

What Makes Hospital Readmissions Unpredictable?

Understanding complex healthcare data can be a daunting task. Without the right tools or methods, what could be an interesting fact-finding journey leading to business improvements could essentially end up becoming a bunch of excel sheets stashed in a folder for later use.

Explore the insights Analance found on a project for a large hospital where the goal was to predict which patients had a higher likelihood of returning with an illness post discharge

HEALTHCARE ANALYTICS

USE CASE



BUSINESS CHALLENGE

The hospital was experiencing large losses in terms of dollars and practitioner's and medical staff time due to unpredictable patient readmissions within 30 days post discharge.



SOLUTIONING PROCESS AT A GLANCE

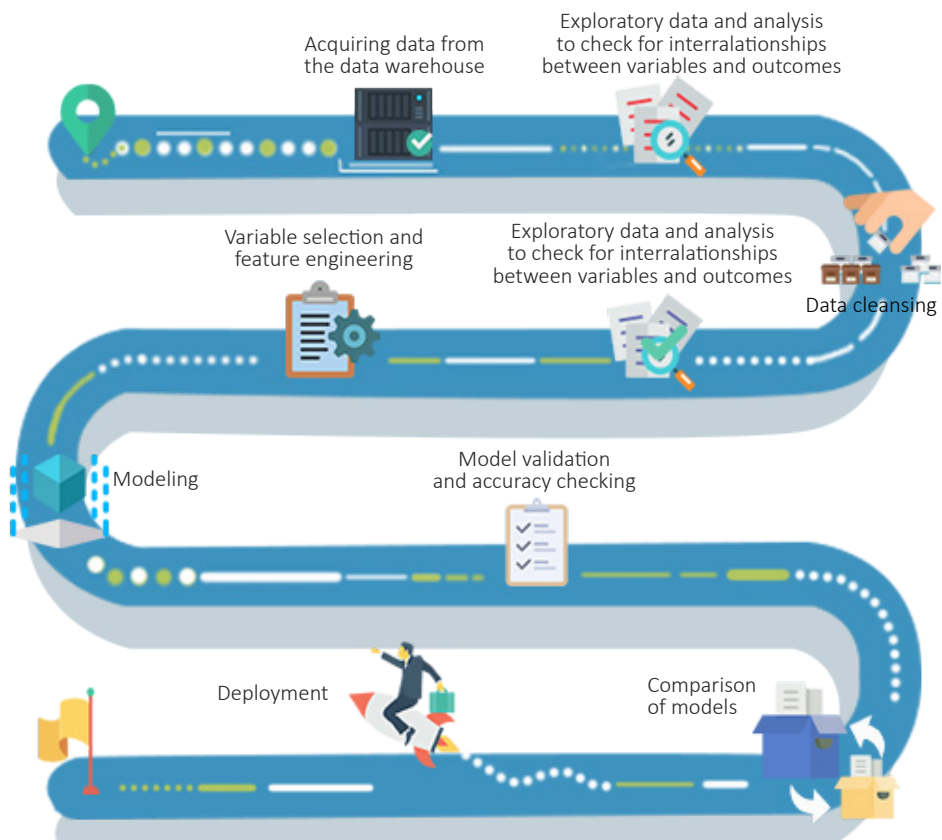
The process of statistical consulting and solutioning starts with a thorough understanding of the business challenge, its impact, and the data available for analysis. With this information, we arrive at a solution to mitigate or control the challenge, offer continued client support, and adjust models over time.



OUR PROCESS

A data dump was acquired and put through a stringent exploratory process before trying to correlate what information was available to solve the challenge at hand.

PREDICTIVE MODELING STAGES



The following variables were available for analysis:

- Ethnicity of patient
- Gender of patient
- Age of patient categorized into tens

Numeric Values

- Time in hospital in days
- Number of lab procedures
- Number of other procedures
- Number of medication administered in the facility
- Number of outpatient visits so far
- Number of inpatient visits so far
- Number of emergency visits so far
- Number of diagnosis so far

Yes / No variables with, Yes = Administered at facility, No = Not administered at facility

- | | | | | |
|---------------------------|-----------------------|-----------------|------------------|-----------------------|
| • DiabetesMed | • Gluburide.Metformin | • Troglitazone | • Tolbutamide | • Nateglinide |
| • Metformin.Pioglitazone | • Insulin | • Miglitol | • Glyburide | • Metformin |
| • Metformin.Rosiglitazone | • Citoglipton | • Acarbose | • Acetohexamide | • HBA1c result |
| • Glimepride.Pioglitazone | • Examide | • Rosiglitazone | • Glimepiride | • Glucose serum level |
| • Glipizide.Metformin | • Tolazamide | • Pioglitazone | • Chlorpropamide | • DiabetesMed |

Readmission outcome (within 30 days)

- Readmission status (Yes within 30 days / No)

ICD 9 coded medical diagnosis

- Diagnosis ID 1
- Diagnosis ID 2
- Diagnosis ID 3



THE MODELING PROCESS

A Mathematical/Statistical/Econometric predictive model is a method of representing a variable of interest (outcome) as a function of other variables (predictors) with an assumed relationship between the outcome and predictors.

