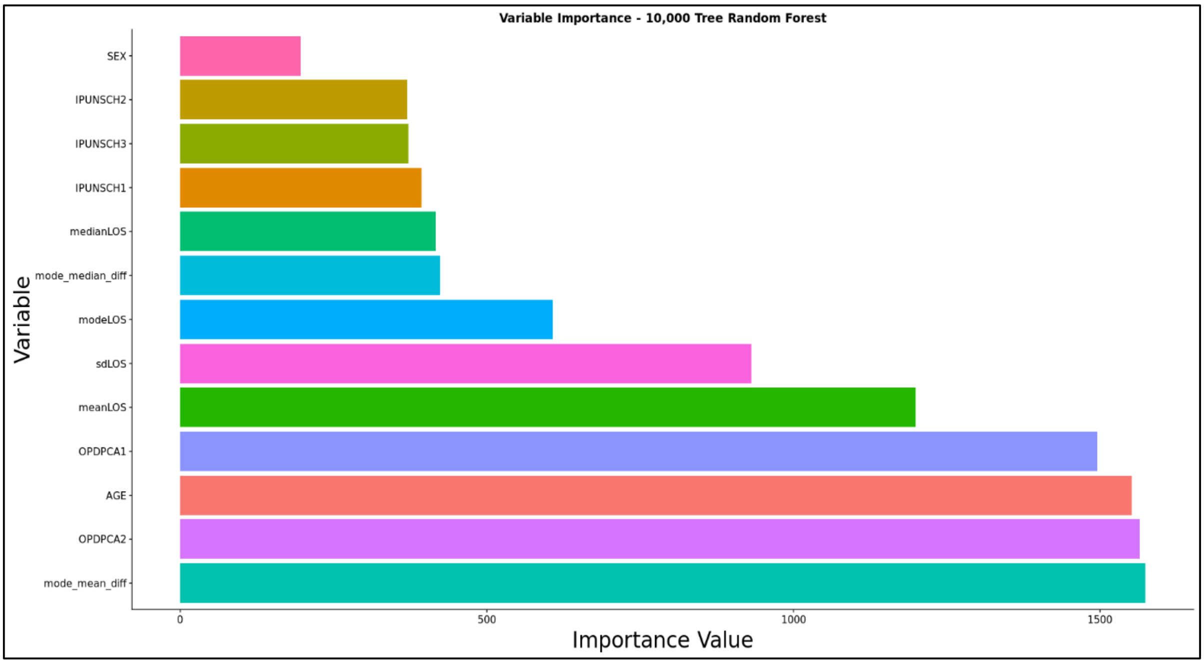


Figure 7 – Variable Importance based on 1000 tree random forest

Beyond the correlation between the independent variables, our outcome variable, LOSclass, shows weaker correlations with age being the biggest contributor outside of the procedure level summary statistics. It's notable that the correlation is highest for those which reflect the variability of Length of Stay (standard deviation, mode and median difference and mode and mean difference).

The variable importance measure, based on impurity of node split, from a random forest of 10,000 trees built using the ranger package (Wright, Wager and Probst, 2019) is shown in Figure 7. This was built to predict LOSclass from the remaining variables. Like the correlation matrix it shows that the Age and length of stay summary statistic variables have significant importance in predicting the outcome of the length of stay classification. It is perhaps interesting to note that it ranks both outpatient principal components variables ahead of all 3 inpatient principal components.

From both the correlation matrix and the variable importance measures it becomes more justifiable to infer the most influential factors which appear to be age, the summary statistics for the procedure concerned (which is an inherent reflection of procedure complexity and expected associated length of stay post operatively) and patient background history, with OPD apparently more influential a factor than inpatient history.

3.5 Architectural / Technical Design

The proposed technical design outlining dataset management and processing within the KDD process is shown in Figure 8 – Outline of Dataset path through modified KDD process, with the primary work after initial dataset extraction being completed within R.

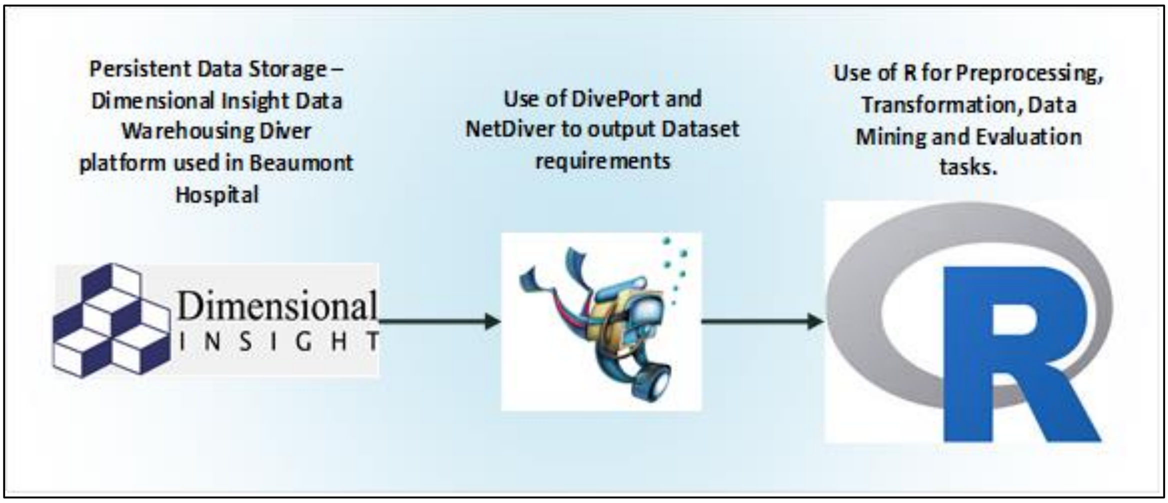


Figure 8 – Outline of Dataset path through modified KDD process

By completing as much of the post dataset output tasks as possible within R, the methodology will be open and transparent within the R scripts and facilitate commenting and annotation where relevant.

