Using Machine Learning to Predict Hospital Re-admissions for Diabetic Population

Avishek Choudhury*, Jmaes Yongjik Eakins, Gregory J Richardson, Redmond Sean O'Hanlon

Syracuse University, Syracuse, USA

Abstract

Background: One of the main factors contributing to the healthcare expenses is unplanned hospital re-admissions. Soon after the launch of the Affordable Care Act in 2010, patient readmission has been a significant metric to measure quality of care. According to the School of Public Health, hospitals can save about $2,140 per patient by minimizing unplanned re-admissions. Other studies have shown that 15% to 25% of discharged patients are re-admitted in less than 30 days. Fierce Healthcare reported that in 2011, hospitals spend $41.3 billion to treat unplanned re-admitted patients. However, a study published by Harvard Business Review reported that improved doctor-patient communication and complying to evidence-based care standards could check patient readmission rate by 5%.**Objective:** To address the re-admission concern, this study leverage supervised machine learning algorithms and provides with an efficient prediction model **Materials and Methods**:The data set used in this study consists of 55 patient attributes and 100,000 observations collected from 130 hospitals across USA. This study follows CRISP-DM approach to analyze the data set. This study implements several machine learning algorithms not limited to support vector machine, decision trees, association rule, and clustering. **Results**: Each classifier used gave different outcomes and identified number of medication, number of diagnosis are important predictor of re-admission.**Conclusion**:This study indicates the presence of biases within doctors, resulting in poor diagnosis, a major cause of hospital re-admission. Some research directions can be sought by trying different variable selection techniques such as LASSO or Non-negative Garrote for better subset regressions. Also, in presence of high right censored data, it is interesting to consider some health care cost measures from which it may be possible to statistically estimate the mean population cost for readmission.
Keywords: `Machine learning, predictive analytics, hospital re-admission, diabetes`

1. Introduction

One of the main factors contributing to the healthcare cost is the avoidable patient readmission. Unplanned hospital re-admission has been a significant metric to measure quality of care. According to the School of Public Health, Veterans Administration can save $2,140 per patient by managing patients prone to readmission (Kathleen Carey, 2016). Moreover, studies have shown that 15 to 25 percent of discharged patients are readmitted in less than 30 days. According to the Agency for Healthcare Research and Quality, about 1.8 million patients were readmitted (Anika and Hines, 2014). Fierce Healthcare reported that in 2011, hospitals spend $41.3 billion to treat unplanned readmitted patients (Shinkman, 2014).

*Corresponding author
Email address: `achoud02@syr.edu` (Avishek Choudhury*)

A study published by Harvard Business Review stated that prioritized and effective communication with the patient and complying to evidence-based care standards could check patient readmission rate by 5 percent (Claire Senot, 2015). However, fostering desired communication within a hospital is arduous due to the complexity of the system. This study focuses on predicting patient readmission. Individuals with a high risk of readmission can be provided with alternative preventive measures such as intensive post-discharge care or home care (Davood Golmohammadi, 2015). This study defines patient readmission as the readmission caused due to poor discharge planning resulting in re-occurrence of the treated disease and worsening health condition. When an individual requires readmission within 90 days' post-discharge for the same cause for which she or he was admitted to a hospital in the very first place is termed as the patient readmission. The reason behind considering readmission within 90 days is since the patients during the first three-month post-discharge are susceptible to the diseases and have suicidal behavior among individuals who have a mental disorder (Appleby, 2013).

1.1. ***Alarming Hospital Discharge Concerns***

This section briefly discusses three crucial factors that encourages patient readmission.

1.1.1. ***Premature Discharge***

The preliminary decision any healthcare providers need to take is whether a patient has recovered enough to leave the hospital independently. Poor decision making at this instance deters patient safety, resulting in frequent readmission or death.

"A man died after a hospital failed to treat sepsis" and discharge the patient before time (Ombudsman, 2003).
According to Homeless Link, more than 70% of underprivileged people were discharged without any housing and addressing underlying health conditions (IHSMHL, 2012).

1.1.2. ***Poor Assessment***

Often physically capable patients are not mentally fit enough to cope with their ailment at home. This is very common problem among mentally challenged patients. These patients after discharge often fail adhere to medical recommendations and lose mental health which in turn enhances the likelihood of readmission. Such conditions are common among elderly patients (age 65 and above) who are not capable of independently maintaining their health either due to cognitive or financial constraints.

According to King's Fund, "being discharged without proper support is an invitation to relapse, worsening of the condition and readmission" (Maguire, 2015).
During 2002 and 2012, 3,225 suicides were recorded by The National Confidential Inquiry into Suicide and Homicide by People with Mental Illness, 2014.

1.1.3. ***Absence of Home Care Plans***

Inefficient communication and coordination between hospitals and community healthcare providers is another concern that needs attention. Due to scarcity of home care facilities, discharged patients are left alone at home which leaves the patient susceptible to health deterioration and emergency readmission.

During 2002 and 2012, 3,225 suicides were recorded by The National Confidential Inquiry into Suicide and Homicide by People with Mental Illness, 2014.

To minimize such occurrences, NHS recommended care providers to follow up with their discharged patients within 7 to 21day post discharge and ensure availability of crisis support (Assessment, 2013).

2. Problem Statement

There exist several possible causes responsible for unplanned patient readmission. However, our study does not focus on identifying the responsible cause, but it provides with an efficient prediction model that can be deployed to a clinical scenario and help healthcare units to be prepared for the unavoidable re-admissions and provide alternative care to preventable re-admissions. The proposed model provides healthcare providers with a decision support system to identify individuals prone to readmission and thus minimize early discharge and ensure follow up with the discharge patients.