All variables available were studied to understand distributions. Data was cleaned by the means of handling outlying values, missing values, and looking for inter-relationships between predictors before looking to see if any data had a significant relationship with the outcome.

Outlying values were removed from the data after which missing values were replaced with column means (average value) for numeric data and column modes (most frequent label) for categorical or ordered data. The industry standard Box Plot was used to check for outliers and a tabulation or summary of data was used to check for missing values. If two or more predictors were inter-related (for instance height in cm, height in inches, height in feet), only one was used.

The industry standard metric VIF (Variance Inflation Factor) was used as a deciding factor. Predictors with VIF values above 6, indicating a significant inter-relationship with another predictor/s, were removed from the analysis.

The industry standard statistical test used to check for relationships between predictors and the outcome of interest is the Chi-Squared test of association. A p-value less than .05 indicates that the probability of association occurring in the population is less than 5% and is of no significance. Using this test for all predictor-outcome combinations helped in restricting the analysis for only those predictors that majorly influenced readmissions.

A further exhaustive exploratory analysis of data using graphs and advanced visualizations found that:

- The facility offered a very high quality of elder care
- Administering medication for diabetes in the facility significantly reduced the risk of readmission. Offering a high number of diagnosis (4 to 9) or prescribing a high number of lab procedures (5 to 11), reduced the risk of readmission.
- Performance of a high number of procedures (2 to 5) in the facility reduced the risk of readmission.
- Drugs offered at the facility could be classified into readmission safe/unsafe drugs.
- Certain diagnoses were found to increase the risk of readmission.





MODEL USED

The modeling process began with using the industry preferred choice (namely, Binomial Logistic Regression), designed specifically for modeling binary outcomes (i.e. Yes / No). Other models, such as Decision Trees, Neural Networks, and Support Vector Machines, were also considered and results from them were analyzed to find the top performing model.

A total of 6 different models were built to find the winning model in terms of findings from the confusion matrix. A confusion matrix tabulates the predicted values from the model with the actual values in the data to see how close we are at capturing the true relationship between the outcome and predictors by means of using derived metrics such as the following:

- **Accuracy:** The model's ability to predict correct values and incorrect values correctly.
- **Kappa value:** A comparison of observed accuracy with expected accuracy.
- **Sensitivity:** The sensitivity of a test (also called the true positive rate) is defined as the proportion of readmissions that has been correctly classified as readmissions.
- **Specificity:** The specificity of a test (also called true negative rate) is defined as the proportion of non- readmissions that have been correctly classified as non- readmissions.
- **Positive predictive value:** The probability that those values predicted as readmissions are truly readmissions.
- **Negative predictive value:** The probability that those values predicted as non-readmissions are truly non- readmissions



CONCLUSION AND RECOMMENDATIONS

Knowing which patients have a higher propensity for readmission within a set period of time or are at risk of falling ill again within set periods after discharge are critical insights to hospitals and, in general, the healthcare industry. Once a model is deployed in a facility, each patient leaving the facility after care can be classified into risk categories based on the likelihood of their return. This can be determined by an algorithmic deduction of their propensity to return based on the predictive factors and factor levels present in a particular facility and/or with a particular provider. Further, by using Analance for modeling the data and creating relevant reports, alerts can be scheduled which alert the check-out staff about the risk of return for the patient being discharged from the facility.

Based on these insights, the organization's administration can plan their service offerings and create evidence based policies which may result in increased ROI. Further, with the help of SMEs, the hospitals or healthcare facilities can develop and offer services such as counselling patients about quality of life maintenance post discharge, home care, and prescription filling. For example, from the analysis done by experts at Ducen on the available data, we found that administering medication in the facility decreases the likelihood of patient readmission. With this kind of insight, an organization can implement a policy that instructs check out staff to administer relevant medication in the facility and counsel patients on its benefits. From the analysis that was done on the data found on the internet, quality of life counseling has been consistently showing up as an unmeasured covariate that decreases the likelihood of readmission.

ABOUT DUCEN

Ducen IT helps Business and IT users of Fortune 1000 companies with advanced analytics, business intelligence and data management through its unique end-to-end data science platform called Analance. Analance is an enterprise-class, state of the art integrated platform that delivers power and ease of use to business users and data scientists with a seamless experience and platform scalability to support business growth and strategy.