There will be several R scripts which perform various functions to facilitate the development and subsequent deployment of the application. The final deployment will contain scripts dedicated to specific tasks such as one dedicated to the pre-processing and transformation processes, another dedicated to the development, testing and evaluation of the models and finally the script which will ultimately support the primary deployment which will load the saved data models and make LOS predictions as required and save the results.

This final deployment script's place in the intended final deployment workflow is shown in Figure 9, with more detail in section 3.7 Deployment.

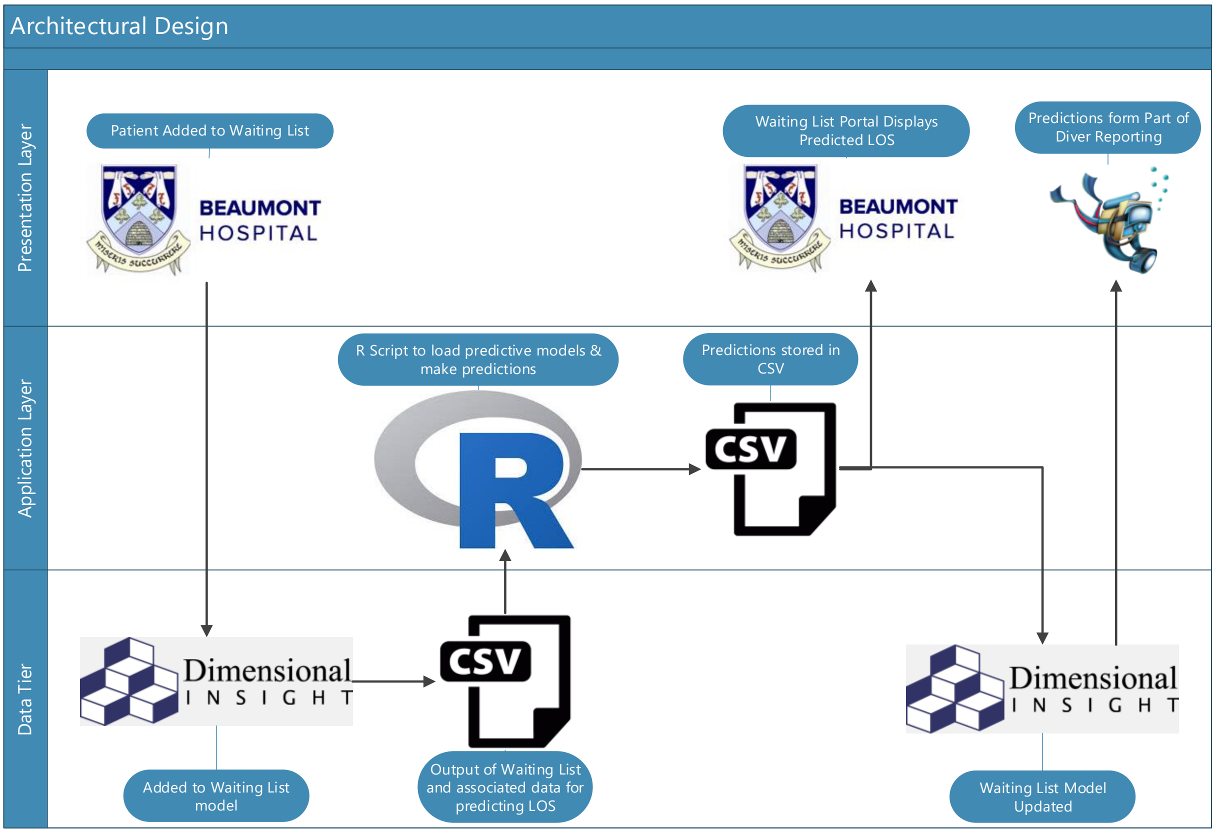


Figure 9 – Architectural Design of Deployment

3.6 Evaluation

Given the business requirements to classify patients' length of stay it is perhaps most appropriate to evaluate the performance of each model based on accuracy and specficity. Initial training and testing will be done with 10-folds cross validation using 80% of the dataset with the final evaluation being on the remaining 20%. Performance on the held-out samples will determine model choice for deployment.

3.7 Deployment

As shown in Figure 9 the proposed deployment would integrate directly with existing hospital user level and data / reporting systems. The hospitals data warehousing and reporting platform is Diver by Dimensional Insights which can interact with R as part of its platform (Dimensional Insight, 2017).

The waiting list portal is hosted on SharePoint with the use of the FlowForma software package (FlowForma, 2019) (waiting list data is also currently mirrored to legacy systems however these are due for decommissioning by the end of 2020).

At a business process level, a user will complete a form within the FlowForma SharePoint portal which adds a patient to the waiting list (stored within SQL Server backend of SharePoint) which is then ingested by the Diver reporting platform. The Diver platform can then collate the necessary information, save it to an excel or CSV file and call the prediction R script to read the file, make the predictions and save an updated CSV with the predictions. This is then ingested and updates both the FlowForma front end for viewing of the patient's waiting list details and the waiting list model within Diver for reporting to end users from the DivePort or NetDiver user interfaces. A sample of the deployment script will be included within the configuration manual of this research project which is subject to amendment based on input from the subject hospital's IT team.

