Hospital Readmission Analysis

Team Members:

Saké Group

Syed Omar Ahmed
Anushua Banerjee
Kunal Kadam
Eric S. Viacrusis

Date: May 17, 2025

Project Summary

The primary objective of this project is to predict hospital readmission using data for hospital admission. The secondary objective is to evaluate whether an initial diagnosis of diabetes can serve as a reliable predictor for hospital readmission.

The data contains patient information - age, initial duration at the hospital, tests, diagnosis (primary, secondary and tertiary) and readmission status. Given the target variable is categorical (yes/no) for readmission, two classification models were employed: Logistic Regression and Neural Networks.

For the Logistic Regression model, due to a high number of predictors, feature selection techniques were applied to enhance model performance, like Backward Elimination, Forward Selection, and Stepwise Selection. While the Exhaustive Search method was excluded due to its high computational cost, a Grid Search was implemented to optimize the neural network model by identifying the most effective number of hidden layer nodes for improved performance.

For the secondary objective, to assess if diabetes can be used to predict hospital readmission, a graphical presentation was used to compare the results to other diagnoses to aid the data analyst to understand the effect to readmission; while the coefficients were used to finalize the conclusion.

The application of the model shall help in better hospital operations - readmissions are costly and often indicative of gaps in care delivery. High readmission rates are associated with increased healthcare expenses and poor patient outcomes. Predicting which patients are at high risk of readmission can enable healthcare providers to deliver preventative interventions and reduce overall readmission rates.

Introduction

Hospital readmission rates are widely recognized as a critical indicator of healthcare quality, efficiency, and patient outcomes. Elevated readmission rates often point to underlying systemic issues, such as insufficient discharge planning, inadequate patient education, or a lack of proper follow-up care. In the context of a healthcare industry moving steadily toward value-based care models, reducing avoidable readmissions has become a key priority. Addressing this issue not only improves the quality of care delivered to patients but also helps healthcare providers avoid penalties, reduce operational costs, and enhance long-term patient satisfaction and health outcomes.

Creating a comprehensive data analysis can significantly aid in identifying the factors contributing to readmissions, allowing healthcare providers to implement targeted interventions. This proactive approach not only benefits patients but also aligns with industry standards aimed at fostering accountability and excellence in healthcare delivery.

By leveraging the Kaggle hospital readmissions dataset, we were able to simulate real-world hospital scenarios and develop predictive models that offer meaningful insights into readmission patterns.

Main Chapter

A. Develop the Understanding

- The primary objective of this project is to predict hospital readmission using data for hospital admission.
- The second objective is to determine if patients with diabetes can be used to predict hospital readmission.
- Once the model is created, it can be used to identify admitted patients' likelihood of readmission, and they may use it to improve hospital services to reduce it.
- One of the outputs of the project is to identify diabetes as a criterion of readmission, if this is determined, hospitals can adjust patient care targeting diabetes patients improved health care services.

B. Obtain Data for Analysis

To conduct our analysis on hospital readmissions, we first identified and sourced a publicly available dataset that provides relevant and comprehensive healthcare data. After exploring various repositories, we selected a dataset from Kaggle titled "hospital Readmissions"

This dataset includes detailed patient-level information from over 100,000 hospital encounters for diabetic patients. It contains variables related to demographics, admission and discharge details, medical diagnoses, length of stay, medications, and whether the patient was readmitted. Specifically, it allows us to investigate factors influencing hospital readmission, making it suitable for predictive modeling and exploratory analysis in our case.

C. Explore, Clean and Preprocess Data / Reduce the Data Dimension

1. Number of Records: 25,000

2. Number of Columns: 17

```
Hospital Readmissions Dimension: (25000, 17)
```

3. Columns:

```
Index(['age', 'time_in_hospital', 'n_lab_procedures',
'n_procedures','n_medications', 'n_outpatient', 'n_inpatient',
'n_emergency','medical_specialty', 'diag_1', 'diag_2', 'diag_3',
'glucose_test','A1Ctest', 'change', 'diabetes_med',
'readmitted'],dtype='object')
```

4. Column Data Types

```
age                  object
time_in_hospital      int64
n_lab_procedures      int64
n_procedures          int64
n_medications         int64
n_outpatient          int64
n_inpatient           int64
n_emergency           int64
medical_specialty    object
diag_1               object
diag_2               object
diag_3               object
glucose_test         object
A1Ctest              object
change               object
diabetes_med         object
readmitted           object
dtype: object
```

5. Converted Column Data Types

The fields that are non-numeric are converted to dummy variables to be used in the analysis. The result is 46 columns.

```
time_in_hospital                               int32
n_lab_procedures                               int32
n_procedures                                   int32
n_medications                                  int32
n_outpatient                                   int32
n_inpatient                                    int32
n_emergency                                    int32
age_[50-60)                                    int32
age_[60-70)                                    int32
age_[70-80)                                    int32
age_[80-90)                                    int32
age_[90-100)                                   int32
medical_specialty_Emergency/Trauma             int32
medical_specialty_Family/GeneralPractice       int32
medical_specialty_InternalMedicine             int32
medical_specialty_Missing                      int32
medical_specialty_Other                        int32
medical_specialty_Surgery                      int32
diag_1_Diabetes                                int32
diag_1_Digestive                               int32
diag_1_Injury                                  int32
diag_1_Missing                                 int32
diag_1_Musculoskeletal                         int32
diag_1_Other                                   int32
diag_1_Respiratory                             int32
diag_2_Diabetes                                int32
diag_2_Digestive                               int32
diag_2_Injury                                  int32
diag_2_Missing                                 int32
diag_2_Musculoskeletal                         int32
diag_2_Other                                   int32
diag_2_Respiratory                             int32
diag_3_Diabetes                                int32
diag_3_Digestive                               int32
diag_3_Injury                                  int32
diag_3_Missing                                 int32
diag_3_Musculoskeletal                         int32
diag_3_Other                                   int32
diag_3_Respiratory                             int32
glucose_test_no                                int32
glucose_test_normal                            int32
A1Ctest_no                                     int32
A1Ctest_normal                                 int32
change_yes                                     int32
diabetes_med_yes                               int32
readmitted_yes                                 int32
dtype: object
```

