Paper 2819-2018

# Examining the drivers of hospital readmissions of Type-2 Diabetic patients

Shashank Reddy Gudipati, Oklahoma State University


## ABSTRACT

In an uncertain business environment, there is an increasing demand by healthcare companies for data and resources about patient's health. Analyzing this data and building models will help healthcare service providers to improve their efficiency and effectiveness. Healthcare industry is moving from only reporting facts out of this data to discovery of insights and predicting prospective trends by studying past patterns. Diabetes Mellitus is one of the major health hazards in developing countries. A large proportion of patients, diagnosed with type-2 diabetes, readmit in hospitals after their first admission. These readmissions are expensive and are caused generally by either patient's bad health or because of bad healthcare provider. Predicting these potential readmissions can help in improving patient care, quality of care by health care providers. In this paper we will look at the drivers that cause potential re-admissions of patients and a way to predict these readmissions using the variables that are easily available for a healthcare provider/ hospital. A dataset containing approximately 165000 records containing quantitative and qualitative information related to many patients residing in the state of Oklahoma provided by the Center for Health Systems Innovation at Oklahoma State University was used for this analysis. Various classification models were built to predict potential readmission using variables like Diagnosis information, patient demographics, hospital demographics etc. Decision Tree yielded the best result predicting the output with misclassification rate of 0.134. Variables category-that shows the prevention group of Type-2 diabetes patient, diagnosis code -that shows diagnosis information of patient, hazard rate are the most important variables for predicting the readmission using the built decision tree model.


## INTRODUCTION

According to Centre of Disease Control and Prevention, over 9.3% of the US population has diabetes. Currently, one out of three people will develop diabetes in their lifetime. Diabetes is mainly of two types, Type 1 - when the body does not make enough insulin- and Type 2 - when the body cannot use insulin properly. As Type 1 is genetic and involves hereditary transmission, we will concentrate on Type 2 diabetes in our analysis. Treatment regimens in Type 2 diabetes are complicated surrounding lifestyle adaptions and social behavior.

According to Medicare Payment Advisory Commission (MedPAC), around one-fifth of the Medicare beneficiaries discharged from a hospital get re-admitted within 30 days. Some of the readmissions are related to initial reason the patient came to visit the hospital and some are not related. Predicting these potential readmissions can help in improving patient care, quality of care by health care providers.

Objective of this paper is to assess the drivers the cause these potential re-admissions and to predict potential re-admission based on factors that are easily available to the healthcare providers. A dataset of around 160,000 observations and 70 variables, containing quantitative and qualitative information of patients provided by the Center for Health Systems Innovation at Oklahoma State University was used for this analysis. A limited number of features which may directly affect the patient's health are used for building the model. Even though the lab test results such as glucose levels and A1C levels are available they were only used to classify patients into diagnosis categories and are not used in the model. These values were not included as patients might have different lab results at various levels of diagnosis and patient is generally bound to admit in a hospital when these lab results are high or low. Only general hospital and patient level attributes which are easily available to any health care provider such as patient demographics and hospital demographics were used. Factors such as gender, race, diagnosis information, cumulative hazard rate and other patient, hospital level factors are used for predicting readmissions.

## DATA DICTIONARY

| Variable Name | Variable type | Description |
|---|---|---|
| Patient_ID | Nominal | Unique Identifier for patient |
| Encounter_ID | Nominal | Unique Identifier for encounter |
| Age_in_years | Interval | Age in years |
| Race | Nominal | Race of patient |
| Gender | Nominal | Gender of patient |
| Marital_Status | Nominal | Marital status of patient |
| Category | Nominal | Diagnosis category of patient |
| Diagnosis_Code | Nominal | Diagnosis code during admission |
| Payer_Code_Desc | Nominal | Payment method (Self or through insurance) |
| Average_lengthstay | Interval | Average length of stay of patients in the hospital patient is admitted |
| African_American_prop | Interval | Average number of African American patients in the hospital patient is admitted |
| Hazard_fn | Interval | Cumulative hazard rate of hospital utilization of hospital that patient is admitted |
| Length_of_stay | Interval | Length of stay of patient |
| Total_admissions | Interval | Total admissions in the hospital patient is admitted |
| Readmission | Nominal (Target) | Readmission indicator Yes/ No |

**Table 1: Data Dictionary**

## DATA PREPARATION

Center for Health Systems Innovation at Oklahoma State University holds a large repository of Cerner Health facts data of Type-2 diabetic patients. Data that is provided was divided into three segments; type-2 diabetes data, glucose test data and A1C data. Type-2 diabetes data had patient details with their diagnosis, encounters and visits. A1C data had A1C level of patients. Glucose test data had the glucose test details of the patients.