3.8 Conclusion

The research methodology supports a thorough approach with clustering of procedures attempted to enhance predictive model performance and was envisaged to result in different model types being deployed for different clusters based on performance evaluation. Ultimately the inclusion of the mode in the summary statistics available for model building removed the benefit of clustering. The dataset request was heavily influenced by expert clinical guidance (Appendix 1 – Expert Interview Notes) and through the literature review and is expected to be appropriate to address the research requirements, together these also address the research in to the current state of the field and engagement of clinical expert stakeholders. Section 3.4, in particular the discussion of the correlation and importance values with the goal of identifying the influential factors on length of stay which addresses the research objective 2 from Table 1 – Research Objectives.

4 Implementation of Patient Length of Stay Predictive Models

4.1 Introduction

Several predictive models were developed in order to address the classification task. To provide a comprehensive approach the methods selected covered both bagging and boosting sampling methods and ensemble methods. The classification was treated as a two-class problem. From an operational management perspective this was appropriate for the expected usage given that a patient will either stay the up to the expected or typical length of stay (the mode for that procedure) or they will exceed this, with patients who exceed the typical length of stay likely to be the ones which present scheduling and bed capacity issues for procedures planned for the subsequent days.

4.2 Implementation of Classification Models

Each model was trained with 3 repeats using repeated cross validation with 10 folds. Additionally, each model was trained using a variety of tuning parameters in order to identify the best performing model. Each was trained using the caret package (Max Kuhn, 2019b). The caret package provided the option to train for optimisation of kappa value or accuracy, both were trialled in this exercise for relevant models to assist in best model identification.

4.2.1 Implementation of Random Forest Model

A random forest was built using the ranger (Wright, Wager and Probst, 2019) package. The forest was modelled with 500 trees. Both the kappa and accuracy optimised models had a model with an mtry (number of features per node split) of 2, a split rule based on gini and a minimum node size of 1 identified as the best fitting model.

4.2.2 Implementation of Naïve Bayes Model

A naïve Bayes model was built using the klaR (Roever *et al.*, 2018) package. The model was trained for prioritisation of kappa and accuracy for comparative purposes. Both the optimisation for kappa and accuracy converged on a best fit of 0 Laplace correction, kernel use and a bandwidth adjustment of 1.

4.2.3 Implementation of an Artificial Neural Network

A neural network model (Ripley and Venables, 2016) was built using an extensive variety of values for the parameters for both the size of the hidden layer and for the decay of the weights in order to identify the most effective model. The accuracy focused build presented a best fit of size 20 and decay of 1e-07 with the kappa focused model providing 20 and 0.1 respectively as best fit.

4.2.4 Implementation of Support Vector Machine

A support vector machine using the radial basis kernel from the kernlab (Karatzoglou *et al.*, 2018) package was developed with a tuneLength parameter of 9 to ensure randomisation impact was minimised in the selection of an appropriate sigma value. Pre-processing of centering and scaling, as provided within the caret package, was used. The kappa optimised model provided a best fit of sigma 0.2133806 and C of 32. By contrast the accuracy optimised model provided a best fit of 0.2064321 and C of 16.

4.2.5 Implementation of Gradient Boosted Decision Tree

A gradient boosted decision tree was developed using the XGBoost (Chen *et al.*, 2016) package. An extensive combination of parameters was tested to assess and determine the most effective model which included max depth of tree, up to 100 rounds of passing the data, subsampling and by tree column sampling. The kappa and accuracy optimised models were in agreement with max depth of 10, eta of 0.01, column sampling per tree of 0.4 and with minimum child weight of 1.

4.2.6 Implementation of Generalized Linear Model

A generalized linear model (R-Core, 2019) from the base R stats package was developed for class prediction with a cut-off value of 0.50 or lower being classed as "Good" and predictions above 0.50 classified as "Long Stay"

4.2.7 Implementation of C5.0 Decision Tree Model

A C5.0 (Kuhn *et al.*, 2018) decision tree model was developed with a number of parameter options for winnow (feature selection) and trials (number of boosting iterations) set and cycled through. The accuracy and kappa optimised models both provided a best fit with 20 trials with feature sub setting.

4.2.8 Implementation of a Gradient Boosted Ensemble Model

In order to assess whether an ensemble model may improve classification performance an XGBoost model was developed. In order to maintain the integrity of the test set, while still providing enough data to train the component models and the overall ensemble model, the training set was split in two to train both stages of the ensemble model build. The first half was used to train the respective components of the ensemble model (random forest, naïve Bayes, neural network, support vector machine, XGBoost decision tree, and C50 decision tree) with each component model taking their tuning parameter from the known best fit of the previous training for the respective model. The models are then asked to predict (probabilities where possible, classifications are used where necessary) for the other half of the training set. These are then used to train an XGBoost ensemble model which incorporates the dependent variable from the second half of the training set. This methodology while complex allows a fair comparison of performance against the individual models. Like the previous XGBoost model the accuracy and kappa optimisations were in agreement with max depth of tree of 8, eta of 0.01, columns sampling by tree of 0.4, minimum child weight of 10 and subsample of 1.