D. Determine the Data Mining Task

- The identified output column is “readmitted” with two outputs “Yes” or “No.” From “readmitted” it was converted to “readmitted_Yes” with either “0” or “1” value.
- The initial predictor variables are listed below:

```
Index(['time_in_hospital', 'n_lab_procedures', 'n_procedures', 'n_medications',
       'n_outpatient', 'n_inpatient', 'n_emergency', 'age_[50-60)',
       'age_[60-70)', 'age_[70-80)', 'age_[80-90)', 'age_[90-100)',
       'medical_specialty_Emergency/Trauma',
       'medical_specialty_Family/GeneralPractice',
       'medical_specialty_InternalMedicine', 'medical_specialty_Missing',
       'medical_specialty_Other', 'medical_specialty_Surgery',
       'diag_1_Diabetes', 'diag_1_Digestive', 'diag_1_Injury',
       'diag_1_Missing', 'diag_1_Musculoskeletal', 'diag_1_Other',
       'diag_1_Respiratory', 'diag_2_Diabetes', 'diag_2_Digestive',
       'diag_2_Injury', 'diag_2_Missing', 'diag_2_Musculoskeletal',
       'diag_2_Other', 'diag_2_Respiratory', 'diag_3_Diabetes',
       'diag_3_Digestive', 'diag_3_Injury', 'diag_3_Missing',
       'diag_3_Musculoskeletal', 'diag_3_Other', 'diag_3_Respiratory',
       'glucose_test_no', 'glucose_test_normal', 'A1Ctest_no',
       'A1Ctest_normal', 'change_yes', 'diabetes_med_yes'],
      dtype='object')
```

E. Partition the Data

To ensure there is no overfitting, initial data was partitioned into 60% Training data, and 40% Validation data using the ***train_test_split*** function. Other models was adjusted accordingly - 80%/20% split, and 70%/30% split.

F. Techniques

(1) Since the identified output variable is categorical rather than numeric, a classification model was used. It was identified to use Logistic Regression as the initial model, and the Neural Net was selected to ensure complicated relationships can be covered.

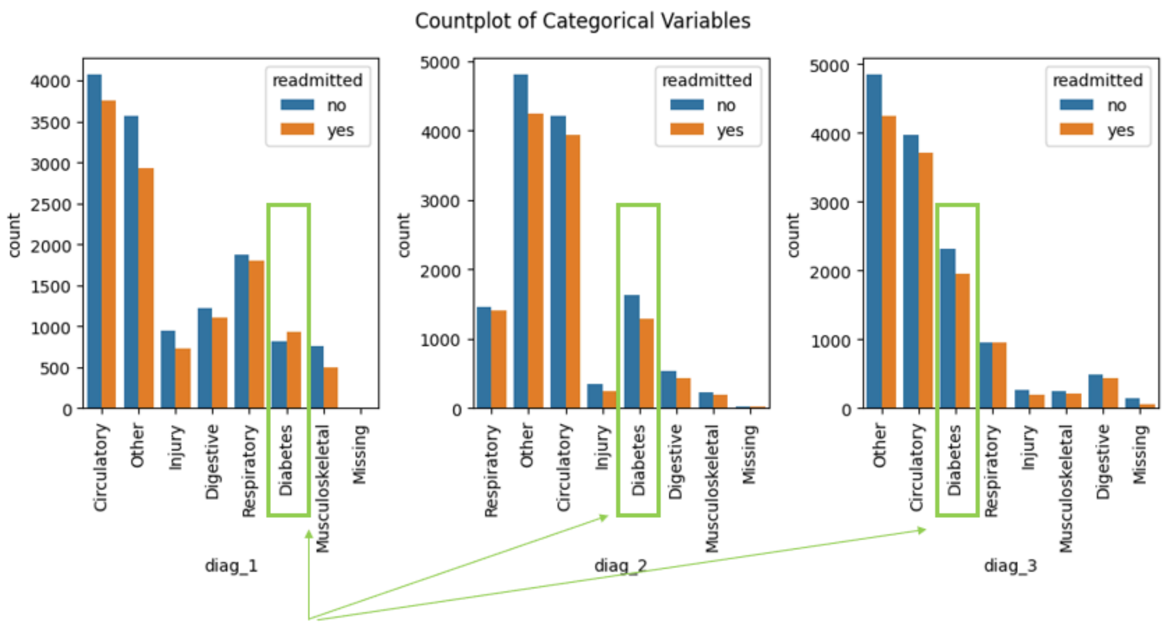
(2) Logistic Regression - a method of reduction through Forward Elimination, Backward Elimination and Stepwise was used to improve the accuracy of the model.

(3) Neural Nets - to refine the model, GridSearch was used to better improve the result.

G. Chart, Algorithm and Measures

LOGISTIC REGRESSION MODEL

<u>Bar Chart Graph for Data Visualization of Diagnosis Readmission</u>



The bar chart shows primary, secondary and tertiary diagnosis and their corresponding readmission count.