3. Analysis and Models

This study follows "Cross-Industry Process for Data Mining" (CRISP-DM) methodology for all analysis. It consists of the following steps (see Figure3):

- Business understanding
- Data understanding
- Data preparation – Feature selection and Normalization
- Modeling
- Evaluation
- Deployment

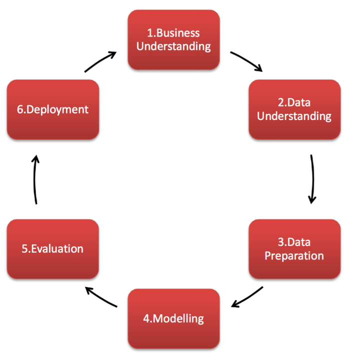

Figure 1: CRISP-DM

3.1. Data Description

The dataset consists of 55 attributes and a sample size of 100,000 instances and represents 10 years of data collected from 130 US hospitals (Avishek, 2018; Strack et al., 2014). The original database contains curtailed, superfluous and noisy information as expected in most of the real-world data (Strack et al., 2014). There were some attributes that could not be treated directly since they had a high percentage of missing values. These features were “weight” (97% values missing), “payer code” (40%) and “medical specialty” (47%). “Weight” attribute was too sparse to be considered and was not included in further analysis. “Payer code” was neglected since it had a high percentage of missing values and it was not considered relevant to the outcome. “Medical specialty” attribute was accounted for analysis, adding the value “missing” in order to account for missing values. Large percentage of missing values of the “weight” attribute can be explained by the fact that prior to the HITECH legislation of the American Reinvestment and Recovery Act in 2009 hospitals and clinics were not required to capture it in a structured format (Strack et al., 2014).

