PREDICTION OF READMISSION RATE IN CARDIAC TREATMENT USING CHSI DATA

Table of Contents

Authors

Kisorkumar Varadarajan   Surya Bhaskar Ayyalasomayajula

Executive Summary

Reducing the hospital readmission rate is critical for any health care provider especially those in the cardiac health care segment. If a person is admitted and given treatment for cardiac – related issues, the visit following by the treatment should be a scheduled follow up versus an emergency case which is an example of readmission. Our aim is to identify the readmission rate of various treatments.

In this paper, we are going to examine various demographics among heart treatment methods (medication, stent or ballooning) when trying to find predictors for unscheduled readmission. We have taken into the input parameters such as marital status, race, age, BMI, gender, geographic location will be evaluated as important predictors of hospital readmission.

Normally in heart attack cases, patients are asked to come to the hospital for preventive check-ups and hospital staffs are instructed to mark these visits as "Recurring". For this study, a readmission is considered an unscheduled admission within 60 days of the first cardiac related issues. Our aim is to identify potential information that could be used by the health care providers during the initial and preventive checkup phase that could lead to a reduction in unscheduled readmissions.

We have obtained data from CERNER corporation (Cerner Corporation is a supplier of health information technology solutions, services, devices and hardware) through the Oklahoma State University's Center for Health Systems Innovation (CHSI). The data are from 2008 to 2013 for four regions in the U.S. (Midwest, Northeast, South and West). The data file contains 161,620 admission cases with 52 variables.

Problem Understanding

Research Question 1

Determination of the factors that affect the readmission rate after the cardiac treatment is given for a particular patient. Factors like age, height, weight, living style, day to day habits are considered in this situation. Regression is used to find the importance of variables which affect the rate of readmission.

Research Question 2

To determine the seasonality and trend of cardiac readmission. Time series analysis is used to find out this trend.

Description of Dataset

Data Owners

The data is originally owned by Cerner. Cerner Corporation is an American supplier of health information technology (HIT) solutions, services, devices and hardware. As of April 2015 its products were in use in approximately 18,000 facilities around the world and the company had about 22,000 employees globally. CERNER provides data to Center for Health Systems Innovation located at the Oklahoma State University where the data are analyzed by CHSI staff, faculty and students. It focuses specifically on rural health innovation, patient care innovation, and data analytics.

Authors

Kisorkumar Varadarajan   Surya Bhaskar Ayyalasomayajula

We have obtained 5 datasets from the CHSI as follows

Patient encounter info table

This dataset contains the basic information about the patients who are admitted to the hospital.

Diagnosis Info table

This dataset contains diagnostic information for each of the encounter ID is the foreign key and it helps in linking the encounter table and the diagnosis table

Procedure Info & Medication Table

Once the patients are diagnosed with cardiac event, they are sent for medication or to surgical procedures. This information is captured in these tables. The patients who undergoes surgery are also given medication. This information will be found in procedure or medication tables.

Data Exploration

We have examined a total patient encounters of 161,620 and out of which there were 141,458 unique patients. These patients belong to both Urban & Rural areas and they are in the Midwest, Northeast, South & West regions of the United States. The followings races were used for this study.

- African American
- Asian
- Biracial
- Caucasian
- Hispanic
- Native American
- Not Mapped
- NULL
- Other
- Pacific Islander

We have put the value as 1 for readmission variable if the readmission is before 60 days and considered as 0 if the readmission is after 60 days. The following surgical procedures are considered by us

Procedure code	Procedure
12579	Transluminal balloon angioplasty, venous (e.g., subclavian stenosis), radiological supervision and interpretation
8882	Transluminal balloon angioplasty, percutaneous; venous
9172	Trans catheter placement of intravascular stent(s), cervical carotid artery, percutaneous; with distal embolic protection
Medical Treatment	NULL

Below is the distribution of patients who were readmitted based on the above criteria.
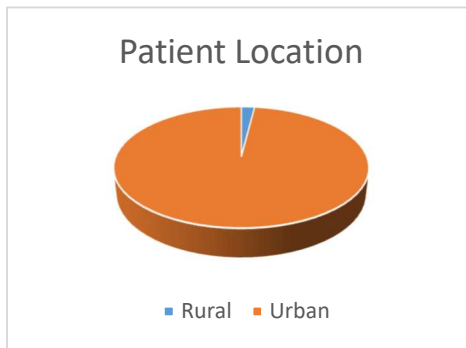
Authors