The Logistic Regression Model with Multiple Predictors

```
Parameters of Logistic Regresion Model with Multiple Predictors
Intercept: -0.663
Coefficients for Predictors
       time_in_hospital  n_lab_procedures  n_procedures  n_medications  \
Coeff:            0.017             0.002        -0.028          0.001

       n_outpatient  n_inpatient  n_emergency  age_[50-60)  age_[60-70)  \
Coeff:        0.119        0.391         0.25       -0.025        0.111

       age_[70-80)  age_[80-90)  age_[90-100)  \
Coeff:       0.152        0.225        -0.106

       medical_specialty_Emergency/Trauma  \
Coeff:                             -0.003

       medical_specialty_Family/GeneralPractice  \
Coeff:                                   -0.006

       medical_specialty_InternalMedicine  medical_specialty_Missing  \
Coeff:                             -0.196                      0.021

       medical_specialty_Other  medical_specialty_Surgery  diag_1_Diabetes  \
Coeff:                   -0.17                     -0.196            0.242

       diag_1_Digestive  diag_1_Injury  diag_1_Missing  \
Coeff:           -0.013         -0.206          -0.801

       diag_1_Musculoskeletal  diag_1_Other  diag_1_Respiratory  \
Coeff:                 -0.098        -0.165              -0.035

       diag_2_Diabetes  diag_2_Digestive  diag_2_Injury  diag_2_Missing  \
Coeff           -0.027            -0.231         -0.232           0.135

       diag_2_Musculoskeletal  diag_2_Other  diag_2_Respiratory  \
Coeff:                 -0.175         -0.08              -0.098

       diag_3_Diabetes  diag_3_Digestive  diag_3_Injury  diag_3_Missing  \
Coeff           -0.011            -0.141         -0.135          -0.858

       diag_3_Musculoskeletal  diag_3_Other  diag_3_Respiratory  \
Coeff:                 -0.129        -0.051              -0.032

       glucose_test_no  glucose_test_normal  A1Ctest_no  A1Ctest_normal  \
Coeff:          -0.044                -0.08       0.055          -0.081

       change_yes  diabetes_med_yes
Coeff:      0.005             0.277
```

Based on the coefficient for predictors, patients that has an initial diagnosis of Diabetes has a positive change on the readmission, whereas Secondary or Tertiary diagnosis has inverse increase.

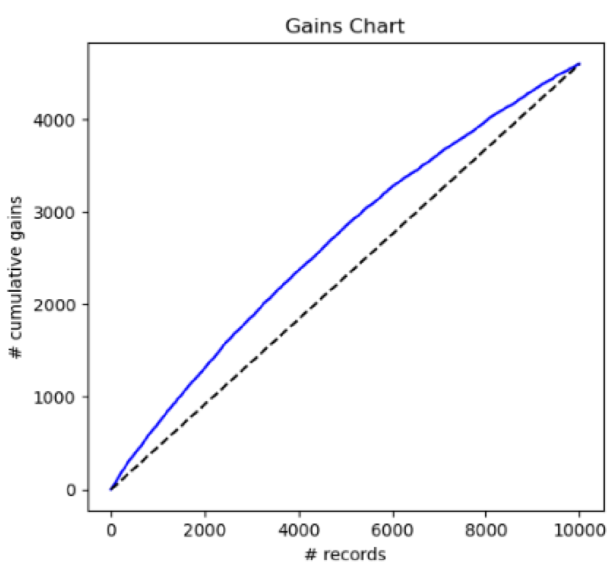
Further checking other predictors - A1Ctest_no, change_yes, and diabetes_med_yes has a positive change on the readmission.

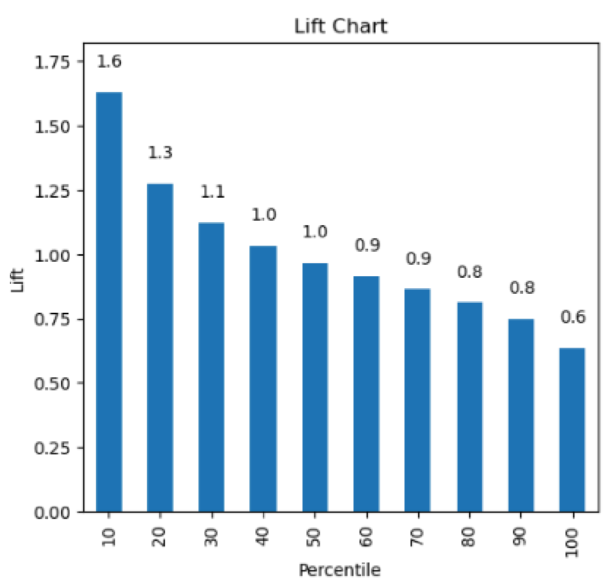
<u>**The Logistic Regression Model - Actual vs Classification:**</u>

Classification for Validation Partition

	Actual	Classification	p(0)	p(1)
21492	0	0	0.6487	0.3513
9488	0	0	0.6551	0.3449
16933	0	0	0.6117	0.3883
12604	1	0	0.6269	0.3731
8222	0	0	0.5476	0.4524
9110	1	1	0.0063	0.9937
21196	1	1	0.4976	0.5024
17193	1	0	0.5673	0.4327
23846	1	1	0.2786	0.7214
10415	0	0	0.5606	0.4394
9018	0	1	0.4652	0.5348
24056	0	0	0.5196	0.4804
19992	0	0	0.5554	0.4446
11464	0	0	0.6402	0.3598
10641	1	0	0.6017	0.3983
7192	0	0	0.6156	0.3844
2164	0	0	0.6636	0.3364
2277	0	0	0.5558	0.4442
6038	0	0	0.6985	0.3015
15100	0	0	0.5593	0.4407

The image shows that some records are misclassified.





Gains Chart:

- The blue curve shows the cumulative number of actual readmissions correctly identified as we move through the top-ranked predicted records (sorted by predicted probability)
- The model is better than random, as the blue line stays above the baseline.
- But it doesn't rise steeply - indicating limited lift.

Lift Chart:

- Top 10% (decile 1) has a lift of 1.6
- The top 10% of predicted patients are 1.6× more likely to be true readmissions than random guessing.
- The model performs reasonably well in the top decile.
- Lift quickly drops in later deciles, indicating limited separation power.