First step for this project is to classify the type-2 diabetic patients into primary, secondary and tertiary prevention groups. Primary group has patients who are pre-diabetic, secondary group has patients who are diabetes diagnosed and does not have any complications and tertiary group has patients who are diabetic and have complications. A1C and Glucose tests are considered for classifying the patients into these prevention groups. The A1C and glucose datasets contain various lab procedures from which the diabetes concentrated lab records are considered. The glucose procedures that are taken into consideration are OGTT procedures and the glucose fasting procedures at different times are taken into consideration. Apart from these tests, the glucose serum tests are also taken into account. Pre-diabetic patients (Primary prevention group) are not considered for this readmission analysis.

Data preparationData preparation

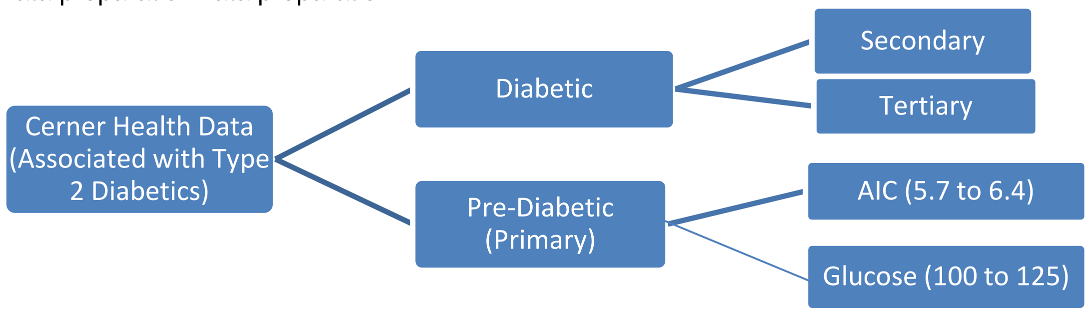


**Figure 1: Prevention group creation flow**

Secondary task for data preparation involves the merging the tables horizontally that contains the unique columns such as patient ID and the encounter ID. They are sorted by using these columns. Also, all these combined datasets and some of the important variables are considered from all these internal datasets.

After the data merging is done a flag for re-admission is created checking if the same patient re-admits into the hospital for the same reason in a span of 30 days. We only considered the readmissions that are caused for the same reason as first admission. Reason for admission is checked based on the ICD9 codes i.e. encounters with same ICD9 information count as readmission for same reason. Admission into the hospital is classified from general depending on whether the encounter is an Inpatient encounter or Out-patient encounter. Emergency or unplanned encounters are ignored. Data independency is maintained by using only the data related to first two readmissions to create the target variable of flag showing if there is a re-admission and then retaining only one observation for each patient.

Cumulative hazard rate variable is created using the existing variables. The cumulative hazard rate is the cumulative number of hospital utilization events over total discharges over 30 days. This variable illustrates how the risk of hospital utilization changes over time for each group. In the case of readmission analysis this variable is calculated as number of readmissions in the hospital over total number of discharges over a period of 30 days. In general, this variable gives the proportion of readmissions in a particular hospital.

Data was mainly present in the form of patient level indicators and hospital level indicators. Some new variables were derived from other variables available in the data. As the main purpose of this paper is to build a predictive readmission model that can be helpful for healthcare providers/ hospitals, only the variables that are directly available to the healthcare provider are used for modelling and other variables are ignored. Another reason for considering only these variables is to eliminate the correlations between variables in the model.

## METHODOLOGY

The modelling approach followed for this project is SEMMA (Sample, Explore, Modify, Model and Assess). Data was partitioned into two stratified samples – training (70%) and validation (30%). Dividing data into training and validation helped in reducing overfitting of the model. A nominal target variable indicating readmission as “Yes” or “No” is used. The training data is used to build various models and validation data is scored based on the models built to assess the performance of the model. Misclassification rate is used as the model assessment metric. This provides honest assessment of the models built.