4.3 Conclusion

Model fitting was quite computationally expensive and time consuming, this was in part due to the comprehensive approach taken to attempt to identify the most appropriate hyperparameters for each model. However, model build would be expected to be an infrequent occurrence once the solution is deployed with increased training data availability and enhancement of patient post-operative pathways likely to be motivating factors in such instances. The computation time impact of the ensemble was minimised by using the hyperparameters of the previously identified best fit, avoiding a second tuning grid search for each model. This comprehensively addresses research objective 3 - incorporating 3(a) to 3(h) - from Table 1 – Research Objectives.

5 Evaluation and Results

5.1 Introduction

The predictive model performance will be compared using established metrics to identify the most appropriate model for deployment. Table 3 is amended from typical predictive model assessment and informs the formulas below;

Table 3 – Confusion Matrix

	Actual	
Prediction	Good	Long Stay
Good	A	B
Long Stay	C	D

$$Sensitivity = \frac{A}{A + C}$$

$$Specificity = \frac{D}{B + D}$$

$$Balanced\ Accuracy = \frac{Sensitivity + Specificity}{2}$$

The table and formulas above are a subset of the metrics available within the caret package confusion matrix function (Max Kuhn, 2019a).

5.2 Evaluation and Results

The models with the highest percentage balanced accuracy are shown in Table 4 with the optimisation metric which produced the model noted.

Table 4 – Performance Metrics for Predictive Models

Predictive Model	Optimisation Metric	Sensitivity	Specificity	Balanced Accuracy
Random Forest	Accuracy	75.05%	79.45%	77.25%
Naïve Bayes	Accuracy	83.91%	42.04%	62.98%
Neural Network	Kappa	70.74%	75.37%	73.06%
Support Vector Machine	Accuracy	74.22%	73.24%	73.73%
Gradient Boosted Decision Tree	Accuracy	72.32%	82.45%	77.38%
Generalised Linear Model	Kappa	73.55%	62.95%	68.25%
C5.0 Decision Tree	Accuracy	74.34%	80.72%	77.53%
Gradient Boosted Decision Tree Ensemble	Kappa	72.05%	80.07%	76.06%

Many of the models perform at a similar level with the Naïve Bayes model perhaps standing out as the poor performer along with the generalised linear model, albeit to a lesser extent. The best performing models, in terms of balanced accuracy, are the Gradient Boosted Decision Tree, C5.0 Decision Tree, the Random Forest and the ensemble model. Within this subset the decision regarding which model is best for deployment must be considered within the context of the domain problem.

For hospital management the potential biggest potential impact is to identify patients most likely to exceed the typically expected length of stay and determine if interventions can be put in place to reduce the variance and inform enhanced patient management strategies. Considering this, the specificity metric is the most relevant for model choice with consideration to be made for balanced accuracy. Given the relatively small differences between the 3 models in terms of balanced accuracy the specificity of the gradient boosted decision tree being the maximum value becomes the most appropriate candidate for deployment within the business context.

5.3 Conclusion

There is a justifiable argument for either the C5.0 decision tree model or Gradient Boosted decision tree model to be the deployed option. Given the operational intent the additional identification of long stay patients is an acceptable trade-off for the similar decrease in identification of normal patient length of stay. As such the deployment will be based on the gradient boosted decision tree due to its higher performance in specificity. This addresses the evaluation and selection of a predictive model for deployment which is research objective 4 in Table 1 – Research Objectives.

6 Discussion

6.1 Introduction

There are several elements to consider in the retrospective review of the research project which primarily fall in to three categories; data pre-processing methods, research reflection and potential applications and research and data limitations.

6.2 Data Pre-Processing Methods Discussion

As is expected throughout a research project, and as defined within the methodology, there has been several reviews of the project and the data pre-processing methods ultimately chosen were reflective of this iterative process. The initial use of hierarchical and K means clustering are an excellent example of this. This was an attempt to provide the predictive models with some grouping or indication of underlying variability of each procedure without the need to provide individual procedure types, which would have been a factor with hundreds of levels which led to concerns of potentially overfitted models.

The decision to use principal components analysis for both outpatient and unscheduled inpatient history was reflective of the nature of the data. In both instances without PCA analysis it would have required training the model with dozens of additional columns, many of which carried a zero value. While this requires an additional step for deployment (the columns need to be reduced using the relevant PCA predictions before being presented for prediction) it has reduced the model complexity significantly. It is also an inherent reflection of the data available for the review, with no specific comorbidity flags available within the hospital dataset, it is necessary to infer this information from the outpatient and inpatient background for each patient.

The decision to use the mode of the length of stay for each procedure is also reflective of the iterative nature of the project. The mode and its difference from the median and mean was shown to be significantly more influential than the k means cluster group on reviewing the importance in a random forest which was generated for review. As such the cluster group variable was dropped and the mode information included. This improved predictive model performance and brought the best performing models from balanced accuracies in the mid to high 60s to the results in

Table 4 with several models exceeding 75% balanced accuracy.

6.3 Research Reflection and Potential Applications

While the research itself is part of a growing interest in the application of machine learning methods being applied in various ways to healthcare around the world there was a lack of similar material available from Irish or European based researchers. As a result, many of the research papers were from sources around the world with different socio-demographic issues, healthcare systems and associated delivery models and associated information technology resources. Many of the research papers reviewed had data available from electronic patient records or the benefit of specific datasets related to the speciality or procedure group of interest, many of which used data on a much smaller cohort of patients. Additionally, this project attempted to move away such groupings and look at the broader service delivery for a wide range of elective admissions.