Confusion Matrix for Logistic Regression (Model 1)

```
Training Partition
Confusion Matrix (Accuracy 0.6123)

       Prediction
Actual    0    1
     0 6056 1783
     1 4033 3128

Validation Partition
Confusion Matrix (Accuracy 0.6157)

       Prediction
Actual    0    1
     0 4125 1282
     1 2561 2032
```

NEURAL NETS MODEL

Final Intercept for Hospital Readmission Neural Network Model

```
Final Intercepts for Hospital Readmission Neural Network Model
[array([-0.13382692, -0.32705677, -0.32341174,  0.32812926,  0.01950466,
        0.01413828,  0.29372219, -0.39160841, -0.23475831, -0.06540316,
        0.57575916, -0.58383911, -0.09282722,  0.08503826,  0.2956193 ,
        0.13718262, -0.0961369 ,  0.50938553, -0.11376415,  0.08993338]), array([-0.24266256])]
```

Network Weights for Hospital Readmission Neural Network Model

```
Network Weights for Hospital Readmission Neural Network Model
[array([[ 1.94016029e-01, -4.46890264e-02, -2.68760235e-01,
         -1.63798950e-01, -5.55263297e-01,  8.71055694e-02,
         -2.70130139e-02, -5.20525018e-02,  1.03467347e-01,
         -1.11274938e-01,  1.74896651e-01,  4.31969465e-01,
          9.19083814e-02, -6.02009461e-02, -3.07351936e-01,
         -1.36132203e-02, -5.88076878e-01,  1.38134826e-01,
         -4.37431583e-01, -2.53376680e-01],
        [-1.85019704e-01,  3.50987629e-02, -6.41478261e-01,
          1.85257091e-01,  1.00912429e-01,  7.89835399e-01,
         -1.44232533e-01, -1.52201999e-01, -4.13748906e-01,
          3.99289554e-01,  3.81634423e-02, -1.78569045e-01,
          6.03877590e-01, -2.34513673e-01, -3.86433522e-01,
         -4.54954680e-01,  1.31249721e-01,  1.82061684e-01,
         -1.09675391e-01, -1.91822169e-01],
        [ 1.09287582e-01,  2.30238093e-01, -4.52444802e-01,
          3.75361109e-02, -2.98246823e-01,  1.98916549e-01,
         -4.43787386e-02, -5.63525933e-01, -2.21275173e-01,
         -2.59439898e-01,  7.87384638e-03,  1.88758876e-01,
          1.59432649e-02, -9.94052772e-02,  1.55931821e-01,
         -3.38527315e-02,  7.28329159e-04, -3.85991990e-02,
          1.92253741e-01, -9.06551452e-02],
        [-1.15801924e-01,  7.21778597e-02, -5.01380612e-02,
          1.83327727e-01,  2.33801120e-01, -3.36098205e-01,
          2.41737706e-01, -1.89126602e-01, -4.01904390e-01,
          7.83622569e-01,  5.50769587e-01, -2.83354324e-01,
         -7.32478655e-01,  1.80487528e-01,  5.44497643e-01,
         -1.55096666e-01,  2.63340524e-01,  5.61828804e-01,
          5.70417286e-01,  4.49116931e-01],
        [ 1.78104443e-01, -1.07161569e-02,  7.54182784e-02,
          5.84221683e-02,  4.16197077e-01, -2.62973653e-01,
         -2.81164779e-01,  1.34180607e-01, -8.05097515e-01,
          2.41548252e-01, -5.19096982e-01,  2.68600437e-01,
         -7.04398304e-01,  4.38786727e-02, -1.56194369e-02,
          1.49937949e-01,  3.22842256e-01,  6.44998059e-02,
         -1.49695542e-01,  7.93751955e-02],
        [-1.02486360e+00, -4.82679413e-01,  4.17170470e-01,
         -2.36444943e-01,  3.09136953e-01, -5.21076906e-01,
          2.20886831e-01,  3.56693122e-01, -1.22471475e-01,
          6.51038103e-01, -5.64810465e-01, -3.99594868e-01,
         -4.87261102e-01, -2.97303653e-01, -3.05006822e-01,
         -1.35029983e-01, -2.19272221e-01,  5.26928223e-01,
          1.90317921e-01, -2.38580492e-01],
        [ 1.07259599e-01, -1.13874804e-01, -2.45423091e-01,
         -3.33320887e-01, -4.10430570e-01, -2.59666496e-01,
         -3.60410073e-01, -5.38376903e-01, -1.64838846e-01,
         -5.99996239e-01, -5.76439517e-02,  2.45327790e-01,
         -3.63926776e-01, -1.32342113e-01,  1.82053696e-02,
          3.77414817e-01, -4.84767499e-01, -6.40619285e-02,
         -5.06093764e-05,  2.98520709e-01],
        [-1.06761140e-01, -1.41548198e-02, -2.95933399e-02,
          2.05004801e-01, -2.23637294e-01, -2.75362886e-01,
         -2.93639156e-01, -1.94306203e-01,  1.40728698e-01,
         -1.75765526e-01, -1.05777816e-01, -3.82993893e-02,
         -2.12707545e-01, -7.23912861e-02,  3.73472010e-02,
          3.04375151e-01,  1.09250680e-01, -9.76434280e-02,
          3.17120136e-01, -1.67561130e-02],
        [ 1.99309147e-01,  2.15239483e-01,  2.26262704e-01,
         -1.25587121e-01,  5.77774257e-02,  8.98527834e-02,
         -5.30305296e-01,  4.43530001e-01, -1.70629115e-01,
          2.49275642e-01, -5.67555225e-01, -4.34960212e-02,