Various models were built using SAS® Enterprise Miner™14. The top-3 models were shown in the below Enterprise Miner node diagram.

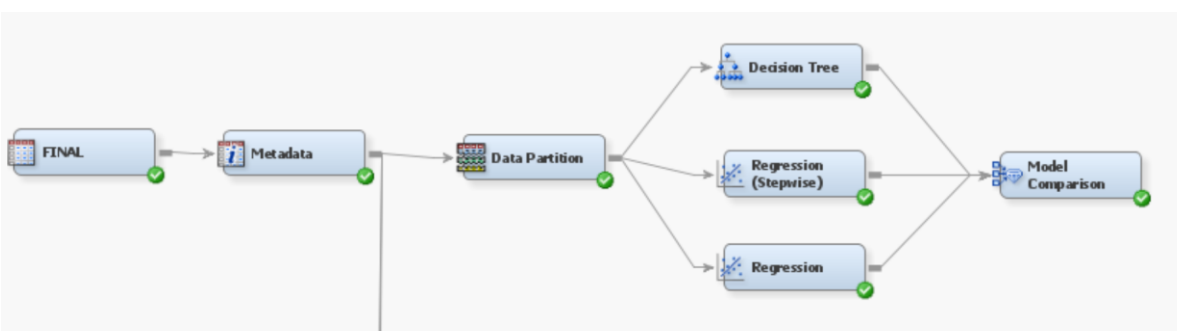


**Figure 2: Enterprise Miner node diagram**

## MODEL ASSESMENT

Decision Tree, Logistic Regression Model with stepwise variable selection and Logistic regression model without variable selection are the three best models that were selected based on least Misclassification rate.

model assesment

| Selected Model | Predecessor Node | Model Node | Model Description | Target Variable | Target Label | Selection Criterion: Valid: Misclassification Rate | Train: Sum of Frequencies | Train: Misclassification Rate | Train: Maximum Absolute Error | Train: Sum of Squared Errors | Train: Average Squared Error | Train: Root Average Squared Error |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Y | Tree | Tree | Decision Tr... | readmission | | 0.134 | 116179 | 0.133828 | 0.876545 | 26236.83 | 0.112916 | 0.336029 |
| | Reg2 | Reg2 | Regression | readmission | | 0.134181 | 116179 | 0.134052 | 0.988183 | 26471.24 | 0.113924 | 0.337527 |
| | Reg | Reg | Regression... | readmission | | 0.134201 | 116179 | 0.134138 | 0.972086 | 26480.12 | 0.113963 | 0.337583 |

**Table 2: Model assessment parameters**

| Predecessor Node | Model Node | Model Description | Target Variable | Target Label | Selection Criterion: Valid: Misclassification Rate |
|---|---|---|---|---|---|
| Tree | Tree | Decision Tree | readmission | | 0.134 |
| Reg2 | Reg2 | Regression | readmission | | 0.134181 |
| Reg | Reg | Regression (Stepwise) | readmission | | 0.134201 |

**Table 3: Selected Model**

Based on the results, decision tree is the best model selected with a Misclassification rate of 0.134. Stepwise Logistic regression and logistic regression without a variable selection also produced good results with almost the same Misclassification rate as before (0.134181).