Given a final dataset of less than half the size of the MIMIC-III (Johnson *et al.*, 2016) dataset it is encouraging to see the better performing models in this research perform at a similar level to the research conducted on that dataset (Gentimis *et al.*, 2017). While the dataset and approach are different, which makes a direct comparison difficult, this research project also outperforms several within the reviewed literature which focused on a specific cohort of patients. The research focused on diabetic patients (Morton *et al.*, 2014) was outperformed by this research project by approximately 10% in balanced accuracy despite the inferred nature of patient background history and the less cohort specific nature. The research specific to cardiology patients (Tsai *et al.*, 2016) was outperformed at an approximately similar level, which again is without the cohort or condition specific selection of patient cases for the research.

Importantly, this research project shows what can be achieved while desirable information was not available in all aspects. This research project, I feel, exemplifies what can be achieved considering gaps in information which is of particular relevance in healthcare given the underlying uncertainty involved in the process of healthcare service delivery.

While the hyperparameter searches provided with the use of tuning grids were quite extensive it does not necessarily exclude the identification of better performing ones or even better performing models outside those considered within this research. This project attempted to cover many of the most common predictive model types with a linear model and a variety of boosting and bagging methodologies. Deep learning methods were not attempted, and it is possible these may have provided enhanced predictive performance although they may have required significantly more data and specific hardware not available to this research. The benefit of using more explainable approaches should also be considered in contrast to deep learning methods. While a boosted decision tree is more complex, it has roots in a decision tree which is quite explainable and providing a context and discernible explanation of the boosting process for any interested stakeholders is expected to be quite achievable.

The eventual removal of the k means clustering due to the lack of benefit within this research project does not necessarily mean that it is not necessary or desirable for other similar research projects. It was simply not influential towards the end of the project's iterative process when the mode and its differences from median and mean proved to be much more important predictors and helped improve the predictive performance of all the models. At the very least the scatter plot of k means clusters, Figure 3, against the procedure mean and median lengths of stay showed there is some potential use of clustering within this area of research as the groups are visually identifiable with the use of colour.

The performance of the selected model provides a significant insight with over 4 of every 5 long stay patients correctly identified within the test set.

In terms of the application of this project's output. Initial application will be within the hospital which provided the dataset with predictions made as patients are added to the waiting list for surgery. This has the potential to inform management and scheduling strategies to better align elective scheduling with the post procedure needs of the patient. This has the potential to influence which procedures are performed on which days (elective procedures are currently typically provided Monday to Friday only), as well as highlighting patient dependency post procedure which may prove useful in the rostering of nursing staff where experience and skill mix are important considerations within the rostering process for each shift. The models and deployment scripts will be provided to the hospital for integration into the workflow set out in Figure 9.

6.4 Research and Data Limitations

Without the benefit of an electronic health record it has not been possible to have as an extensive and accurate a dataset as would be desired, choices made in the design and implementation phases have reflected this limitation.

In particular, the data social history (i.e. database records pertaining to home care package and long-term care / nursing home applications) was not considered of appropriate quality, by the management information team in the hospital, to include for this analysis. The entire dataset for this is less than 4,000 records so as a maximum it would have matched less than 20% of the final training set used for the research. Nonetheless it would have been preferable to have a thorough and reliable dataset for these patients to ascertain impact, if any, on length of stay.

The volume of inpatient activity is approximately 5,000 per year which resulted in a final dataset of just over 27,000 cases post data pre-processing. In an ideal scenario model building would have significantly more cases on which to base the analysis and predictive models.

I feel the steps taken as regards to inpatient and outpatient background history are appropriate within the context of the dataset available. It would have been preferable to have a dedicated binary flag for many of the comorbidities as outlined by the clinical specialists in Appendix 1 but this is not currently available.

It is important to note that aspects of the intended final deployment are outside the scope of this project. The project is intended to facilitate the proposed deployment with the development of the appropriate predictive models and R scripts for the deployment outlined in Figure 9 – Architectural Design of Deployment.

6.5 Conclusion

This research project, I feel, significantly encourages further research on the topic to assess if a larger dataset both in terms of patient cases and patient background information such as social history and specific comorbidity flags may produce better performing predictive models which may enhance the application of the solution as well as providing a strategy with which to approach similar or related research questions.

7 Conclusion and Future Work

The changes in coming years as regards electronic patient records has the potential to enhance this research significantly but I feel this research project has, at the very least, provided an incentive to investigate other areas of potential application as well as expand its use to other hospitals in Ireland.

This research project has identified the current state of the art in applications of this type while being informed by relevant clinical expertise. It has identified age, procedure variability and an inference of patient background history as significant contributors to patient length of stay variance, addressing Research Question 2 comprehensively. It has, within the context of the dataset available, developed and evaluated 8 predictive models and identified a gradient boosted decision tree as the most appropriate for the intended business application and led to the development of an integration pathway for this application in one of Ireland's largest acute hospitals, which addresses Research Question 1.

The insights and predictive model are to be reviewed within the hospital and deployed within the workflow of the management information platform. This is expected, in the longer term, to enable better informed elective procedure scheduling and significantly enhance the internal understanding of the influential factors which impact patient length of stay post procedure. Additionally, it may serve as a practical example which enables the exploration of areas where machine learning can be applied within the hospital.

The limitations identified are primarily centred on enhanced and greater volume of data to move away from inference of patient history and provide a greater volume for predictive model training which may enhance performance.