          5.37443651e-02, -2.34075153e-02,  1.22924371e-01,
          3.24435464e-01, -7.33885034e-02, -3.38688567e-01,
          2.99827929e-01,  3.08123755e-01],
        [ 1.60843120e-01,  6.45276049e-03,  3.00210681e-01,
         -2.06872002e-01,  1.71023378e-01, -1.76070126e-01,
         -2.12004353e-01,  9.01159832e-03,  6.91338518e-02,
          4.22442328e-01, -1.46728842e-01,  4.19701495e-01,
         -1.93370487e-01, -2.23819947e-01, -4.46007998e-01,
         -7.34161047e-02, -2.95641565e-01,  3.23022427e-01,
         -2.04367957e-01,  6.46718674e-03],
        [ 4.15722323e-02, -3.45124651e-01, -3.15526836e-01,
          1.19835750e-01,  8.78050426e-02,  3.28159637e-02,
         -3.27385836e-01,  2.00721296e-02,  3.77802772e-01,
          6.55644929e-01, -1.00245127e-01, -4.96719787e-02,
         -1.47428519e-01, -8.87858560e-02, -4.06370851e-01,
         -7.33272947e-01,  1.15058076e-01, -1.48066129e-01,
         -1.49151395e-01,  4.18342763e-01],
        [ 8.81375473e-02,  6.08513108e-02, -1.82177929e-01,
         -5.12588021e-01, -5.16056237e-02,  3.06746730e-01,
         -5.00869324e-02,  1.10701213e-01,  2.07123444e-01,
         -2.24118385e-01, -8.28182484e-02,  1.97782137e-01,
          1.79383007e-01,  2.24211140e-01,  8.18251292e-03,
         -2.14406220e-01, -5.44082492e-02, -3.32797389e-02,
          1.13425274e-01, -2.32886459e-01],
        [ 1.57106620e-01,  4.92873003e-01,  4.04018994e-02,
         -3.05561972e-01, -3.25834449e-01, -1.60112319e-02,
         -1.85587294e-01, -1.57398448e-01,  3.41194223e-01,
         -4.62903523e-01, -5.60368711e-02, -6.97781785e-01,
          7.58596724e-02,  3.18171785e-01,  4.47922249e-01,
          4.68832551e-01, -3.09751902e-01, -1.09682331e-01,
         -6.53554496e-03,  1.02133098e-01],
        [ 1.94036188e-02,  2.98270587e-01,  3.05904560e-02,
         -4.24765906e-01,  1.57814118e-02,  7.99756134e-01,
         -5.56466486e-02,  3.93493271e-01,  2.80672049e-01,
          2.27455397e-01,  4.89868504e-02, -3.60143561e-02,
          6.76165854e-01,  3.38893041e-01,  2.61785676e-01,
         -3.71821814e-01,  9.03981500e-02,  6.95504643e-02,
          5.38678328e-01,  2.11330448e-02],
        [ 1.18643945e-01,  7.93104118e-02, -6.56163154e-01,
         -4.37292086e-01, -5.56148407e-01,  4.65612901e-01,
         -9.78473592e-03, -1.69108045e-01, -1.60317739e-01,
         -4.29989729e-01,  2.03533898e-01,  3.08383616e-01,
         -1.64508060e-03, -1.41185108e-01, -7.09745112e-03,
          2.70128586e-02, -1.91992445e-01, -7.90744802e-02,
         -1.23564083e-01,  1.40696467e-01],
        [ 3.02989345e-01, -2.39837037e-02, -7.60920620e-02,
         -8.52982563e-01, -1.78866015e-01, -4.17466068e-02,
         -5.96984909e-01,  1.97866964e-01,  4.37482254e-01,
         -9.57063335e-02,  1.13528884e-01,  2.97769242e-01,
         -2.12722301e-01,  1.51054269e-01, -6.68807455e-02,
         -5.52647622e-01, -4.58586011e-01,  4.02584806e-01,
          6.62617379e-02,  1.15954902e-01],
        [ 4.99938724e-02,  8.52768567e-02,  1.11016529e-01,
          6.93196852e-02,  7.11833499e-02, -5.29323573e-01,
          2.06698024e-01,  1.97848480e-01, -2.49513876e-01,
         -1.47123164e-02, -7.50242911e-02,  6.62247417e-02,
         -7.72204089e-01,  5.10372335e-03, -5.54128601e-02,
         -1.43106867e-01,  3.10097048e-01, -5.28012017e-01,
         -5.92872881e-02,  2.18603288e-01],
        [-8.92405878e-02,  2.18327156e-01,  7.01119420e-01,
          1.42933323e-02,  3.74682615e-01, -2.39550931e-01,
          1.42421251e-01,  3.30413679e-01, -1.62291372e-01,
          2.90505078e-02, -1.63637146e-01, -3.44226362e-01,
          2.27951834e-01,  5.49432819e-01,  1.26980217e-01,
          2.80461699e-01,  3.18456157e-01, -2.57360424e-01,
          7.12309340e-02, -3.75086801e-01],
        [ 2.09717322e-02,  1.86633055e-01,  2.86658871e-03,
         -4.29808401e-01, -8.69541601e-02, -9.60959247e-02,
         -1.72122342e-01, -5.28667480e-02,  3.55074103e-01,
          5.41039173e-02, -1.38897863e-01, -2.37557659e-01,
         -2.08786255e-02, -3.45076776e-01,  1.29258315e-01,
          1.05490540e-01, -4.69314056e-01,  2.88333601e-01,
         -3.70970703e-01,  1.95004738e-01],
        [ 4.13002765e-01,  3.72964468e-01, -2.02936841e-01,
          2.04545799e-01,  1.36518083e-01,  2.65168455e-01,
         -4.37145012e-02, -1.08307839e-01, -1.20476304e-01,