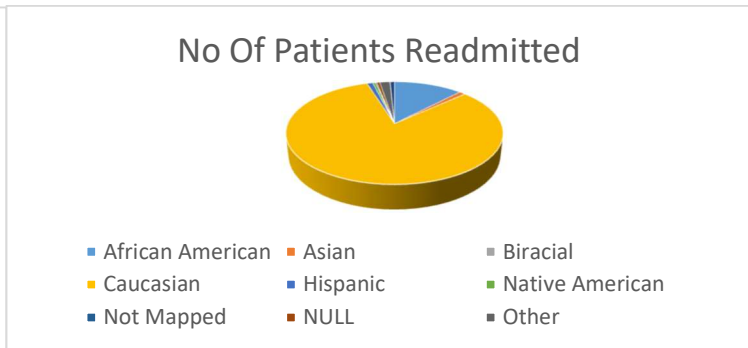Kisorkumar Varadarajan   Surya Bhaskar Ayyalasomayajula

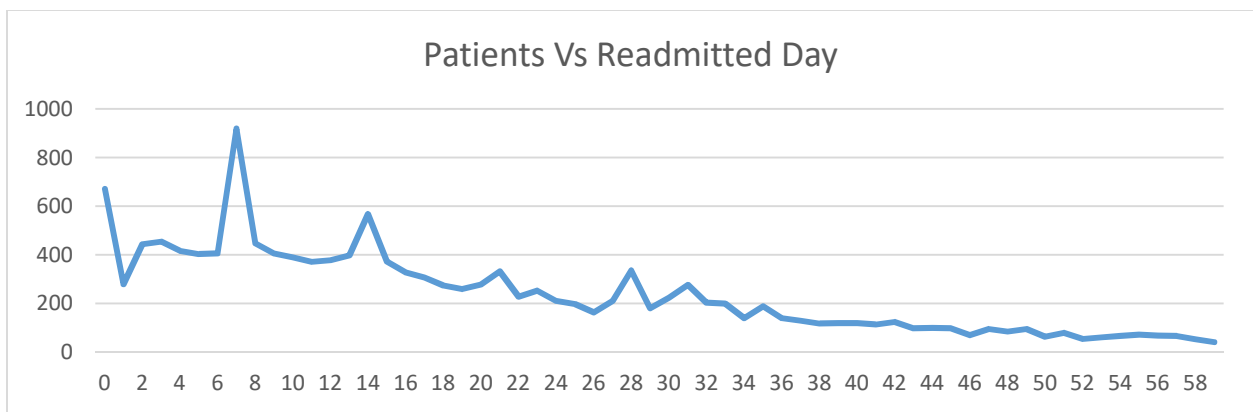# Prediction of Readmission rate in cardiac treatments using CHSI Data

Based on the Patient location   Based on Race

Patient Location

- Rural - Urban

No Of Patients Readmitted

- African American - Asian - Biracial
- Caucasian - Hispanic - Native American
- Not Mapped - NULL - Other