The future research directions are quite extensive with the potential to move beyond this project's pre-elective admission predictions to a variety of solutions which provide a comprehensive suite of insights to both clinical staff and hospital management.

Length of stay prediction of non-elective (i.e. emergency admissions) could better inform planning to curtail elective scheduling in the days following an influx of admissions which are projected to have a longer length of stay. The natural follow on from both this suggestion and this research project would be the updating of these predictions at regular, perhaps daily, intervals following admissions to identify influential factors during a patients length of stay in relation to laboratory tests, diagnostic imaging and for example sentiment analysis of theatre notes post operatively to assess impact on length of stay.

Prediction of social interventions required for discharge – can patients who will require home supports or long-term care be identified at or close to the point of admission, for example? With this insight the interventions could be sought at much earlier stages in the patients' journey enhancing care provision for both individual patients who spend less time in acute hospital setting and the collective patients in the hospital's catchment area who benefit from enhanced service provision and better bed availability.

Overall, I feel this research project, at the very least, shows what can be achieved within the confines of Irish healthcare as it currently stands and provides a meaningful insight into the research problem. I hope to follow on with further research in related areas to enhance and build on the research in the future, particularly in the specific areas mentioned.

8 Acknowledgements

I would like to thank my research supervisor Dr Catherine Mulwa for her support throughout this project.

On a personal level the support of my wife, Claire, and my son, Harrison, throughout this project has been amazing and without them I'd have struggled to get through this process.

On a professional level I would like to thank Mr Darragh Moneley, Prof Chris Thompson, Prof Arnold Hill, Ms Fiona McNally, Ms Claire Noonan and Ms Barbara Keogh Dunne for their professional opinions and input on this project. Additionally, Stephen Winder and Rosita Guildea from the IT and HIPE teams respectively within Beaumont Hospital were vital in the assembling of the necessary data to facilitate the research on live data. Ms Valerie Caffrey, Chief Operating Officer for Beaumont Hospital, was particularly supportive of the research project goals and authorised access to the relevant datasets required to complete the project.

9 References

Aghajani, S. and Kargari, M. (2016) 'Determining Factors Influencing Length of Stay and Predicting Length of Stay Using Data Mining in the General Surgery Department', *Hospital Practices and Research*, 1(2), pp. 51–56. doi: 10.20286/hpr-010251.

Alahmar, A., Mohammed, E. and Benlamri, R. (2018) 'Application of Data Mining Techniques to Predict the Length of Stay of Hospitalized Patients with Diabetes', in *2018 4th International Conference on Big Data Innovations and Applications (Innovate-Data)*. IEEE, pp. 38–43. doi: 10.1109/Innovate-Data.2018.00013.

Azari, A., Janeja, V. P. and Mohseni, A. (2012) 'Predicting Hospital Length of Stay (PHLOS) : A multi-tiered data mining approach', in *Proceedings - 12th IEEE International Conference on Data Mining Workshops, ICDMW 2012*. IEEE, pp. 17–24. doi: 10.1109/ICDMW.2012.69.

Baek, H. *et al.* (2018) 'Analysis of length of hospital stay using electronic health records: A statistical and data mining approach', *PLoS ONE*. Edited by T. Abe, 13(4), p. e0195901. doi: 10.1371/journal.pone.0195901.

Chen, T. *et al.* (2016) *xgboost: Extreme Gradient Boosting*, *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16*. New York, New York, USA: ACM Press. doi: 10.1145/2939672.2939785.

Chuang, M. Te *et al.* (2016) 'The Identification of Prolonged Length of Stay for Surgery Patients', in *Proceedings - 2015 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2015*. doi: 10.1109/SMC.2015.522.

Department of Health (2018) *Evaluation of the 'Pilot Implementation of the Framework for Safe Nurse Staffing and Skill-Mix'*. Available at: https://health.gov.ie/wp-content/uploads/2018/04/Evaluation-of-the-'Pilot-Implementation-of-the-Framework-for-Safe-Nurse-Staffing-and-Skill-Mix'.pdf (Accessed: 3 March 2019).

Dimensional Insight (2017) *Integrating Statistics into Diver with R, 2017-06-06*. Available at: https://www.dimins.com/blog/2017/06/06/integrating-statistics-diver-r/ (Accessed: 21 July 2019).

Elings, J. *et al.* (2015) 'What preoperative patient-related factors predict inpatient recovery of physical functioning and length of stay after total hip arthroplasty? A systematic review', *Clinical Rehabilitation*, 29(5), pp. 477–492. doi: 10.1177/0269215514545349.

Fayyad, U., Piatetsky-Shapiro, G. and Smyth, P. (1996) 'The KDD process for extracting useful knowledge from volumes of data', *Communications of the ACM*. ACM, 39(11), pp. 27–34. doi: 10.1145/240455.240464.

FlowForma (2019) *Digital Process Automation Software I FlowForma*. Available at: https://www.flowforma.com/ (Accessed: 21 July 2019).

Gentimis, T. *et al.* (2017) 'Predicting Hospital Length of Stay Using Neural Networks on MIMIC III Data', in *2017 IEEE 15th Intl Conf on Dependable, Autonomic and Secure Computing, 15th Intl Conf on Pervasive Intelligence and Computing, 3rd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress(DASC/PiCom/DataCom/CyberSciTech)*. IEEE, pp. 1194–1201. doi: 10.1109/DASC-PICom-DataCom-CyberSciTec.2017.191.