          3.71217089e-01,  1.35168979e-01, -1.45749741e-01,
          2.15249597e-01,  1.47626651e-01,  6.30044248e-01,
          2.71887095e-01,  3.98822799e-01,  1.48852932e-01,
          1.56535012e-01,  9.03645352e-01],
        [-2.16871557e-01, -2.47337859e-01, -2.91832240e-01,
         -2.63956572e-01, -2.49296541e-02, -2.36175397e-01,
         -2.86681045e-01,  1.37619453e-01, -6.01578214e-02,
```

```
 2.31855922e-01, -5.63141323e-02, -4.75174140e-02,
-1.07114892e-02, -1.26569466e-02, -2.49743900e-02,
-6.17945127e-02,  1.23842403e-01, -2.86304293e-01,
 2.21679509e-01, -2.28454189e-01],
[-1.83259160e-01,  7.09427788e-02,  3.30844222e-01,
-3.23914264e-01,  1.20222200e-01,  5.30277805e-01,
 3.31550929e-01,  1.90998488e-01,  3.85563054e-01,
-2.81527904e-02, -6.53060580e-02, -6.30990942e-02,
 9.64040322e-01,  3.91756696e-01,  3.45508644e-01,
 1.95279000e-01, -4.24371729e-02, -2.23240747e-01,
 1.07536322e-01, -2.63818299e-01],
[-3.82988003e-01, -1.08468593e-01, -1.03844781e-01,
-2.78278239e-01, -1.36681208e-01, -2.36113718e-01,
 3.67408364e-01, -5.28418463e-01,  3.65751256e-02,
 6.23626687e-02,  1.48237926e-01, -1.99401717e-01,
-1.65255392e-01, -1.50487192e-01,  1.54487901e-01,
-1.52076446e-01, -3.07377889e-01,  3.86012650e-01,
-1.87247171e-01,  7.61853064e-02],
[ 1.55624032e-01,  4.88339978e-01, -1.31165960e-01,
-1.84315411e-01, -6.00668705e-01,  1.86947418e-01,
-1.53729815e-02, -3.84517071e-01, -1.26087922e-01,
-7.77642056e-04,  4.71193711e-01,  2.52317167e-01,
 2.78674174e-01,  2.35280657e-01,  4.03766071e-01,
 1.67958494e-01, -4.80297970e-01,  2.89296282e-01,
 2.47493885e-01,  1.36855588e-01],
[-5.21545460e-01,  1.10327343e-01, -1.25549994e-01,
-2.54095453e-01,  3.92634637e-01,  3.69160065e-01,
 1.72520074e-01, -5.97524539e-01, -1.18826094e-01,
-4.61637944e-02,  4.66073073e-02, -3.58862298e-01,
 1.47601380e-01, -8.71389615e-02,  1.18967375e-01,
 4.16487628e-03,  1.10107772e-01, -3.52922543e-01,
 1.88417609e-01, -1.39980422e-01],
[-3.73062005e-01,  3.67358942e-01,  3.32107185e-01,
 1.97542305e-01,  5.35336234e-01, -1.50038359e-02,
 9.01524611e-02, -3.89768664e-01,  3.18293182e-01,
-1.35860724e-01,  6.79948459e-01, -6.79420371e-01,
 2.50156855e-01,  1.91968847e-01,  1.12398008e-01,
 3.78708164e-01,  2.62965540e-01, -4.40977531e-02,
-1.32918400e-01, -2.25558300e-01],
[ 1.26602423e-01,  6.83296284e-02, -5.53341895e-02,
 1.46599252e-01, -5.40095989e-02,  5.87267386e-02,
-1.75067916e-01, -1.00576213e-01,  3.22731010e-01,
-3.48179017e-02,  5.68845797e-02, -8.53879583e-02,
 1.32972603e-01, -3.68545356e-02, -1.26564519e-01,
-2.17507989e-01,  3.43729075e-02, -5.61651612e-02,
 7.30372406e-02,  7.70525912e-02],
[-2.90288824e-01, -4.52965999e-02,  2.15444034e-01,
-2.97425145e-01,  4.77013872e-02,  7.34846831e-02,
-3.06378332e-02,  1.91380967e-01, -8.25175340e-02,
-3.06228505e-01,  1.04537636e-01,  2.22322099e-01,
 2.71124779e-01,  3.43532679e-02,  5.59551778e-02,
 6.03172407e-02, -1.31348935e-01, -1.39839647e-01,
 8.33149332e-02, -9.15241138e-02]
[ 2.58415103e-01, -4.72886700e-02,  2.01332950e-01,
-2.87300874e-01,  3.54185751e-01, -1.13017361e-01,
 2.51094377e-01,  2.52938624e-02,  1.20481720e-01,
 1.80151603e-01, -1.06854742e-01, -1.64883270e-01,
-1.39568562e-01, -3.12703449e-01, -5.50553313e-01,
-1.49562297e-01, -2.21439117e-01,  9.02190949e-02,
-2.85160995e-01, -4.91965411e-02],
[ 1.87009026e-01,  4.63003074e-01, -4.78351700e-01,
-2.20019858e-01, -8.35276993e-02,  3.28412996e-01,
 1.64586680e-01, -3.53819935e-01,  3.27348725e-01,
 3.88663179e-02, -3.36581510e-01, -2.08570739e-01,
 1.94236151e-01, -1.62545643e-01, -6.24014021e-02,
-1.82851222e-01, -2.18494863e-01, -6.57233039e-01,
-2.18891586e-01,  2.00004414e-01],
[ 2.53437653e-02,  1.52716345e-02, -8.63380908e-02,
 2.93409801e-02,  4.17537054e-01, -1.92935136e-01,
-7.19483297e-02, -1.71273673e-01,  7.75600827e-02,
 9.63990265e-02, -2.87575647e-01,  2.48216683e-02,
-1.40570436e-01,  1.65298061e-01,  7.13710787e-02,
 1.18611713e-02,  1.90891895e-01, -3.79981194e-01,
-8.77140876e-02, -1.34638757e-01],
[ 2.89223323e-01,  4.63236425e-01, -1.67215005e-01,
 3.39820744e-01,  2.72497283e-01,  1.87494242e-01,
-4.83879291e-02,  2.97557107e-02, -1.63744327e-01,
 4.07473803e-02, -4.15590896e-01,  1.17301774e-01,
 2.51051745e-01, -2.20338018e-01,  2.86354161e-02,
-3.21658438e-01,  6.93610281e-01, -4.45787864e-01,
 5.71665142e-01, -1.68967223e-02],
[ 3.02064701e-01,  1.90802028e-01, -5.02343375e-02,