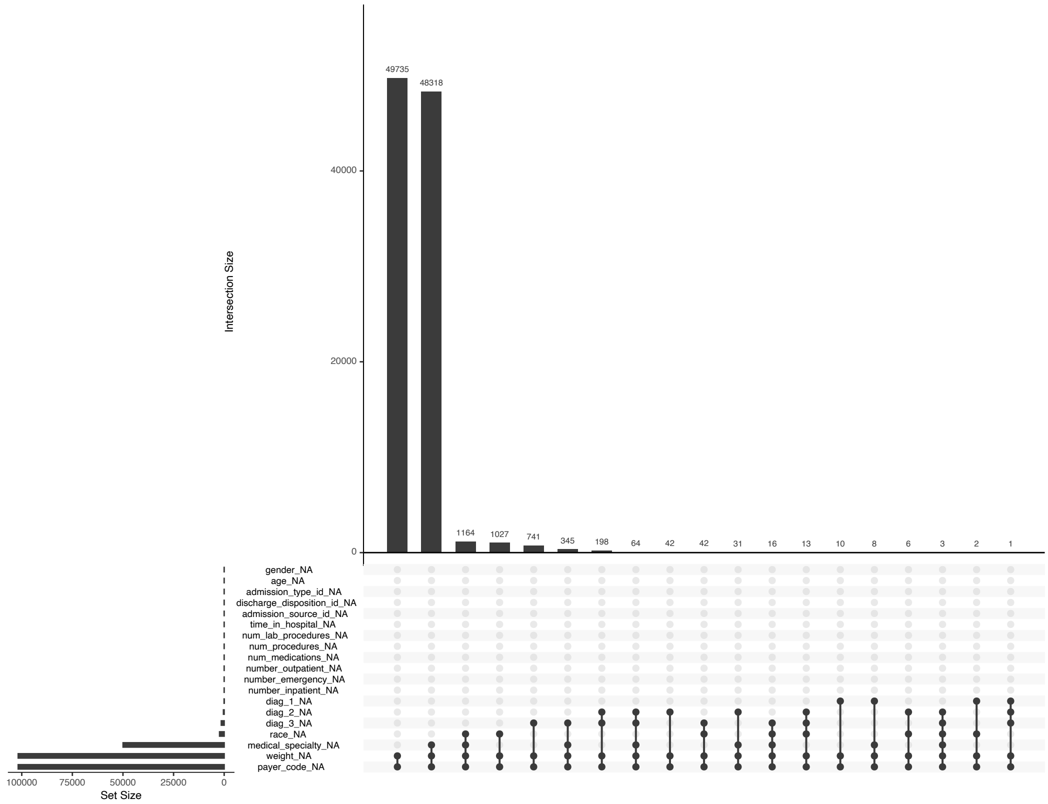

Figure 2: Missing Values

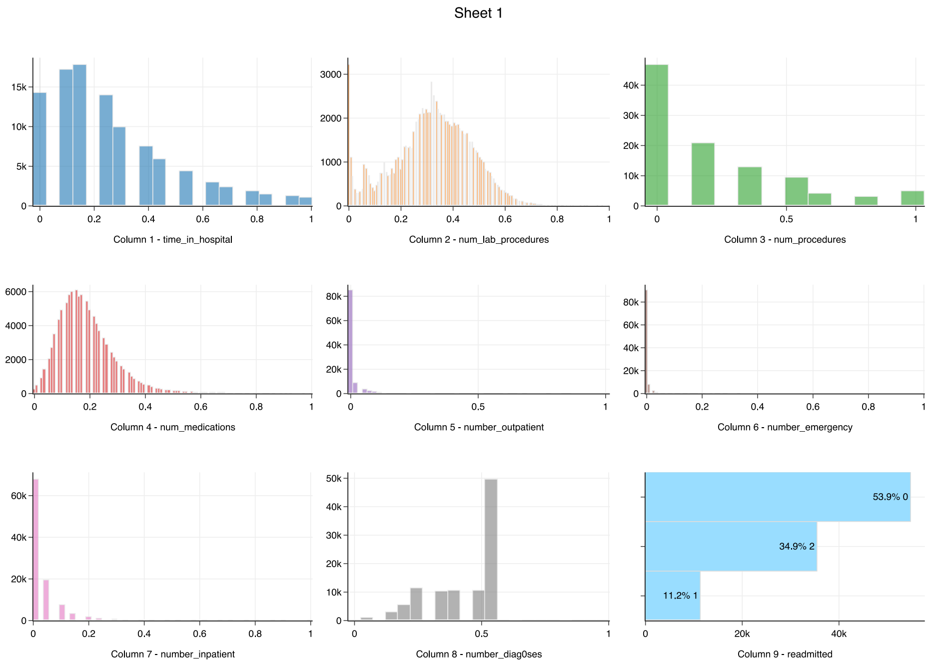

Figure 3: Data Description

3.1.1. Data Visualization

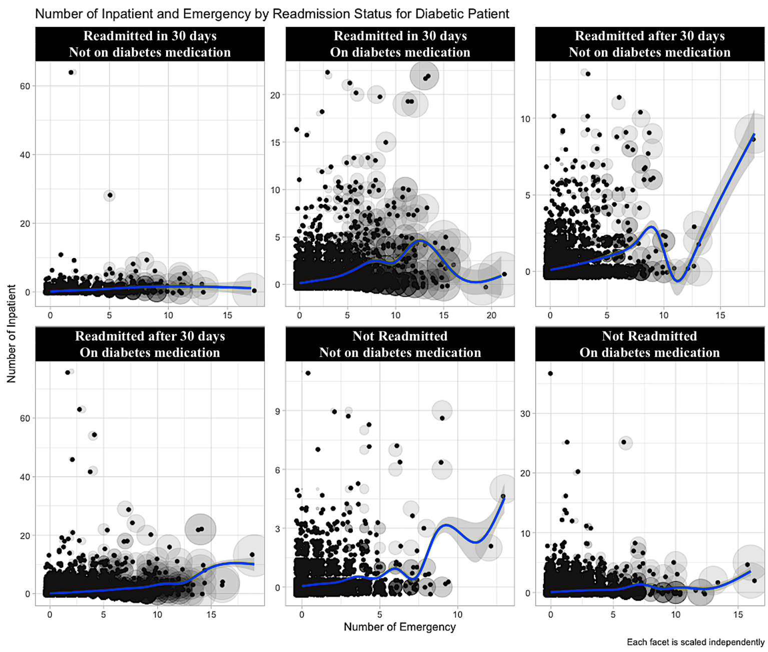

Figure 4: Re-admission rates for Diabetic and Non-diabetic Patients

3.2. Feature Selection

Large datasets hinder the speed of algorithms and even deteriorate classification accuracy (Kohavi and John, 1997). The concern raised due to data size is termed as the minimal-optimal problem (Nilsson, 2007). Our study employs Boruta algorithm and stepwise regression to determine the best features within the dataset.

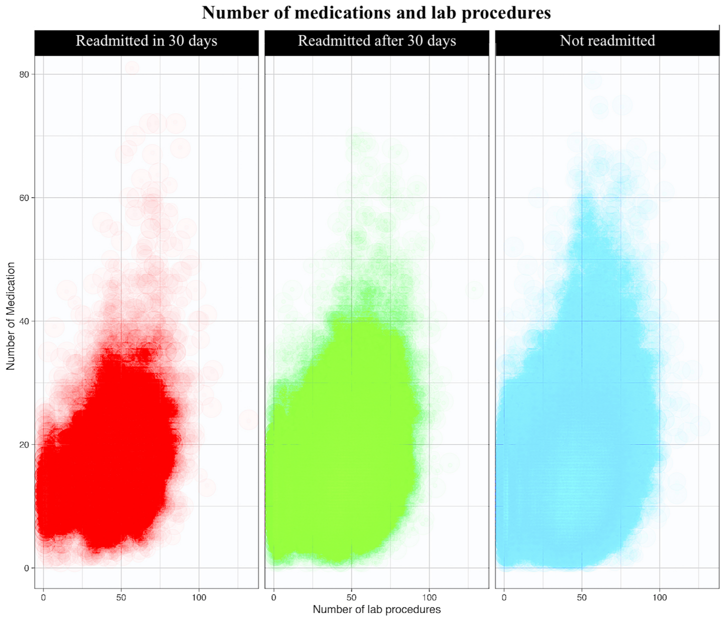

Figure 5: Medication and Procedure

Figure 7 shows Boruta algorithm is a wrapper developed on random forest classification algorithm (Liaw and Wiener, 2002). In this algorithm, the relevance of any attribute is retrieved as the loss of classification accuracy caused due to permutation of attribute values among objects. It calculates the shuffled correlations between the response and the attributes. It also computes the Z-score to determine attributes' relevance by dividing the mean accuracy loss by its standard deviation. In addition to Boruta, stepwise regression was also implemented.

Figure 8 shows Step-wise regression is designed as an automatic computational procedure in which the performance of the regression increases with increase in the input variable (Barnett et al., 1975; Campolongo et al., 2000). Stepwise regression is a different version of the forward selection in which after every step a variable is added, all selected attributes in the model are analyzed to determine any loss in relevance. If an irrelevant variable is found, it is blocked from the model. Stepwise regression mandates two significance levels: One for adding attributes and one for eliminating attributes. The cutoff plausibility for adding an attribute must be less than the cutoff probability for eliminating attributes to avoid an infinite loop trap (Mengchao Wang, 2013).

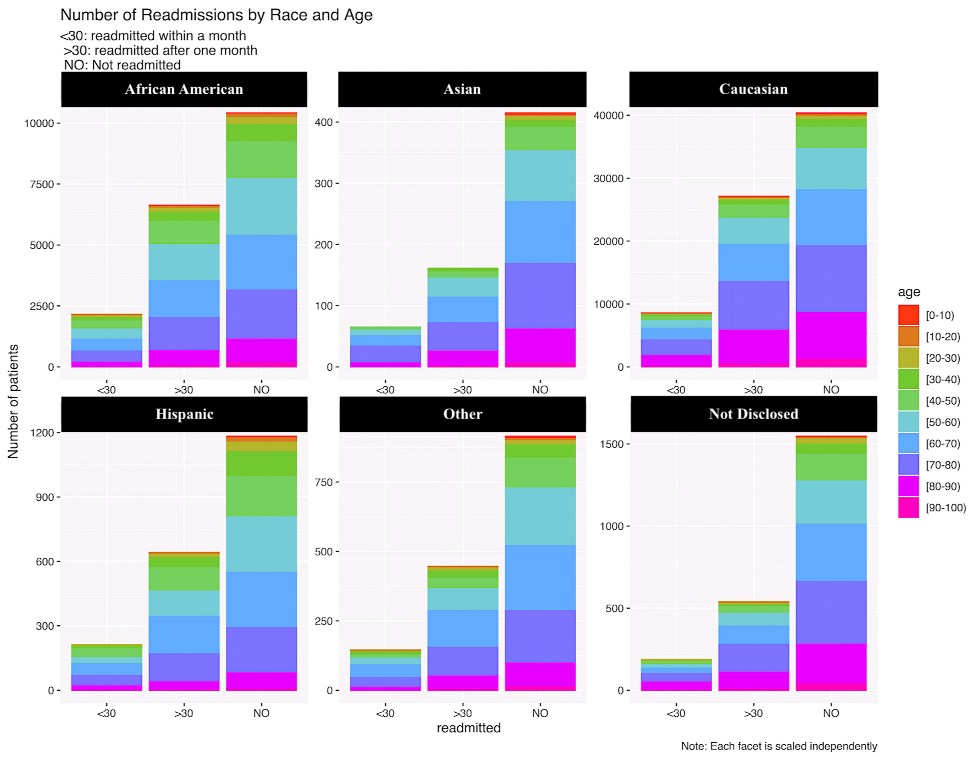

Figure 6: Re-admission by Race

3.3. Models Used

3.4. Association Rules

The Apriori algorithm is an influential algorithm particularly used for mining and analyzing frequent itemsets for boolean association rules. The figure9 shows the top 10 Rules By Lift for Readmittance in 1 Month. A formal definition of association rule can be described as following:

Let $i = (i_1, i_2, i_3, \ldots, i_n)$ be a set of n attributes called *items* and $D = (t_1, t_2, \ldots, t_n)$ be the set of transactions. Every transaction, (t_i) in (D) has a unique transaction ID that contains of a subset of itemsets in (I). A rule obtained using apriori algorithm is defined as an implication, XY where (X) and (Y) are subsets of $I(X, YI)$, and they have no common/shared elements, i.e., (XY). X and Y are the antecedent and the consequent of the rule, respectively.