Number of days to Readmission

Patients Vs Readmitted Day

Medical Treatment Provided

No of Patients
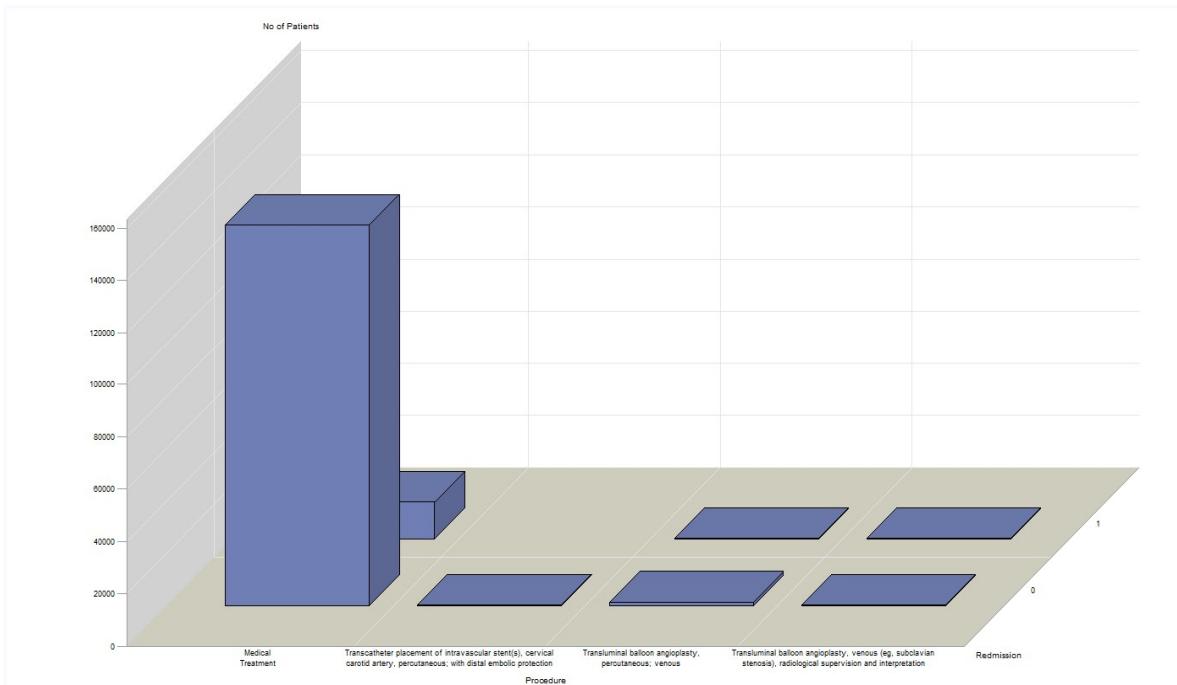
Authors
Kisorkumar Varadarajan   Surya Bhaskar Ayyalasomayajula

Data Preparation

## Join on Diagnosis info table

We have taken the table **query_get_encounter_info** as the base table to get the details of the patients who are admitted in the hospital for various diseases. This table is then joined with the table **query_get_diagnosis** through the encounter ID which is unique for every encounter in order to identify the patients who are being diagnosed with what type of diseases. Following that, the patients with the heart disease were filtered out from the dataset that was obtained from the above two datasets.

Join on Procedure info table

Using the above derived dataset, it is joined with the table **query_get_procedures** using the same encounter ID in order to find out the different types of heart disease cases. The patients come under 4 category Transluminal balloon angioplasty, venous (eg, subclavian stenosis), radiological supervision and interpretation, Transluminal balloon angioplasty, percutaneous; venous,Trans catheter placement of intravascular stent(s), cervical carotid artery, percutaneous; with distal embolic protection, Medical Treatment. For this analysis, we are considering mainly the people who are readmitted after the surgery.

Creation of readmission days and Readmission indicator variable

We created a table by self-joining the **query_get_encounter_info** table to itself on variable patient_id which has multiple records and are sorted in ascending order in accordance with the readmission date. Then the dates are subtracted from each other to calculate the number of days for readmission. In this new table, the readmission indicator is created where the readmission is set as 1 if the readmission day is between 1 to 60. Else it is marked as 0. We have used the logic that if readmission is more than 60 days it is a scheduled checkup and if the readmission day is 0, the patient did not have any problems to come back to work. Below is the demographic distribution of readmissions.