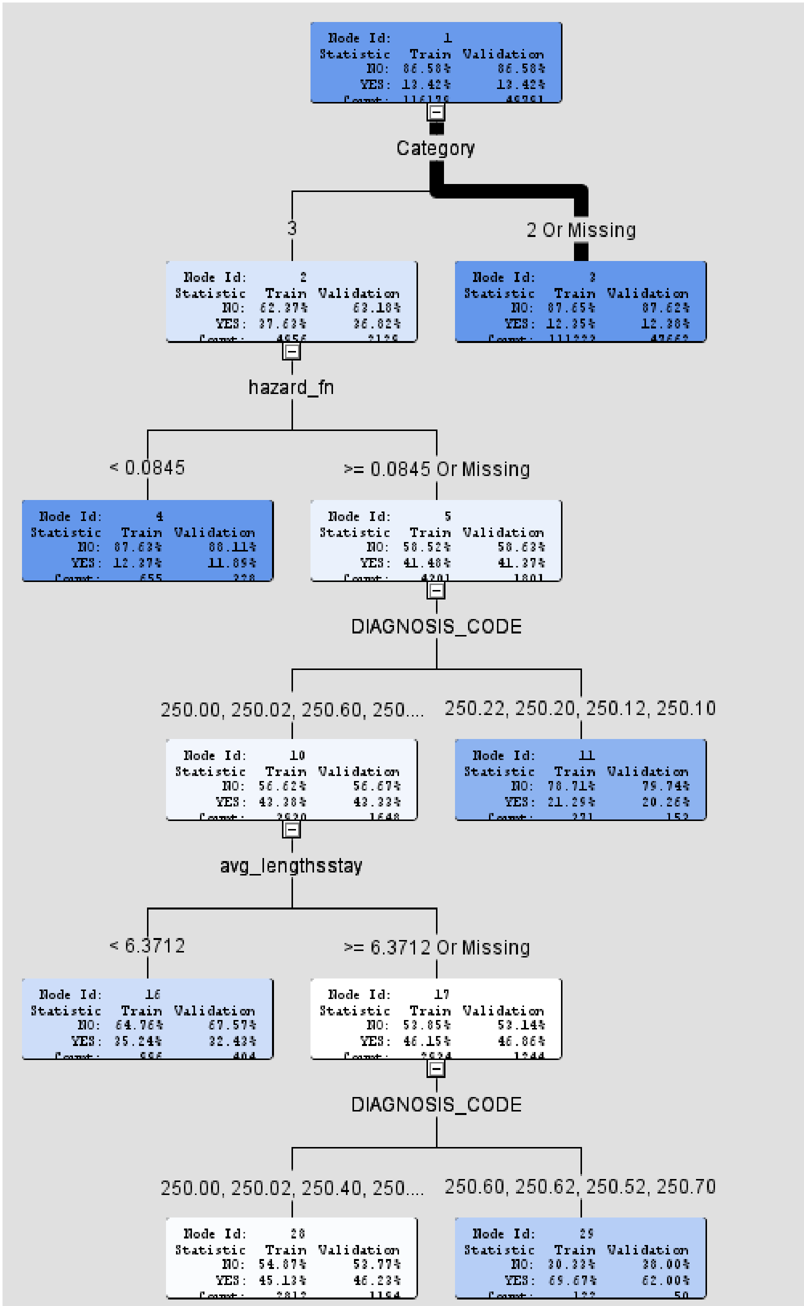


**Figure 3: Decision tree node diagram**

## VARIABLE IMPORTANCE ACCORDING TO DECISION TREE

Decision tree predicted category, hazard function, diagnosis code and average length of stay as the most important variables to predict the potential re-admission of the patient. Category is the patient prevention group category that we created earlier during data preparation. Hazard function is the Cumulative Hazard rate of hospital utilization. Diagnosis code is the ICD9 diagnosis code if the patient. Average length of stay is the average amount of time a patient is generally admitted in that particular hospital that patient is admitted. It can be seen that not only the patient attributes, but also the hospital attributes that gives information of the hospital where the patient is admitted are important for predicting the readmission. This importance is further analyzed in the next steps.

Variable importance according to decision tree

Variable Importance

| Variable Name | Label | Number of Splitting Rules | Importance | Validation Importance | Ratio of Validation to Training Importance |
|---|---|---|---|---|---|
| Category | | 1 | 1.0000 | 1.0000 | 1.0000 |
| hazard_fn | | 1 | 0.3985 | 0.4464 | 1.1200 |
| DIAGNOSIS_CODE | | 2 | 0.2788 | 0.2604 | 0.9340 |
| avg_lengthsstay | | 1 | 0.1708 | 0.2216 | 1.2977 |

**Figure 4: Variable importance**

### Effect of age of the patient on readmission:

Age of the patient doesn't have a significant effect on the probability of readmission. Average age of people who readmit is 66.7 and those who don't readmit is 64.5. This is only a two-year difference and doesn't show the effect. Also, most of the patients considered in this study are over 50 years and this value effect the average significantly.