Association rule generates multiple rules even from a very small database, so to select the most useful and important rule, this study use constraints on various measures of significance such as support, confidence, and lift.

$$Supp(X) = \frac{Number - of - transactions - in - which - X - appears}{Total - number - of - transaction - is - the - proportion - of - transaction} \tag{1}$$

$$Conf(X \longrightarrow Y) = \frac{Supp(X \cup Y)}{Supp(X)} \tag{2}$$

$$Lift(X \longrightarrow Y) = \frac{Supp(X \cup Y)}{Supp(X) \times Supp(Y)} \tag{3}$$

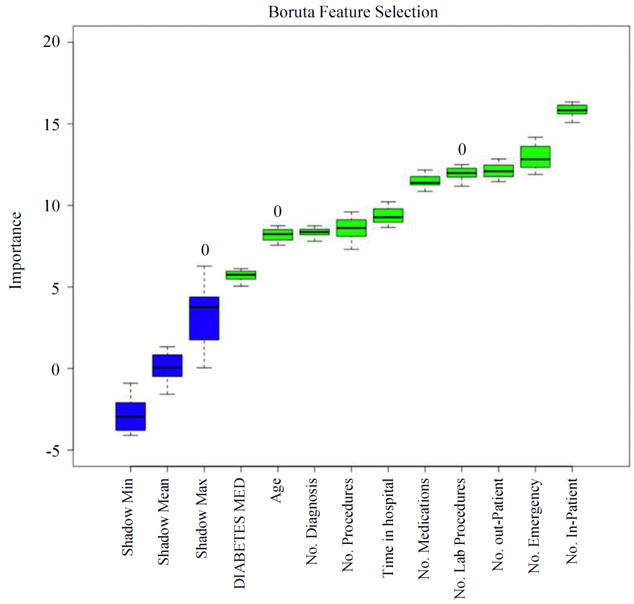

Figure 7: Feature Selection by boruta

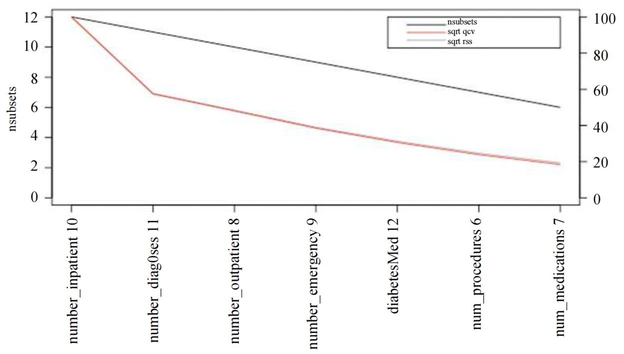

Figure 8: Feature Selection by Step-wise

3.5. Support Vector Machine

A Support Vector Machine (SVM) is a discrete classifier that generates a decision line or plane to segregate different classes using linear, radial or polynomial kernel. This study examines radial and polynomial kernel.

3.6. Multinomial Naïve Bayes

Naive Bayes is a widely used classification technique. It is a probabilistic classifier that uses the Maximum A Posterior decision rule to make classification in a Bayesian setting.

Naïve Bayes are fast and simple algorithm, however, a major limitation of this technique is that it requires the predictors to be independent. But in reality most of the predictors are dependent. This deters Naïve Bayes's classification problem.

3.7. k-Mean Clustering

The k-means clustering algorithm assigns data points to categories, or clusters, by finding the mean distance between data points. It then iterates through this technique in order to perform more accurate classifications

```
> inspect(mod2[1:20])
     lhs                                     rhs             support confidence     lift count
[1]  {num_medications=(10,15],
      number_emergency=[20,25)}          => {readmitted=<30} 3.110839e-05          1 8.714712     3
[2]  {num_medications=(5,10],
      number_inpatient=[15,20)}          => {readmitted=<30} 3.110839e-05          1 8.714712     3
[3]  {num_lab_procedures=(25,50],
      number_inpatient=[15,20)}          => {readmitted=<30} 4.147786e-05          1 8.714712     4
[4]  {number_emergency=[0,1),
      number_inpatient=[15,20)}          => {readmitted=<30} 3.110839e-05          1 8.714712     3
[5]  {age=[30-40),
      number_emergency=[15,20)}          => {readmitted=<30} 3.110839e-05          1 8.714712     3
[6]  {num_medications=(10,15],
      number_emergency=[20,25),
      change=Ch}                         => {readmitted=<30} 3.110839e-05          1 8.714712     3
[7]  {gender=Female,
      num_medications=(10,15],
      number_emergency=[20,25)}          => {readmitted=<30} 3.110839e-05          1 8.714712     3
[8]  {race=Caucasian,
      num_medications=(10,15],
      number_emergency=[20,25)}          => {readmitted=<30} 3.110839e-05          1 8.714712     3
[9]  {num_medications=(10,15],
      number_emergency=[20,25),
      diabetesMed=Yes}                   => {readmitted=<30} 3.110839e-05          1 8.714712     3
[10] {num_medications=(10,15],
      number_emergency=[20,25),
      number_diagnoses=(5,10]}           => {readmitted=<30} 3.110839e-05          1 8.714712     3
```