Hachesu, P. R. *et al.* (2013) 'Use of Data Mining Techniques to Determine and Predict Length of Stay of Cardiac Patients', *Healthcare Informatics Research*, 19(2), p. 121. doi:

10.4258/hir.2013.19.2.121.

Johnson, A. E. W. *et al.* (2016) 'MIMIC-III, a freely accessible critical care database', *Scientific Data*. Nature Publishing Group, 3(1), p. 160035. doi: 10.1038/sdata.2016.35.

Karatzoglou, A. *et al.* (2018) *Kernel-Based Machine Learning Lab [R package kernlab version 0.9-27], 2018/08/10*. Comprehensive R Archive Network (CRAN). Available at: https://cran.r-project.org/web/packages/kernlab/index.html (Accessed: 3 July 2019).

Kelly, M. *et al.* (2012) 'Factors predicting hospital length-of-stay and readmission after colorectal resection: a population-based study of elective and emergency admissions', *BMC Health Services Research*, 12(1), p. 77. doi: 10.1186/1472-6963-12-77.

Kuhn, M. *et al.* (2018) *C50: C5.0 Decision Trees and Rule-Based Models, 2018/05/22*. Morgan Kaufmann Publishers. Available at: https://cran.r-project.org/package=C50 (Accessed: 3 July 2019).

Kumar, A. and Anjomshoa, H. (2018) 'A Two-Stage Model to Predict Surgical Patients' Lengths of Stay from an Electronic Patient Database', *IEEE Journal of Biomedical and Health Informatics*, pp. 1–1. doi: 10.1109/JBHI.2018.2819646.

Lamiri, M., Grimaud, F. and Xie, X. (2009) 'Optimization methods for a stochastic surgery planning problem', *International Journal of Production Economics*. Elsevier, 120(2), pp. 400–410. doi: 10.1016/j.ijpe.2008.11.021.

Li, X. *et al.* (2017) 'Scheduling elective surgeries: the tradeoff among bed capacity, waiting patients and operating room utilization using goal programming', *Health Care Management Science*, 20(1), pp. 33–54. doi: 10.1007/s10729-015-9334-2.

Liu, P. *et al.* (2006) 'Healthcare data mining: Prediction inpatient length of stay', in *IEEE Intelligent Systems*. IEEE, pp. 832–837. doi: 10.1109/IS.2006.348528.

Max Kuhn (2019a) *confusionMatrix function | R Documentation*. Available at: https://www.rdocumentation.org/packages/caret/versions/6.0-84/topics/confusionMatrix (Accessed: 21 July 2019).

Max Kuhn (2019b) *CRAN - Package caret*. Available at: https://cran.r-project.org/web/packages/caret/ (Accessed: 3 July 2019).

Morton, A. *et al.* (2014) 'A Comparison of Supervised Machine Learning Techniques for Predicting Short-Term In-Hospital Length of Stay among Diabetic Patients', in *2014 13th International Conference on Machine Learning and Applications*. IEEE, pp. 428–431. doi: 10.1109/ICMLA.2014.76.

National Treatment Purchase Fund (2019a) *Inpatient/Day Case - Active as at 31/01/2019*. Available at: https://www.ntpf.ie/home/pdf//2019/01/nationalnumbers/in-patient/National01.pdf (Accessed: 3 March 2019).

National Treatment Purchase Fund (2019b) *Inpatient / Day Case TCI as at 31/01/2019*. Available at: https://www.ntpf.ie/home/pdf/2019/01/tci/TCI_National01_na.pdf (Accessed: 3 March 2019).

Ong, P. H. and Pua, Y. H. (2013) 'A prediction model for length of stay after total and unicompartmental knee replacement', *Bone and Joint Journal*, 95 B(11), pp. 1490–1496. doi: 10.1302/0301-620X.95B11.31193.

Pakzad, H. *et al.* (2014) 'Factors associated with longer length of hospital stay after primary elective ankle surgery for end-stage ankle arthritis', *Journal of Bone and Joint Surgery - Series*

A, 96(1), pp. 32–39. doi: 10.2106/JBJS.K.00834.

Quinian, R. and Ross, J. (1993) *C4.5 : programs for machine learning.* Morgan Kaufmann. Available at: https://dl.acm.org/citation.cfm?id=583200 (Accessed: 24 March 2019).

R-Core (2019) *glm function | R Documentation*. Available at: https://www.rdocumentation.org/packages/stats/versions/3.5.3/topics/glm (Accessed: 21 July 2019).

Ripley, B. and Venables, W. (2016) *Feed-Forward Neural Networks and Multinomial Log-Linear Models [R package nnet version 7.3-12], 2016/02/02*. Comprehensive R Archive Network (CRAN). Available at: https://cran.r-project.org/web/packages/nnet/index.html (Accessed: 3 July 2019).

Roever, C. *et al.* (2018) *klaR: Classification and Visualization, 2018/03/19*. Available at: https://cran.r-project.org/web/packages/klaR/index.html (Accessed: 3 July 2019).

Royal College of Surgeons Ireland (2017) *The national launch of NQAIS Clinical occurred on October 24th 2017 at the National Acute Medicine Programme Conference - Royal College Surgeons in Ireland*. Available at: http://www.rcsi.ie/ncps_nqais (Accessed: 21 July 2019).

Solon, J. G., Coffey, J. C. and McNamara, D. A. (2013) 'Length of stay following elective surgery can we improve efficiency?', *Irish Medical Journal*, 106(6), pp. 179–81. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23909155 (Accessed: 1 December 2018).