Authors
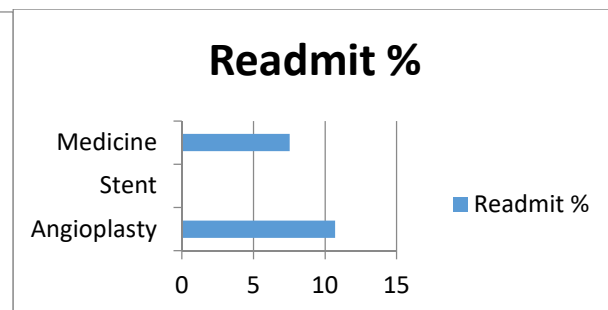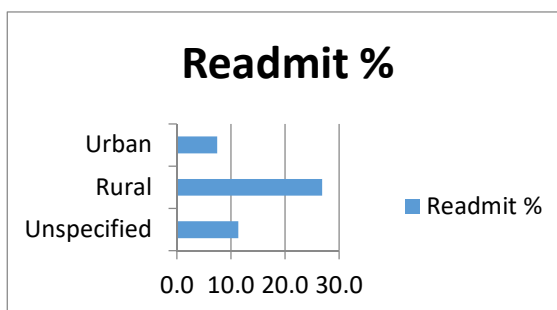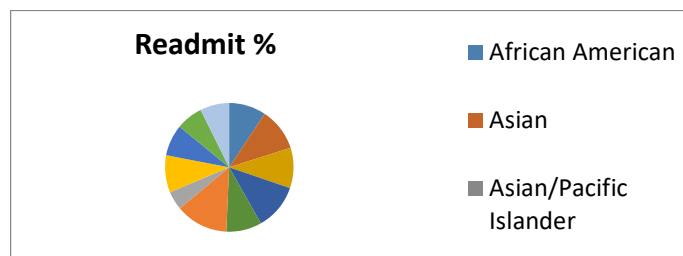
Kisorkumar Varadarajan   Surya Bhaskar Ayyalasomayajula

Preparing Time series data

Using time series, two types of analysis is performed as follows,

1. To find the seasonality and trend of the readmission rates of different months. For this, the data prepared from the above step was used. The month, year and the readmission numbers are pulled from this data and graph was plotted. For e.g.

Month	Year	Number of Patients readmitted
June	2010	234
July	2010	345
Aug	2010	122
.	.	.
.	.	.
.	.	.

2. To find the seasonality and trend of the readmission rates for different days. As above the same was followed but in this case the data is pulled with the number of readmission cases and the day in which the patients were readmitted. For e.g.

| Readmission day	Number of Patients
1	24
2	36
3	33
.	.
.	.
.	.

Modeling

Time series Analysis

Predicting the readmission rate / period for cardiac patients who have undergone treatment (Medicine, Ballooning or Stent) by admission to hospital is a complex question. There are many factors that come into play. Some of them are age, race, ethnicity, income, number of kids, marital status. Data on treatment by medicine and ballooning (angioplasty) is plenty. As there are only 21 stent patients much cannot be inferred about the same.

The number of readmissions per month can be analyzed as a time series. By analyzing we can see if there is a trend or seasonality. The seasonality, if any, would provide us the information to take more precautions during the season and would also provide us a stimulus to analyze further why the seasonality is occurring and how we can eradicate or minimize it. We have accumulated the Data on a monthly basis. Hence, the frequency value in the ARIMA model = 12.

Authors
Kisorkumar Varadarajan   Surya Bhaskar Ayyalasomayajula
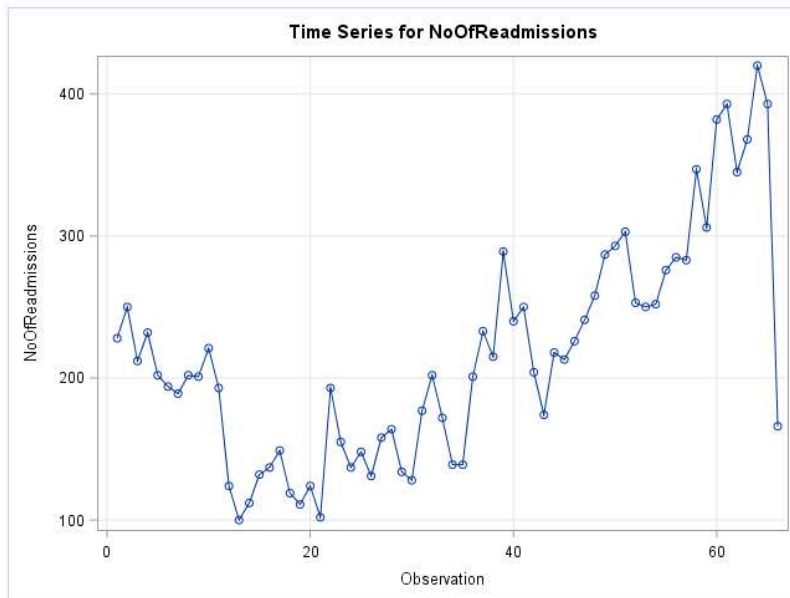
Prediction of Readmission rate in cardiac treatments using CHSI Data