| Readmission | |
|---|---|
| No | 64.470 |
| Yes | 66.707 |

**Table 4: Average age of readmission**

### Effect of average length of stay on readmission:

Through there is no significant difference in average length of stay of patients admitted to the hospital among readmitted and non-readmitted patients, decision tree is using this variable to classify patients.

| Readmission | Days |
|---|---|
| **No** | 6 |
| **Yes** | 6.4 |

**Table 5: Average length of stay of first visit**

### Effect of hospital hazard rate on readmission:

Hazard rate gives the proportion of total number of readmission encounters among total discharges in the particular hospital that the patient is admitted. This value is generally a very sensitive number and slight change in this number generally changes the health parameters drastically.

| Readmission | |
|---|---|
| **No** | 10% |
| **Yes** | 16% |

**Table 6: Average hazard rate**

## Effect of prevention category of disease on readmission:

The table shows the division of readmissions among all of the people, for secondary prevention group of diabetes and for tertiary prevention group. It can be seen that the probability of re-admission increases drastically from secondary prevention group to tertiary prevention group. This was also used the main criteria to predict the readmission by the decision tree model.

| Readmission | Overall division of readmissions | Division of readmissions among secondary prevention group (diabetes with no complications) | Division of readmissions among tertiary prevention group (diabetes with complications) |
|---|---|---|---|
| **No** | 86.5% | 87.6% | 62.6% |
| **Yes** | 13.5% | **12.4%** | **37.4%** |

**Table 7: Probability of readmission for each prevention group**

## Effect of payer of admission fee on readmission:

Most of the payers in the data are insurance providers and Medicare is the largest payer among the data present. It can be seen that patients who are paid via Medicare and Workers compensation are more probable to readmit. These percentages are shown in the table. Missing values are avoided in the below table.

| PAYER_CODE_DESC | Yes (%) | No (%) | Total (%) |
|---|---|---|---|
| Medicare | 15.09 | 84.91 | 41.09 |
| Worker's Compensation | 15.09 | 84.91 | 0.39 |
| Medicaid | 13.08 | 86.92 | 5.37 |
| Self-Pay | 12.63 | 87.37 | 4.39 |
| HMO/Managed Care (undesignated) | 12.53 | 87.47 | 3.01 |
| Other Government | 12.32 | 87.68 | 0.67 |
| Other Commercial Payer | 11.83 | 88.17 | 1.67 |
| PPO (undesignated) | 11.66 | 88.34 | 1.51 |
| Blue Cross/Blue Shield | 11.22 | 88.78 | 6.47 |
| CHAMPUS (Military dependents) | 10.13 | 89.87 | 0.55 |
| Medicare Managed Care (undesignated) | 7.66 | 92.34 | 1.71 |

**Table 8: Probability of readmission for each payer type**

## Effect of disease complications on readmission:

Patients with ICD9 diagnosis codes of "250.1", "250.12", "250.2", "250.22", "250.3", "250.32", "250.4", "250.42", "250.5", "250.52", "250.6", "250.62", "250.7", "250.72", “250.82”, “250.92” during their admission or readmission are those patients who have other complications along with Type-2 diabetes. As it can be seen, for most of these IDC9 codes people are more probable to re-admit. Patient encounters with 250.52, 250.42, 250.40, 250.50, 250.72 diagnosis codes have more probability of readmission. Type-2 diabetic patients with ophthalmic manifestations, renal manifestations and peripheral circulatory disorders (heart related disorders) are more probable to readmit. Readmissions of patients with peripheral circulatory disorders is common, but it is surprising to see that those with renal manifestations and ophthalmic manifestations also have a very high probability for readmission.