-2.07009649e-01, -6.99425096e-01,  7.35518965e-02,
-8.82096041e-02, -1.73190189e-01, -3.88447860e-02,
-1.57957062e-01,  1.00083272e+00, -1.57100915e-01,
 2.57143351e-02,  8.65268892e-01,  6.71546952e-01,
 3.52991253e-01, -4.14261314e-01,  8.54333384e-01,
-2.03824620e-01, -8.48511347e-03],
[ 4.06348203e-01,  2.72574293e-01, -1.53854387e-02,
-5.10193145e-02, -2.33943229e-02,  2.63062630e-01,
-5.17371701e-02,  3.25412314e-01,  2.93721853e-02,
 1.74639196e-01, -2.50561167e-01, -1.78318064e-02,
 1.28610597e-01, -1.34937732e-01, -2.97042905e-01,
-1.93682742e-01, -2.83191186e-01,  1.94005455e-01,
-5.23851002e-01, -6.41373221e-01],
[-2.21302261e-01, -1.04757450e-01, -3.74434154e-02,
-9.75397434e-02, -3.60973516e-03, -1.52866146e-01,
 1.69003698e-01,  1.79774118e-01, -3.10310750e-01,
 4.24516997e-01, -3.04579809e-02, -3.61140595e-02,
-4.03961076e-01, -2.39665588e-01,  7.85162527e-02,
-2.76458302e-02,  9.39100889e-02, -3.02598030e-01,
 4.12993936e-01, -1.43969925e-01],
[-5.67994565e-01,  6.22033801e-02,  1.15904365e-01,
 2.88470025e-01,  1.65152231e-01, -2.06110832e-02,
-1.74139024e-02,  1.42622986e-02, -1.71401263e-01,
 9.50671095e-02, -1.63418937e-01,  9.41852351e-02,
 1.31185102e-01, -1.09157081e-02, -3.71663593e-01,
-1.20644595e-01,  2.67740362e-01, -1.16398245e-01,
-7.34793575e-02, -7.43845103e-01],
[ 1.29357725e-01, -4.34762241e-01, -2.65152170e-02,
-1.44717922e-01, -3.40364350e-01,  2.55997147e-01,
-2.55198068e-02,  4.24300921e-01, -2.60824167e-01,
 1.11253105e-01,  3.03196387e-02, -3.13339230e-01,
 3.25298930e-01,  1.99992019e-01, -1.74153388e-01,
 1.09786758e-01, -1.48623687e-01,  4.42295102e-01,
-2.51693994e-01,  4.26875835e-02],
[-1.40240508e-01,  3.08567185e-01,  7.09795098e-02,
-3.01730692e-01, -3.66115924e-01, -4.29411081e-01,
-1.28282914e-02,  2.85658550e-01,  3.72237636e-01,
-1.79174728e-01,  5.37108776e-02,  1.33198837e-01,
-1.07682910e-01, -2.40091341e-01,  6.87920739e-02,
-4.57494628e-02, -3.46030395e-01,  3.80098265e-03,
 1.82961945e-01, -4.02434264e-01],
[ 3.04027832e-01,  1.11607546e+00,  3.00536504e-02,
 2.83409380e-01, -3.44714859e-01, -3.24811476e-01,
 8.73167671e-02,  6.10162097e-01, -2.42183907e-02,
-2.80081331e-01, -1.42810536e-01,  1.60696732e-01,
-1.92472537e-01, -1.47642266e-02, -8.76213658e-02,
 2.48078526e-01,  2.72143517e-01, -5.22916429e-01,
 1.02455392e-01,  4.09968360e-02],
[-3.30374655e-02, -1.24588567e-01, -5.27311159e-02,
 3.36388088e-01, -6.56290950e-02, -2.87419121e-01,
-9.43886652e-03,  1.27216469e-01,  5.28794965e-02,
-1.65772231e-01,  9.47146108e-02, -4.04018945e-01,
-3.66709559e-01, -6.39052236e-02, -1.59247075e-02,
-3.58045358e-02,  3.99228272e-01, -2.87192376e-01,
 4.11020428e-01, -6.24907339e-02],
[ 1.26715099e-01, -1.79690406e-01,  3.35435288e-01,
-5.55195867e-01, -1.78959093e-01, -7.61419645e-02,
 9.78640687e-02, -5.30153794e-02,  1.68011653e-01,
-5.25583490e-01,  1.93601699e-02,  2.72686413e-01,
-1.68525506e-01,  2.41153255e-01,  2.09861505e-01,
 5.86472756e-01, -3.45979362e-01, -8.16258879e-02,
 2.59470061e-01, -2.67275389e-01],
[-9.37102454e-02,  1.57654472e-01, -1.11896433e-01,
-2.87384102e-01, -7.63497086e-02, -2.02385961e-02,
-6.66185033e-02, -1.24780562e-01,  5.61083091e-02,
-9.95159881e-02,  8.14652148e-02, -2.39790880e-01,
 3.09910100e-01,  2.77985141e-01,  2.01374807e-01,
-1.43064202e-01, -1.01740346e-01,  6.70790309e-01,
 1.66244265e-01, -1.43268029e-01],
[ 5.47132868e-01,  2.64656476e-01,  2.33100493e-01,
-1.55247819e-01,  2.15263281e-01, -1.72241081e-01,
-2.99137018e-02, -1.18345770e-02, -2.25457355e-01,
 1.08663398e-01, -1.16995364e-01,  4.99641896e-02,
-6.18087505e-02,  1.87206861e-01,  2.99480865e-01,
 2.48938045e-01,  2.10225760e-02,  3.48152161e-02,
 2.69082705e-01,  4.08523919e-01],
[-7.14802854e-01, -4.93706625e-01, -1.01983980e-01,
 2.19471521e-01,  3.09519698e-01, -1.61342008e-01,
 1.79620847e-02, -1.94507413e-01, -4.92065565e-02,
 7.67534225e-03, -2.07965274e-01, -3.42163446e-01,
-3.22335303e-01, -3.14968173e-01, -1.59302353e-01,
-1.22842034e-01,  2.95446821e-01, -1.64391140e-01,
 1.25284895e-01, -1.79641887e-01]]), array([[-0.29865964],
[ 0.22372021],
[-0.32219865],
[ 0.31571709],
[ 0.33944127],
[ 0.33121193],
[ 0.24646196],
[ 0.245591  ],
[ 0.17751168],
[-0.24980963],
[-0.19066648],
[ 0.25335414],
[-0.35319777],
[ 0.35669599],
[-0.37486567],
[ 0.24186654],
[-0.40694166],
[ 0.21951073],
[ 0.28613077],
[ 0.26730649]])]
```