The below is the overall model built for the time series analysis
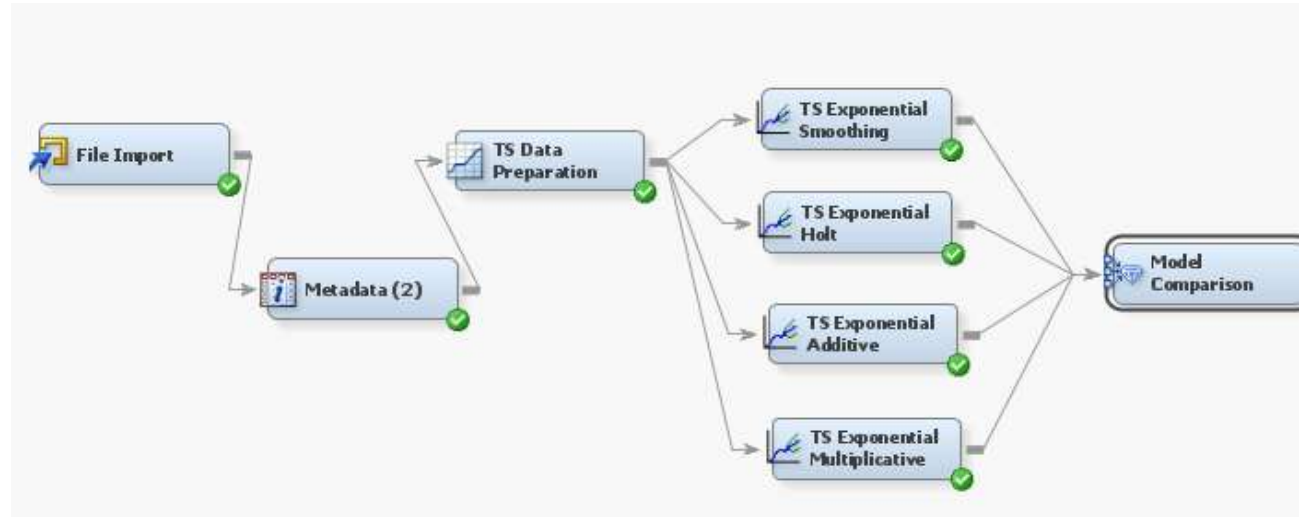
The following is the graph of Readmissions over the period Jan-2008 to Jun-2013.

To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	189.42	6	<.0001	0.845	0.737	0.674	0.618	0.564	0.498
12	261.46	12	<.0001	0.476	0.422	0.410	0.369	0.330	0.310

Autocorrelation Check for White Noise

While test the null hypothesis that "The time series observations are uncorrelated (white noise)", it has been rejected (probability values highly significant), hence it can be concluded that the

Authors
Kisorkumar Varadarajan   Surya Bhaskar Ayyalasomayajula

given time series is not simply white noise and there is meaning to exploit the dependency between observations and TS models can be tried.
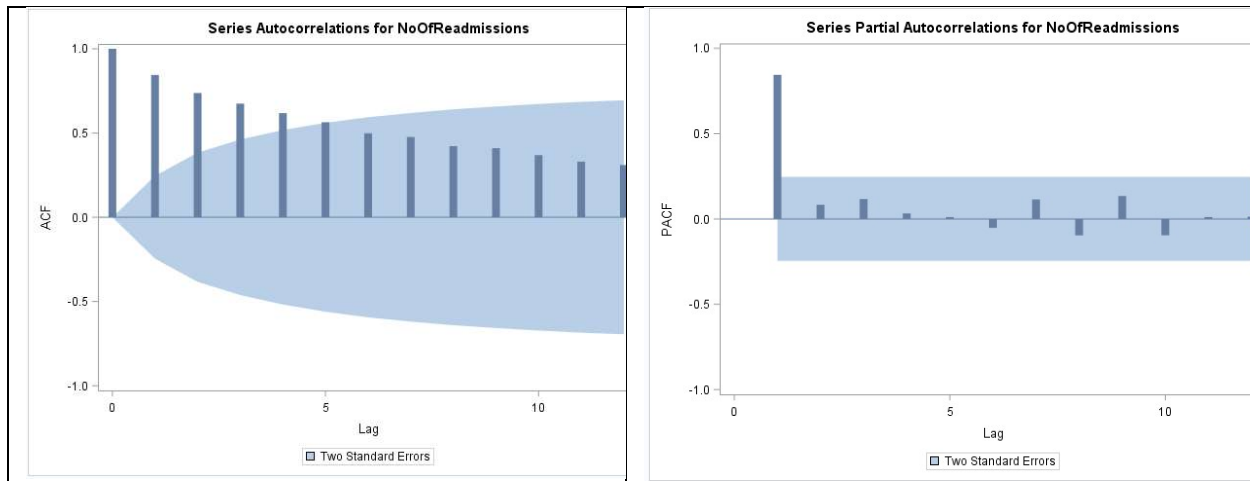
A TS is said to be stationery if its underlying generating process is based on a constant mean and constant variance with its autocorrelation function (ACF) essentially constant through time.