| Diagnosis code | No (%) | Yes (%) | Description of complication |
|---|---|---|---|
| **250** | 90.5 | 9.54 | Without mention of complication |
| **250.02** | 92.7 | 7.35 | Without mention of complication (Uncontrolled) |
| **250.10** | 74.5 | **25.49** | With ketoacidosis |
| **250.12** | 77.8 | 22.19 | With ketoacidosis (Uncontrolled) |
| **250.20** | 92.7 | 7.35 | With hyperosmolarity |
| **250.22** | 82.6 | 17.39 | With hyperosmolarity (Uncontrolled) |
| **250.30** | 100 | 0 | With other coma |
| **250.32** | 100 | 0 | With other coma (Uncontrolled) |
| **250.40** | 58.8 | **41.16** | With renal manifestations |
| **250.42** | 52.7 | **47.27** | With renal manifestations (Uncontrolled) |
| **250.50** | 59 | **41.05** | With ophthalmic manifestations |
| **250.52** | 42.9 | **57.14** | With ophthalmic manifestations (Uncontrolled) |
| **250.60** | 86.8 | 13.19 | With neurological manifestations |
| **250.62** | 78 | 21.99 | With neurological manifestations (Uncontrolled) |
| **250.70** | 87.4 | 12.62 | With peripheral circulatory disorders |
| **250.72** | 62.1 | **37.93** | With peripheral circulatory disorders (Uncontrolled) |
| **250.82** | 85.2 | 14.78 | With other specified manifestations (Uncontrolled) |
| **250.92** | 95.4 | 4.65 | With unspecified complication (Uncontrolled) |

**Table 9: Probability of readmission for each diagnosis category of patient**

## CONCLUSION

Unplanned hospital readmissions are expensive for patients and hospitals. By predicting these readmissions, hospitals and healthcare providers can effectively allocate their resources across departments. Building a model that predicts readmissions by using the data available publicly and through other various health data repositories, and understanding the characteristics of the patients that may be prone to readmission, can help hospitals/ healthcare providers to avoid readmissions. The decision tree model built has a good predictive accuracy with misclassification rate of 0.13. The model predicted that diabetes prevention category, diagnosis code, hazard rate and average length of stay in a hospital as the most important variables for predicting the potential readmission of a patient. It can be seen from the data that type-2 diabetic patients in tertiary prevention group (diabetes with complications) are more probable to readmit and among those patients with ophthalmic manifestations, renal manifestations and peripheral circulatory disorders (heart related disorders) are more probable to readmit.

## REFERENCES


Riikka Airaksinen, Panu Rantakokko, Johan G Eriksson, Paul Blomstedt, Eero Kajantie, Hannu Kiviranta. Association Between Type 2 Diabetes and Exposure to Persistent Organic Pollutants. Epidimeology/Health Services Research.

Matthew J Swaina, Hadi Kharrazib. Feasibility of 30-day hospital readmission prediction modeling based on health information exchange data. International Journal of Medical Informatics. 84 (2015) 1048–1056.

SF Jencks, MV Williams, EA Coleman. Rehospitalizations among patients in the Medicare fee-for-service program. N. Engl. J. Med. 360 (April (14)) (2009) 1418–1428.

Amit D. Raval, Steve Zhou, Wenhui Wei, Sandipan Bhattacharjee, Raymond Miao, Usha Sambamoorthi. 0-Day Readmission Among Elderly Medicare Beneficiaries with Type 2 Diabetes. Population Health Management. 2015 Aug 1; 18(4): 256–264.

Ostling S, Wyckoff J, Ciarkowski SL, Chih-Wen Pai, Hae Mi Choe, Vinita Bahl and Roma Gianchandan. The relationship between diabetes mellitus and 30-day readmission rates. March 22, 2017. Clinical Diabetes and Endocrinology 2017.

Examining the Drivers of Readmissions and Reducing Unnecessary Readmissions for Better Patient Care. American Hospital Association September 2011.

Saurabh S Kokad, Dr. Miriam McGaugh. 2017. Predictive Analysis for Classifying Type-2 Diabetes Patients Using SAS-Enterprise Miner 14.1 and SAS Enterprise Guide 7.1. Proceedings of SAS South Central Regional Conference, Dallas TX. Available at http://www.scsug.org/wp-content/uploads/2017/10/SK21.pdf.

Free ICD-9-CM Codes, Free ICD-10-CM/PCS Codes. http://www.icd9data.com.

## ACKNOWLEDGMENTS

This work was conducted with data from the Cerner Corporation's Health Facts database of electronic medical records provided by the Oklahoma State University Center for Health Systems Innovation (CHSI). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Cerner Corporation.

I thank my professors Dr. Miriam McGaugh and Dr. Goutam Chakraborty for their constant support and guidance through every part of this project.

## CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Shashank Reddy Gudipati
shashankgudipati@gmail.com