Figure 9: Top 10 Rules By Lift for Readmittance in 1 Month

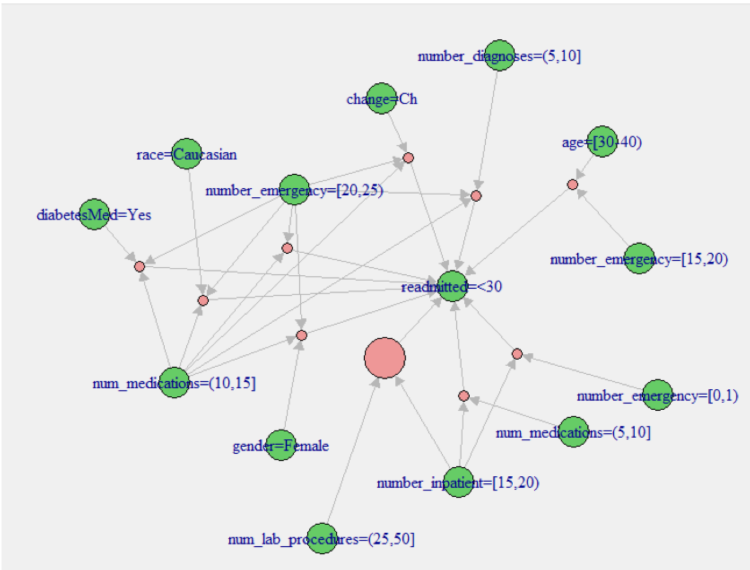

Figure 10: Network Map for Patients Readmitted Within 30 Days

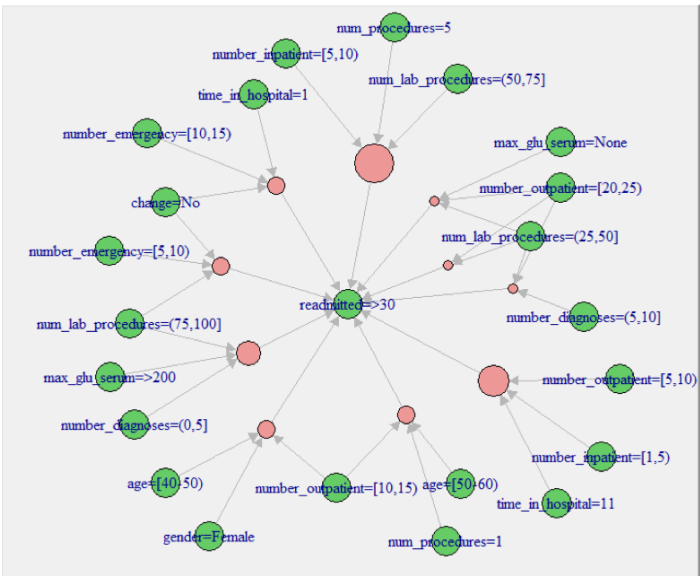

Figure 11: Network Map for Patients Readmitted After 30 Days

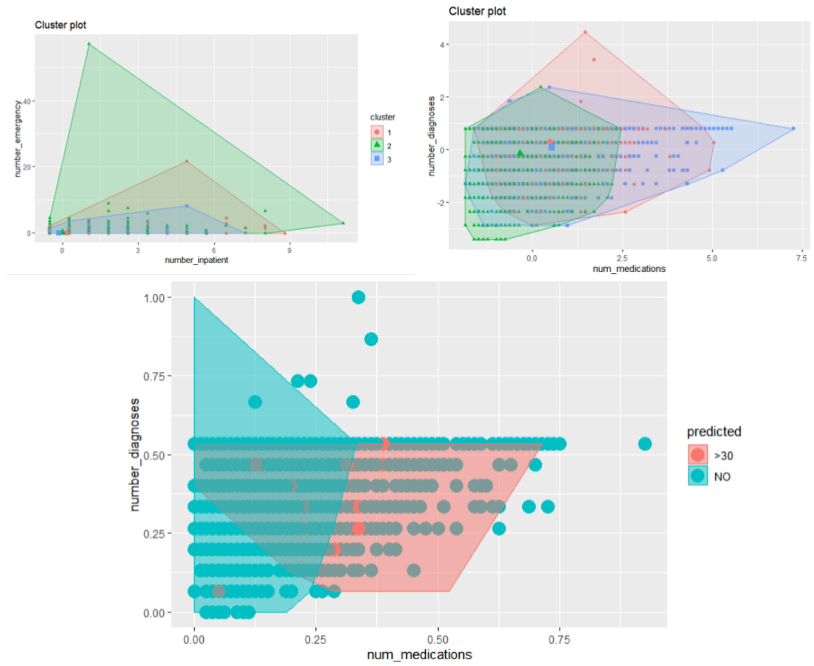

Figure 12: K-mean Clustering

over time. K-mean clustering typically employs different distance metrics. This study uses euclidean distance and cosine similarity. Figure **??** shows k-mean clustering using euclidean distance.

The study also implemented Random Forest algorithm, but yielded error rate of 43.21%.

4. Results

This section contains the best outcome of the analysis. However, Appendix A shows all analysis. Support vector machine and Random forest did not gave good results. SVM generated an accuracy of about 67% where are random forest was found to highly over-fitting (Due to the size of dataset). However, Association rules and data visualization was found to be the two most useful methods to understand the underlying factors causing hospital re-admissions.

Naïve Bayes and Clustering also failed to report acceptable outcomes. In this section, meaning full outcomes generated by Association Rules, Naïve Bayes, clustering and data visualization has been berifly discussed.

4.1. Association Rules

4.1.1. Re-admission within 30 days

Most patients visiting a hospital are effectively treated, discharged, and do not return. But there is a small subset of the population who frequently returns to the hospital. With a dataset as large as this one – nearly 100,000 observations – those who were readmitted in less than 30 days are only a small portion of the population. With that in mind, the Support threshold for the Association-Rules algorithm, where the right hand side was set to "Readmission $<$ 30 days", was set very low. To compensate for the low support minimum, a higher Confidence threshold – 0.9 was used. Ultimate, a support minimum of 0.00003 was used. The model mined 80 unique rules from the data set. Many of the rules were very strong – with Lift measurements of 8.7 or higher. Many of the rules suggested the same intuitive result. Patients who frequent patrons of medical services are likely to be readmitted within a month. Examples of this include patients who visited the ER more than 20 times in the last year, patients who had 15-20 inpatient hospital visits in the previous year, and various other combinations of inpatient/outpatient/emergency medical services over the previous 12 months.

In many of these cases, it is evident that some amount of treatment was performed. Many of the rules include treatment with 11-15 different medications during their hospital stay. When the support was relaxed further – to 0.00001, this number ballooned. In those cases, patients with extended hospital stays – more than 11 days, and subjected to a battery of medications (in some cases as many as 85), were, unsurprisingly, bound to return to the hospital within a month's span.

There were some indications that Caucasians and females with a history of ER visits were more likely to return within a month, but the observation counts are so low, this may simply be an artifact of the population sample.