St James's Hospital (2018) *EPR PROJECT OAK ELECTRONIC PATIENT RECORD*. Available at: http://www.stjames.ie/NewsBanner/ProjectOak/Project Oak patient record leaflet.pdf (Accessed: 24 March 2019).

Steele, R. J. and Thompson, B. (2019) 'Data Mining for Generalizable Pre-admission Prediction of Elective Length of Stay', in *2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC)*. IEEE, pp. 0127–0133. doi: 10.1109/CCWC.2019.8666598.

Al Taleb, A. R., Abul Hasanat, M. H. and Khan, M. B. (2017) 'Application of Data Mining Techniques to Predict Length of Stay of Stroke Patients', in *Informatics Health and technology International conference*. IEEE, pp. 1–5. doi: 10.1109/ICIHT.2017.7899004.

Tsai, P. F. J. *et al.* (2016) 'Length of Hospital Stay Prediction at the Admission Stage for Cardiology Patients Using Artificial Neural Network', *Journal of Healthcare Engineering*, 2016, pp. 1–11. doi: 10.1155/2016/7035463.

Turgeman, L., May, J. H. and Sciulli, R. (2017) 'Insights from a machine learning model for predicting the hospital Length of Stay (LOS) at the time of admission', *Expert Systems with Applications*. Pergamon, 78, pp. 376–385. doi: 10.1016/j.eswa.2017.02.023.

Ucar, I., Smeets, B. and Azcorra, A. (2017) 'simmer: Discrete-Event Simulation for R'. Available at: http://arxiv.org/abs/1705.09746 (Accessed: 12 April 2019).

Wright, M. N., Wager, S. and Probst, P. (2019) *ranger: A Fast Implementation of Random Forests, 2019-03-07*. Available at: https://cran.r-project.org/web/packages/ranger/index.html (Accessed: 3 July 2019).

Appendix 1 – Expert Interview Notes

Prof Chris Thompson, Consultant Endocrinologist, Clinical Director – Medical Directorate, Beaumont Hospital

Prof Thompson highlighted, from his experience, the primary clinical issues which are likely to lead to longer lengths of stay in patients admitted for elective procedures; He outlined that patients with diabetes, both insulin and non-insulin dependent with different levels of issues, a history of heart disease and / or blood pressure issues, patients with a history of stroke, patients with Parkinsons and patients with issues mobilising, such as those wheelchair bound either pre admission or post procedure were all likely to see increased lengths of stay following a planned procedure.

Prof Arnold Hill, Consultant Surgeon, Head of School of Medicine and Professor of Surgery, RCSI & Beaumont Hospital

Prof Hill discussed several potential issues including patients with high BMI's or which have been on the waiting list for a long time, or both. He highlighted that patients with cardiovascular and respiratory issues, perhaps identified by having seen either specialty in the preceding 2 years, can also cause complications and additionally highlighted that patients who have had previous surgery within the previous 3 years can also have longer lengths of stay.

Mr Darragh Moneley, Consultant Vascular Surgeon, Clinical Director – Surgical Directorate, Beaumont Hospital

Mr Moneley discussed a variety of issues that can impact a patient's length of stay many of which are known preadmission but may or may not be available in suitable format for the type of research proposed. He highlighted significant co morbidities such as cardiac, diabetes and respiratory issues. He highlighted age and mobility to likely be two significant contributing factors. Additionally, he discussed how patients can have some cognitive decline post-surgery and during their stay can influence their stay with a drop in mini mental state scoring. While this typically recovers often there is social and support issues preventing sending the patient home in this period with a lack of preadmission interventions from social work and arrangement of home support services not possible until the patient is admitted to the acute setting.

Ms Fiona McNally, Enhanced Recovery Nurse – Clinical Nurse Specialist, Coloproctology, Beaumont Hospital

Ms McNally highlighted that, from her experience, the most influential factors centre around the patient's age and social status. She highlighted that older patients take longer to recover from surgery and if there is no support network socially in the form of relatives or existing healthcare community supports in place it can lead to an increase in their length of stay. She highlighted that a significant cardiac history also has the potential to increase length of stay due to complications. The lack of integration of non-acute / step down beds into the acute healthcare environment was also seen as a challenge which can increase length of stay.

Ms Claire Noonan, Clinical Nurse Manager 3, Surgical Directorate, Beaumont Hospital
Ms Noonan felt that the communication and expectation are significant factors. Patients coming in for surgery often expect to be in hospital for much longer than clinically warranted. She highlighted the work of the Enhanced Recovery Nurses in highlighting the patients' recovery pathway and likely expected length of stay prior to admission as a significant improvement on this issue. She highlighted that age can be a factor with some but not necessarily all procedures and that co-morbidities and issues around patient mobility are important factors. The lack of integration and planning with social work and community supports was highlighted to be an issue also.

Ms Barbara Keogh Dunne, Head of Patient Flow, Beaumont Hospital
The conversation with Ms Keogh Dunne was primarily about the most desirable output format for the predictive models. I spoke about the possibility of treating the issue at hand as either a classification or regression problem. Ms Keogh Dunne highlighted that a linear prediction while interesting is only informative to the point of whether the length of stay will be shorter, as long, or longer than expected. She considered point predictions somewhat less relevant especially given the variability of patient care and highlighted that of particular interest are the instances where a patient is predicted to stay longer than the typically expected length of stay (i.e. the average) for a scheduled procedure as this can help inform the care and supports provided to a patient and help ensure that early interventions are provided to the patient to help reduce the length of stay.