Autocorrelation refers to the way observations in a TS are related to each other and is measured by the simple correlation between current observation and observations from p periods before the event.

Partial autocorrelations are used to measure the degree of association between yt and yt-p when the y-effects at other time lags 1,2,3,,,,p-1 are removed.

The PACF have spikes at lag 1 and cuts off to zero. This tells that the process is autoregressive with order 1.

The Auto Correlation is also not falling down quickly to zero implying that there is a trend.
The ACF and PACF graphs show that there is seasonality and trend.
The model selected by the Varmax is the above one.

Model Parameter Estimates						
Equation	Parameter	Estimate	Standard Error	t Value	Pr > |t|	Variable
NoOfReadmissions	CONST1	6.58735	0.61272	10.75	0.0001	1
	AR1_1_1	0.90450	0.00000			NoOfReadmissions(t-1)
	MA1_1_1	0.43246	0.15531	2.78	0.0075	e1(t-1)
	MA2_1_1	0.57045	0.15245	3.74	0.0005	e1(t-2)

The MA2_1_1 and MA1_1_1 both are significant at the 5% significance level.

To estimate the trend component of a non–seasonal time series that can be described using an additive model, it is common to use a smoothing method, such as calculating the simple moving average of the time series.
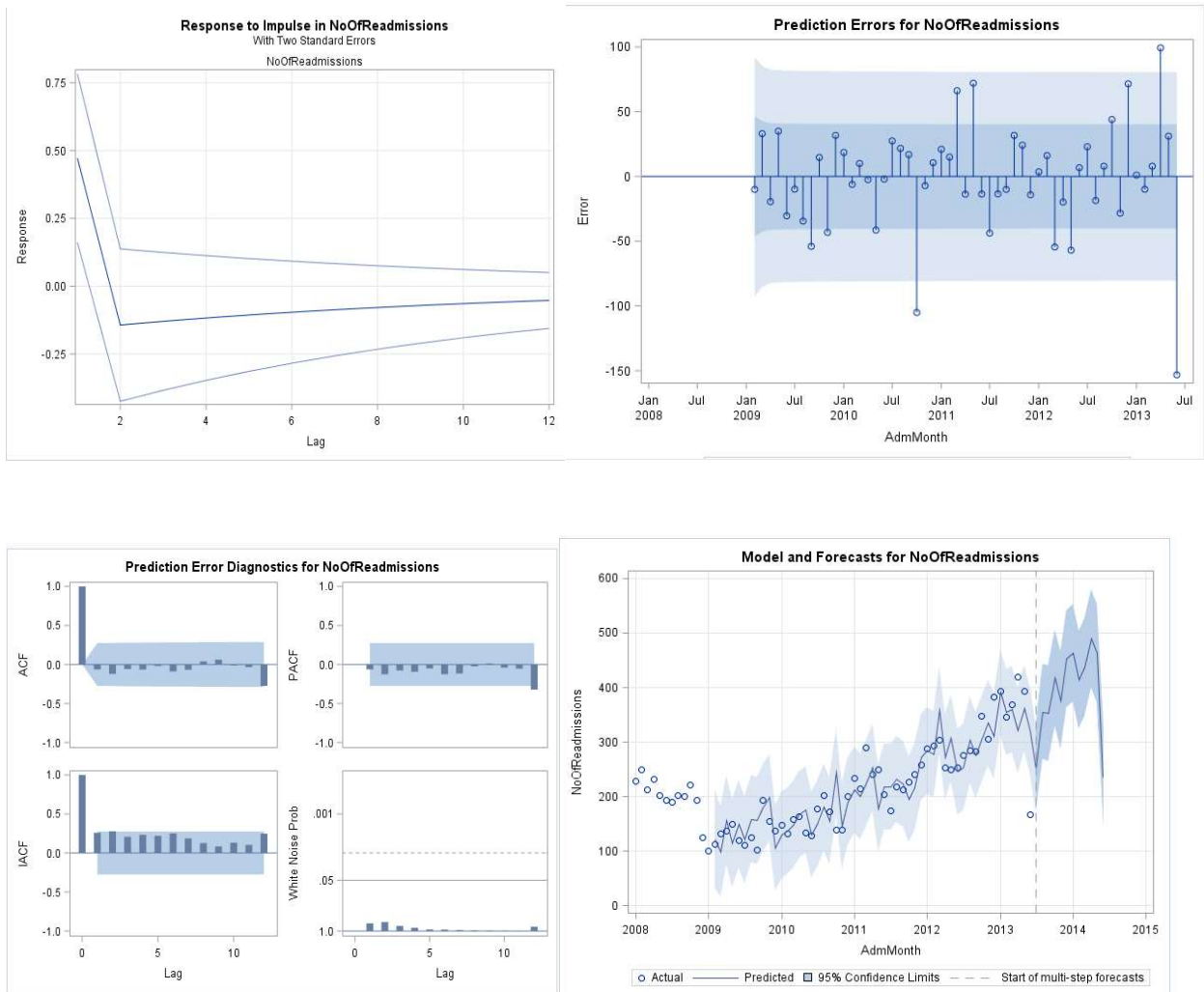
Authors
Kisorkumar Varadarajan   Surya Bhaskar Ayyalasomayajula