4.1.2. Re-admission after 30 days

Like the previous model, patients who are readmitted are a small subset of the population. Though the number of patients readmitted within the 1-6 month window is significantly higher than those readmitted within a month, the overall percentage is still quite low. Therefore, the second Association Rules model – which was set up with a right hand side set to "readmission $>$ 30", used a fairly low support threshold – 0.00005. Like before, the confidence minimum was set to 0.9. This model yielded 92 unique rules.

The rules here paint a different picture. Patients here that are likely to be readmitted spent short times in the hospital, but are still frequent returners to the hospital. For example, one rule, with a lift of nearly 2.8, suggests that patients who have frequented the ER 10-15 times in the last year, and who spent a day in the hospital, with no changes made to their medication are liable to return in 1-6 months.

Similarly, there are rules depicting patients where a high number of lab procedures were performed – between 75 and 100 – but very few diagnoses – less than 5. These patients also had a max glucose measurement of 200 – considered High typically by medical professionals.

The rules suggest that patients are returning to the hospital, having a battery of tests done, but leaving with little done in the way of treatment. Looking at the disposition codes for several of these patients puts another piece of the puzzle together. Often, in these cases, patients are discharged to hospices or other out-patient treatment facilities. It suggests that these patients are entering the hospital, and the hospital is either incapable, or unwilling, to treat them – shunting them off to other services they deem better equipped. But this isn't solving the problem – as these patients are returning back to the hospital in just a matter of months. This is creating expensive headaches as these returning patients take up valuable hospital space, consume resources that could be used for more critical patients, and, most tragically, these patients are not being helped by the medical professionals they rely on.

4.2. Naïve Bayes

The model predicted higher numbers for not readmitted with higher probabilities on number of diagnoses and number of medications. Those two variables are the significant variables to predict the result of patients. Table **??** shows the classification matirx of the model.

	TestSet_labels		
readmission_Prediction	<30	>30	NO
<30	58	91	39
>30	72	294	168
NO	443	1425	2498

Figure 13: Confusion Matrix of Naïve Bayes

4.3. Findings from Data Visualization

From the data visualization, it is evident that doctors pay more attention to elderly patients. In figure 3.1.1, irrespective of patient race, elderly patients were least re-admitted. Additionally, presence of diabetes was not the main cause for hospital re-admission.

In the figure 14, it can be observed that patient with more number of diagnosis where given less medications and vice-versa.

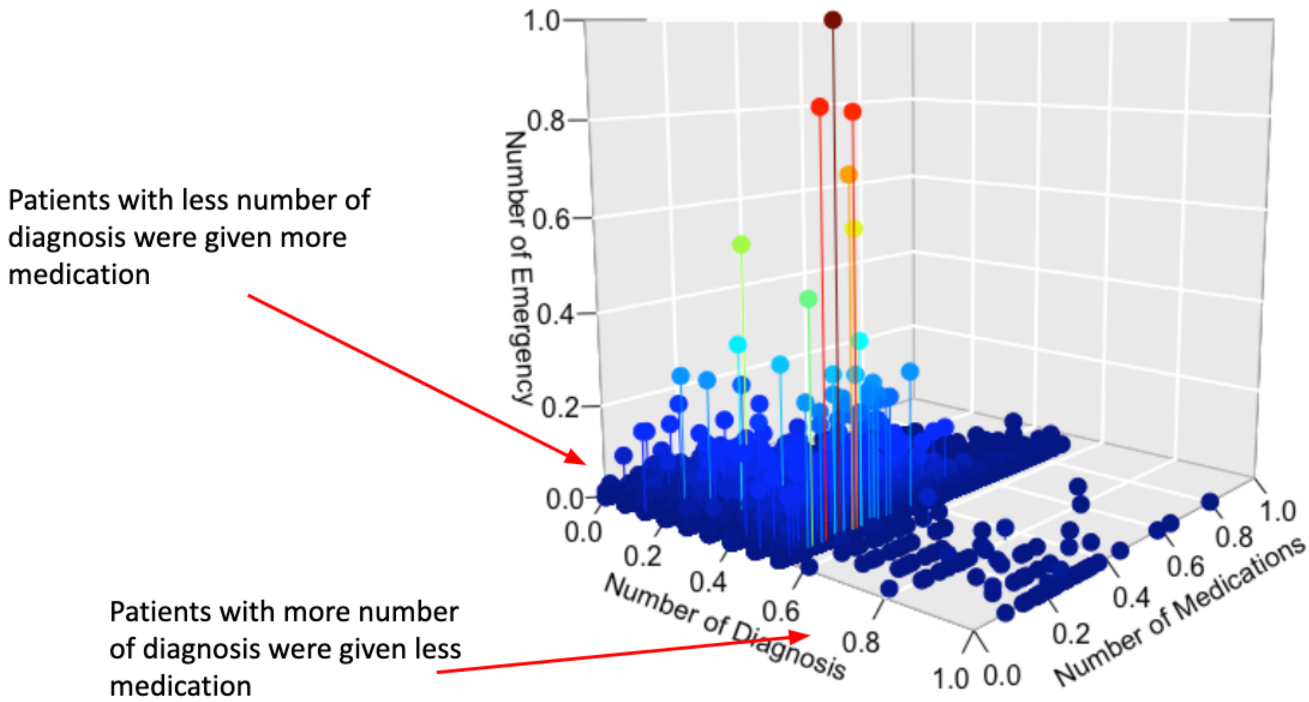

Figure 14: Poor diagnosis indication

5. Conclusion

Readmission rate is a quality evaluation metric customarily used to extrapolate the quality of life index of patient population and the quality of healthcare delivery (Shameer et al., 2017). Irrespective of the developments in biomedical and healthcare research practices, hospital quality control offices still use traditional predefined sets of variables to infer the probability patient readmission (Shameer et al., 2017). However, predictive analytics could provide evidences to improve the quality of healthcare delivery. Uniting predictive analytics with preventive measures would involve patients, physicians and payers to contribute proactively in taming the health and wellness.

This study implement a predictive analytical approach to identify patients prone to readmission and thus, systematically reduce the number of avoidable re-admissions mainly caused by patient non-compliance to medication instruction or early discharge from hospital. Our proposal has the capability of capturing both patient and population-based variations of hospital re-admissions. It incorporates patient with diverse health concerns across 130 US hospitals. The novelty of our method is to directly incorporate patients' history of re-admissions into modeling framework along with other demographic and clinical characteristics. We also verify the effectiveness of the proposed approach by validating training accuracy. Some contributions made in this paper are (i) applying Boruta algorithm and step-wise variable selection and (ii) implementing genetic and greedy ensemble algorithm to optimize our predictive models.