Accuracy Measures for Training Partition for Neural Network

```
Accuracy Measures for Training Partition for Neural Network

Regression statistics

               Mean Error (ME) : 0.0000
Root Mean Squared Error (RMSE) : 0.3932
     Mean Absolute Error (MAE) : 0.3342

Accuracy Measures for Validation Partition for Neural Network

Regression statistics

               Mean Error (ME) : 0.0065
Root Mean Squared Error (RMSE) : 0.5675
     Mean Absolute Error (MAE) : 0.4867
```

Classification for readmission Data for Validation Partition for Neural Network

Classification for readmission Data for Validation Partition

	Actual	p(0)	p(1)	Classification
4758	0	0.6108	0.3892	0
11168	1	0.2747	0.7253	1
18767	0	0.5516	0.4484	0
22415	1	0.5578	0.4422	0
20022	1	0.4883	0.5117	1
18030	1	0.5504	0.4496	0
10106	1	0.7692	0.2308	0
22179	0	0.7033	0.2967	0
22146	1	0.4046	0.5954	1
15688	1	0.5980	0.4020	0

Confusion Matrix for Neural Nets (Model 5)

```
Training Partition for Neural Network Model
Confusion Matrix (Accuracy 0.6176)

       Prediction
Actual    0    1
     0 1645  649
     1 1024 1057

Validation Partition for Neural Network Model
Confusion Matrix (Accuracy 0.6192)

       Prediction
Actual   0   1
     0 702 273
     1 441 459
```

Grid Search for Hospital Readmission:

```
Best score:0.6048
Best parameter:  {'hidden_layer_sizes': 6}
```

Confusion Matrix for Neural Nets (Model 6)

```
Confusion Matrix (Accuracy 0.6402)

       Prediction
Actual    0    1
     0 1720  574
     1 1000 1081

Validation Partition for Neural Network Model
Confusion Matrix (Accuracy 0.6048)

       Prediction
Actual   0   1
     0 685 290
     1 451 449
```

H. Interpret Results

	Logistic Regression					Neural Nets	
	1	2	3	4	5	6	7
Description	45 Predictors	45 Predictors	13 Predictors Backward Selection	22 Predictors Forward Selection	21 Predictors Stepwise	45 Predictors	45 Predictors
Train / Test Split	60 % Training 40 % Validation	80 % Training 20 % Validation	60 % Training 40 % Validation	60 % Training 40 % Validation	60 % Training 40 % Validation	70 % Training 30 % Validation	70% Training 30% Validation
Hidden Layers	— N/A —	— N/A —	— N/A —	— N/A —	— N/A —	1 hidden layer	1 hidden layer
No of Nodes	— N/A —	— N/A —	— N/A —	— N/A —	— N/A —	20 nodes	6 nodes
Confusion Matrix Accuracy - Training	61.23%	61.52%	61.36%	61.00%	61.44%	61.76 %	64.02%
Confusion Matrix Accuracy - Validation	61.57%	60.42%	60.20%	59.76%	60.22%	**61.92 %**	60.48%

- All models' respective confusion matrix results indicate no overfitting (the training and validation result is close to each other)
- Logistic Regression comparison - Model #1 to Model #5 provided almost the same results of approximately 60%. The Backward Selection, Forward Selection and Stepwise reduction method provided no significant improvement.
- Neural Nets - initial model performed better even if the second model was enhanced using the Gridsearch.

I. Conclusion

Primary Objective

The initial Logistic regression model (Model #1), with a confusion matrix accuracy (validation) of around **61.57%**, yields acceptable results using the full predictors of the data set.

The revised model (Model #2), and models using backwards elimination algorithm, forward elimination and stepwise resulted in almost similar accuracy of around 60% but still below the initial model - which provided no advantage.

The Neural Nets model (Model #6) yielded a confusion matrix accuracy (validation) of **61.92%**, a little higher compared to the best Logistic Regression model (difference of 0.35%). Though GridSearch was used to improve performance, it did not improve the accuracy making the original ***Neural Net (Model #6) performance better than all other 6 models***.

Secondary Objective

Based on the coefficient for predictors, patients that have an initial diagnosis of Diabetes have a positive change on the readmission, whereas Secondary or Tertiary diagnosis has an inverse increase.

Further checking other predictors - A1Ctest_no, change_yes, and diabetes_med_yes has a positive change on the readmission.

Overall, the project has achieved a model that can predict a 61.92% accuracy for current admitted patients with the focus on patients with initial diagnosis of diabetes (with no changes on A1C test, with changes in diabetes medication and prescribed diabetes medication). With this the hospital can improve providing better health care services by predicting which patients are at high risk of readmission and to enable healthcare providers to deliver preventative interventions and reduce overall readmission rates.

Bibliography

Data Set Link: https://www.kaggle.com/datasets/dubradave/hospital-readmissions/data