The selected model has 4 variables to predict a future value.

1. Intercept of 6.58735.
2. Autoregressive component (t-1) with a weight of 0.90450
3. Moving average component with weight 0.43246
4. Moving average component with weight 0.57045

We see a trend that it is growing which is proportional the increased number of cases.
There seems to be seasonality also. Most of the peaks are in and around April. Care should be taken for Patients being treated in February and March.

Authors
Kisorkumar Varadarajan   Surya Bhaskar Ayyalasomayajula

The Data of the last month is either incomplete or there is a real drop in number of cases.
The AIC and SBC are very small for the model selected by VARMAX procedure in comparison to
the (AR 1,1) model.

Information Criteria	
AICC	459.4256
HQC	461.8616
AIC	458.1213
SBC	467.8775
FPEC	1859.207

Constant Estimate	1.350934
Variance Estimate	2979.63
Std Error Estimate	54.58599
AIC	576.3452
SBC	580.2857
Number of Residuals	53

* AIC and SBC do not include log determinan

Correlations of Parameter Estimates		
Parameter	MU	AR1,1
MU	1.000	-0.090
AR1,1	-0.090	1.000
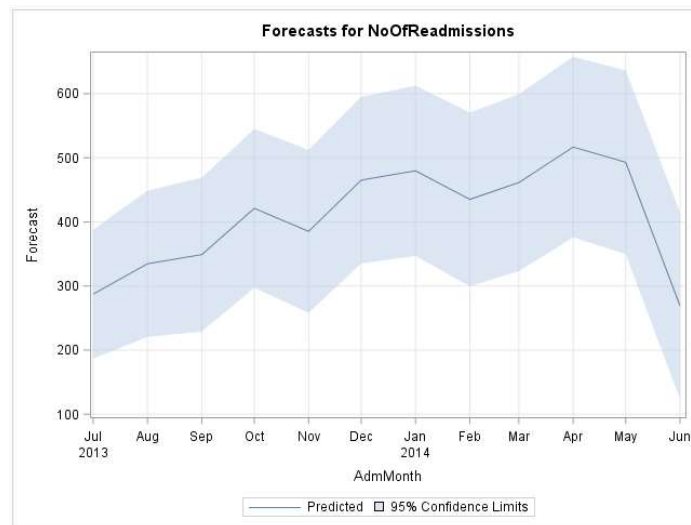
The forecast for the next year also show that there is a peak in the month of April and the dip is
in Jun. It needs further statistical and clinical investigation for the reasons of the April peak and
June dip.