Some research directions can be sought by trying different variable selection techniques such as LASSO or Non-negative Garrote for better subset regressions. Also, in presence of high right censored data, it is interesting to consider some health care cost measures from which it may be possible to statistically estimate the mean population cost for readmission.

6. Reference

Ajorlou, S., I. Shams and K. Yang, 2014. Predicting patient risk of readmission with frailty models in the Department of Veteran Affairs. Proceedings of the IEEE International Conference on Automation Science and Engineering, Aug. 18-22, IEEE Xplore Press, Taipei, Taiwan, pp: 576-581. DOI: 10.1109/CoASE.2014.6899384

Allison, N., 2012. Examining the likelihood of readmission to inpatient pulmonary rehabilitation using a variety of predictors. ProQuest Dissertations and Thesis, ProQuest Dissertations Publishing.

Al-Wesabi, Y.M.S., A. Choudhury and D. Won, 2018. Classification of cervical cancer dataset.

Amarasingham, R., F. Velasco, B. Xie, C. Clark and Y. Ma et al., 2015. Electronic medical record-based multicondition models to predict the risk of 30-day readmission or death among adult medicine patients: Validation and comparison to existing models. BMC Med. Inform. Dec. Mak., 15: 39-39. DOI: 10.1186/s12911-015-0162-6

Anika, L. and M.L. Hines, 2014. Conditions with the largest number of adult hospital readmissions by payer. HCUP Statistical Brief 172. Rockville: Agency for Healthcare Research and Quality.

Appleby, L.K.N., 2013. The national confidential inquiry into suicide and homicide by people with mental illness. Healthwatch England.

Avishek, C. and C.M. Greene, 2018. Prognosticating autism spectrum disorder using artificial neural network: Levenberg-Marquardt algorithm. J. Bioinform. Syst. Biol., 1: 001-010. DOI: 10.26502/fjbsb001

Avishek, C.K., 2018. Decision support system for renal transplantation. Proceedings of the IISE Annual Conference, (AC' 18), IISE, Orlando.

Bakal, J.A., F.A. Mcalister, W. Liu and J.A. Ezekowitz, 2014. Heart failure readmission: Measuring the ever-shortening gap between repeat heart failure hospitalizations. PLoS ONE, 9: 106494-106494. DOI: 10.1371/journal.pone.0106494

Bayati, M., M. Braverman, M. Gillam, K.M. Mack and G. Ruiz et al., 2014. Data-driven decisions for reducing readmissions for heart failure: General methodology and case study. PloS one, 9: e109264- e109264. DOI: 10.1371/journal.pone.0109264

Betihavas, V., P.M. Davidson, P.J. Newton, S.A. Frost and P.S. Macdonald et al., 2012. What are the factors in risk prediction models for rehospitalisation for adults with chronic heart failure? Australian Critical Care, 25: 31-40. DOI: 10.1016/j.aucc.2011.07.004

Campolongo, F., S. Tarantola and A. Saltelli, 2000. Sensitivity anaysis as an ingredient of modeling. Stat. Sci. DOI: 10.1214/ss/1009213004

Choudhury, C., 2018. Identification of cancer: Mesothelioma's desease using logistic regression and association rule. Am. J. Eng. Applied Sci.

Claire Senot, A.C., 2015. What has the biggest impact on hospital readmission rates. Harvard Business Review.

Davison, B.A., M. Metra, S. Senger, C. Edwards and O. Milo et al., 2016. Patient journey after admission for acute heart failure: length of stay, 30-day readmission and 90-day mortality. Eur. J. Heart Failure. DOI: 10.1002/ejhf.540

Davood Golmohammadi, N.R., 2015. Prediction modeling and pattern recognition for patient readmission. Int. J. Product. Econom., 171: 151-161. DOI: 10.1016/j.ijpe.2015.09.027

Dixon Woods, M.A., 2005. Synthesizing qualitative and quantitative evidence: A review of possible methods. J. Health Services Res. Policy, 10: 45-53. DOI: 10.1177/135581960501000110

Futoma, J., J. Morris and J. Lucas, 2015. A comparison of models for predicting early hospital readmissions. J. Biomed. Inform., 56: 229-238. DOI: 10.1016/j.jbi.2015.05.016

Golas, S.B., T. Shibahara, S. Agboola, H. Otaki and J. Sato et al., 2018. A machine learning model to predict the risk of 30-day readmissions in patients with heart failure: A retrospective analysis of electronic medical records data. BMC Med. Inform. Dec. Mak. DOI: 10.1186/s12911-018-0620-z

Golmohammadi, D. and N. Radnia, 2016. Prediction modeling and pattern recognition for patient readmission. Int. J. Product. Econom. DOI: 10.1016/j.ijpe.2015.09.027

Greenhalgh, 2005. Effectiveness and efficiency of search methods in systematic reviews of complex evidence: Audit of primary sources. BMC, 331: 1064-1065. DOI: 10.1136/bmj.38636.593461.68

Hebert, C., C. Shivade, R. Foraker, J. Wasserman and C. Roth et al., 2014. Diagnosis-specific readmission risk prediction using electronic health data: A retrospective cohort study. BMC Med. Inform. Dec. Mak. DOI: 10.1186/1472-6947-14-65

Hodgson, R.E., M. Lewis and A.P. Boardman, 2001. Prediction of readmission to acute psychiatric units. Soc. Psychiatry Psychiatric Epidemiol., 36: 304- 309. DOI: 10.1007/s001270170049

IHSMHL, 2012. Improving hospital admission and discharge for people who are homeless. Inclusion Health St Mungo's Homeless Link, Healthwatch England

Inouye, S., V. Bouras, E. Shouldis, A. Johnstone and Z. Silverzweig et al., 2015. Predicting readmission of heart failure patients using automated follow-up calls. BMC Med. Inform. Dec. Mak., 15: 22-22. DOI: 10.1186/s12911-015-0144-8

Jiang, S., K.S. Chin, G. Qu and K.L. Tsui, 2018. An integrated machine learning framework for hospital readmission prediction. Knowledge-Based Syst., 146: 73-90. DOI: 10.1016/j.knosys.2018.01.027

Kang, C., K. Kim, J.H. Lee, Y.H. Jo and J.W. Park et al., 2015. Predictors of revisit and admission after discharge from an emergency department in acute pyelonephritis. Hong Kong J. Emergency Med., 22: 154-162. DOI: 10.1177/102490791502200304