Regression Modeling

The output of the SQL join is saved to CSV file for further analysis in Enterprise Miner.
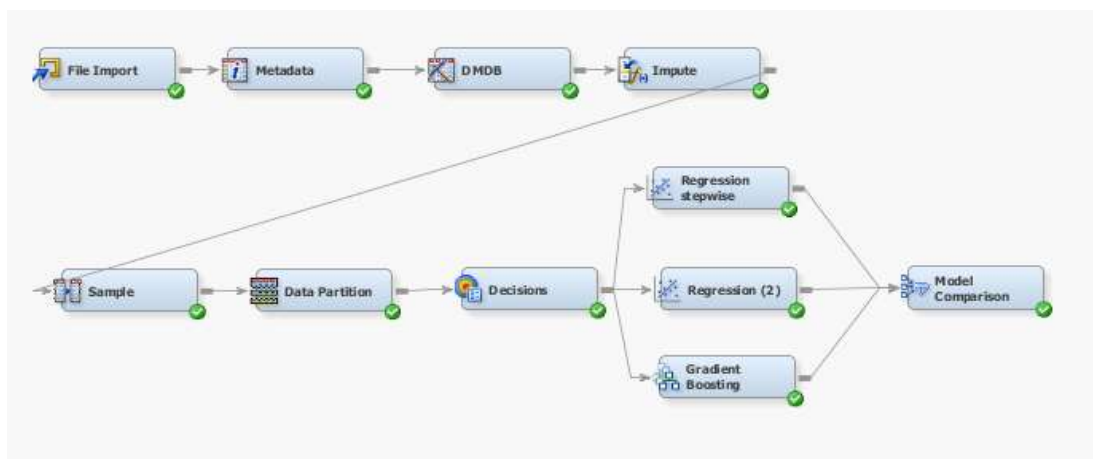
Authors
Kisorkumar Varadarajan   Surya Bhaskar Ayyalasomayajula

Imported the CSV data using File Import Node. Using metadata set correct roles of the variables. Imputed the missing values using the Impute Node. As the number of events to be analyzed in the population is very less we have oversampled the data. Data is partitioned into Test and Validation. The oversampling has been taken care by use of the Decisions node.

We have run three different models. Stepwise Regression, Regression and Gradient Boosting. Stepwise Regression was selected as the best model based on Average Profit for Target. It is the best model based on other statistics also namely: Averaged Square Error, Mean Square Error, and Misclassification Rate.

The 13th iteration of the Stepwise Regression was selected. As per the regression:

The predictor model consisted of the following variables:

Intercept, BED_SIZE_RANGE, IMP_ACUTE_STATUS, IMP_CENSUS_REGION, IMP_DRG_TYPE, IMP_MEDICAL_SPECIALTY, IMP_URBAN_RURAL_STATUS, PAYER_CODE, TEACHING_FACILITY_IND TOTAL_CHARGES_NUM, UNIT_DESC, marital status, race and weight

The following category patients have higher rate of readmission. This was obtained as result of regression.

- Bed Size Range < 5
- Acute Status - Acute
- Region - Northeastern
- Drug Type = DRG
- Who are transferred from the specialties like Radiology, Nephrology, and Psychology.
- Race - Asian
- Area - Rural
- Payer Code= "MR"
- Hospitals w/o teaching facility
- Marital Status – Divorced, Life Partner

Conclusions

Patients discharged during the months of February to April have a higher tendency of readmission. Preventive care must be taken before discharging during this period. Similarly, Asian patients must be taken care of as the relapse rate is higher. Patients of Rural regions due to non-availability of Medicare nearby are prone to readmission. As divorced patients do not have a partner to support them in the process of recovery are prone to relapse. It would help if they are provided with medical /social support after the discharge.

Recommendations

Further analysis should be done to reason out why the readmission rate is higher around the month of April.

Authors
Kisorkumar Varadarajan   Surya Bhaskar Ayyalasomayajula