Kathleen Carey, T.S., 2016. The Cost of hospital readmissions: Evidence from the VA. Health Care Manage. Sci., 19: 241-248. DOI: 10.1007/s10729- 014-9316-9

Kohavi, R. and G.H. John, 1997. Wrappers for feature subset selection. Artificial Intell., 97: 273-324. DOI: 10.1016/S0004-3702(97)00043-X

Kroeger, A.R., J. Morrison and A.H. Smith, 2018. Predicting unplanned readmissions to a pediatric cardiac intensive care unit using pre-discharge Pediatric Early Warning Scores. Congenital Heart Dis., 13: 98-104. DOI: 10.1111/chd.12525

Lewis, C.M., Z.L. Cox, P. Lai and D.J. Lenihan, 2016. Evaluation of two different models to predict heart failure readmissions. Heart Lung: J. Acute Critical Care, 45: 374-374. DOI: 10.1016/j.hrtlng.2016.05.008

Leys, C., 2013. Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the mean. J. Exp. Soc. Psychol., 49: 764-766. DOI: 10.1016/j.jesp.2013.03.013

Liaw, A. and M. Wiener, 2002. Classification and Regression by random Forest. R News.

Lum, H.D., S.A. Studenski, H.B. Degenholtz and S.E. Hardy, 2012. Early hospital readmission is a predictor of one-year mortality in communitydwelling older medicare beneficiaries. J. General Internal Med. DOI: 10.1007/s11606-012-2116-3

Maguire, D., 2015. The king's fund. Premature discharge: Is going home early really a Christmas gift? kingsfund.org.uk

Mahajan, S., P. Burman and M. Hogarth, 2016. Analyzing 30-day readmission rate for heart failure using different predictive models. Nurs. Inform., 2016: 225: 143-147. PMID: 27332179

Mahajan, S.M., P. Burman, A. Newton and P.A. Heidenreich, 2017. A validated risk model for 30- day readmission for heart failure. Stud. Health Technol. Inform., 245: 506-510. PMID: 29295146

Mengchao Wang, J.W., 2013. A comparison of approaches to stepwise regression for global sensitivity analysis used with evolution optimization. Proceedings of the 13th Conference of International Building Performance Simulation Association, (PSA' 13), France, pp: 2551-2558.

Miller, D.J., N.C. Beck and C. Fraps, 1984. Predicting rehospitalization at a community mental health center: A „double-crossed" validation. J. Clin. Psychol., 40: 35-39.

Nilsson, R., 2007. Consistent feature selection for pattern recognition in polynomial time. J. Mach. Learn. Res.

Ombudsman, T.P.O., 2003. The parliamentary ombudsman and the health service ombudsman website. Parliamentary Ombudsman and the Health Service Ombudsman.

Pack, Q.R., A. Priya, T. Lagu, P.S. Pekow and R. Engelman et al., 2016. Development and validation of a predictive model for short-and medium-term hospital readmission following heart valve surgery. J. Am. Heart Assoc., 5: e003544-e003544. DOI: 10.1161/JAHA.116.003544

Rowe, J.E., 2015. Genetic Algorithms. In: Springer Handbook of Computational Intelligence, Kacprzyk, J. and W. Pedrycz (Eds.), Springer, Berlin, pp: 825-844.

Shameer, K., K.W. Johnson, A. Yahi, R. Miotto and L. Li et al., 2017. Predictive modeling of hospital readmission rates using electronic medical recordwide machine learning: A case-study using mount sinai heart failure cohort. Biocomputing.

Shams, I., S. Ajorlou and K. Yang, 2015. A predictive analytics approach to reducing 30- day avoidable readmissions among patients with heart failure, acute myocardial infarction, pneumonia, or COPD. Health Care Manage. Sci., 18: 19-34. DOI: 10.1007/s10729-014-9278-y

Shinkman, R., 2014. Questex. Fierce Healthcare.

Shipeng Yua, F.F., 2015. Predicting Readmission Risk with Institution-Specific Prediction Models. Artificial Intell. Med., 65: 89-96. DOI: 10.1016/j.artmed.2015.08.005

Strack, B., J.P. Deshazo, C. Gennings, J.L. Olmo and S. Ventura et al., 2014. Impact of HbA1c measurement on hospital readmission rates: Analysis of 70,000 clinical database patient records. BioMed Res. Int. DOI: 10.1155/2014/781670

Tong, L., C. Erdmann, M. Daldalian, J. Li and T. Esposito, 2016. Comparison of predictive modeling approaches for 30-day all-cause non-elective readmission risk. BMC Med. Res. Methodol. DOI: 10.1186/s12874-

016-0128-0

Turgeman, L. and J.H. May, 2016. A mixed-ensemble model for hospital readmission. Artificial Intell. Med., 72: 72-82. DOI: 10.1016/j.artmed.2016.08.005

Wang, H., C. Johnson, R.D. Robinson, V.A. Nejtek and C.D. Schrader et al., 2016. Roles of disease severity and post-discharge outpatient visits as predictors of hospital readmissions. BMC Health Services Res., 16: 564-564. DOI: 10.1186/s12913-016-1814-7

Wang, H., R.D. Robinson, C. Johnson, N.R. Zenarosa and R.D. Jayswal et al., 2014. Using the LACE index to predict hospital readmissions in congestive heart failure patients. BMC Cardiovascular Disorders, 14: 97-97. DOI: 10.1186/1471-2261-14-97

Wasfy, J.H., K. Rosenfield, K. Zelevinsky, R. Sakhuja and A. Lovett et al., 2013. A prediction model to identify patients at high risk for 30-day readmission after percutaneous coronary intervention. Circulation: Cardiovascular Quality and Outcomes.

Yu, S., F. Farooq, A. van Esbroeck, G. Fung and V. Anand et al., 2015. Predicting readmission risk with institution-specific prediction models. Artificial Intell. Med. DOI: 10.1016/j.artmed.2015.08.005

Zheng, B., J. Zhang, S.W. Yoon, S.S. Lam and M. Khasawneh et al., 2015. Predictive modeling of hospital readmissions using metaheuristics and data mining. Expert Syst. Applic., 42: 7110-7120. DOI: 10.1016/j.eswa.2015.04.066

7. Appendix A

Submitted alongside as R file

THANK YOU FOR THIS AMAZING COURSE - WE ALL LEARNED A LOT FROM YOU