Length of Stay in hospital for Diabetic Patients

Siddharth Sharma
Mohammad Shahid Memon
Mohammed Abou Hasan
Martijn Cool

Introduction

- Irregularities of hospital visits
- Hospitalisation may depend on the severity of the illness or the impact on a patients health.
- Diabetes

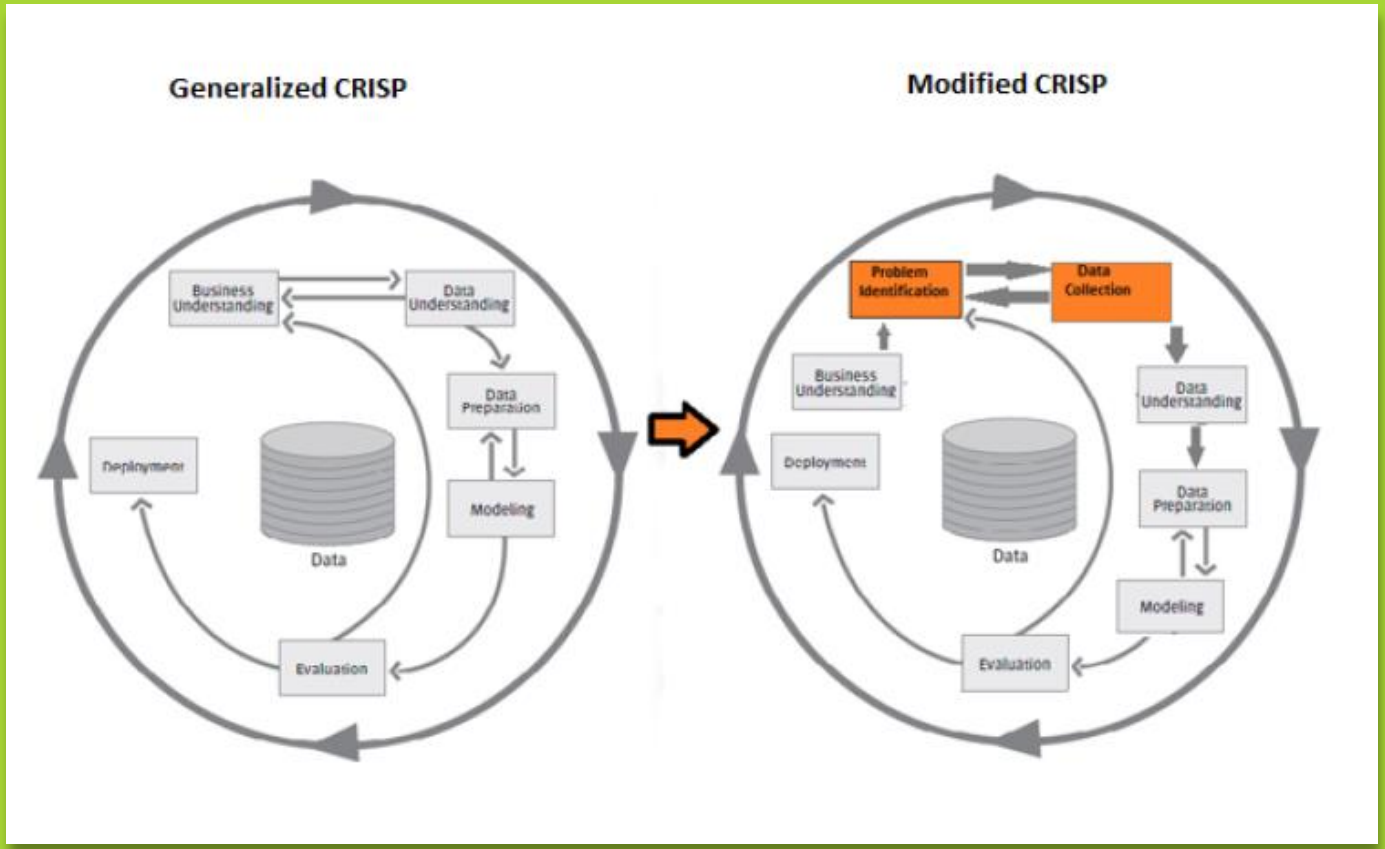
Motivation And Research Question

- No research has been performed on length of stay for diabetic patients
- Many different perspectives.
- Research question:
  - How can the duration of hospitalisation for diabetic patients be predicted?

Methodology

We are using CRISM- DM Approach.

1. Business Understanding and Problem Identification
2. Data Collection
3. Data Understanding and Preparation
4. Data Modelling
5. Evaluation
6. Deployment



Data: http://downloads.hindawi.com/journals/bmri/2014/781670.f1.zip/

Methodology - continued

- Class imbalance
- Simpler algorithms -> more sophisticated, complex algorithms
- Binary classification

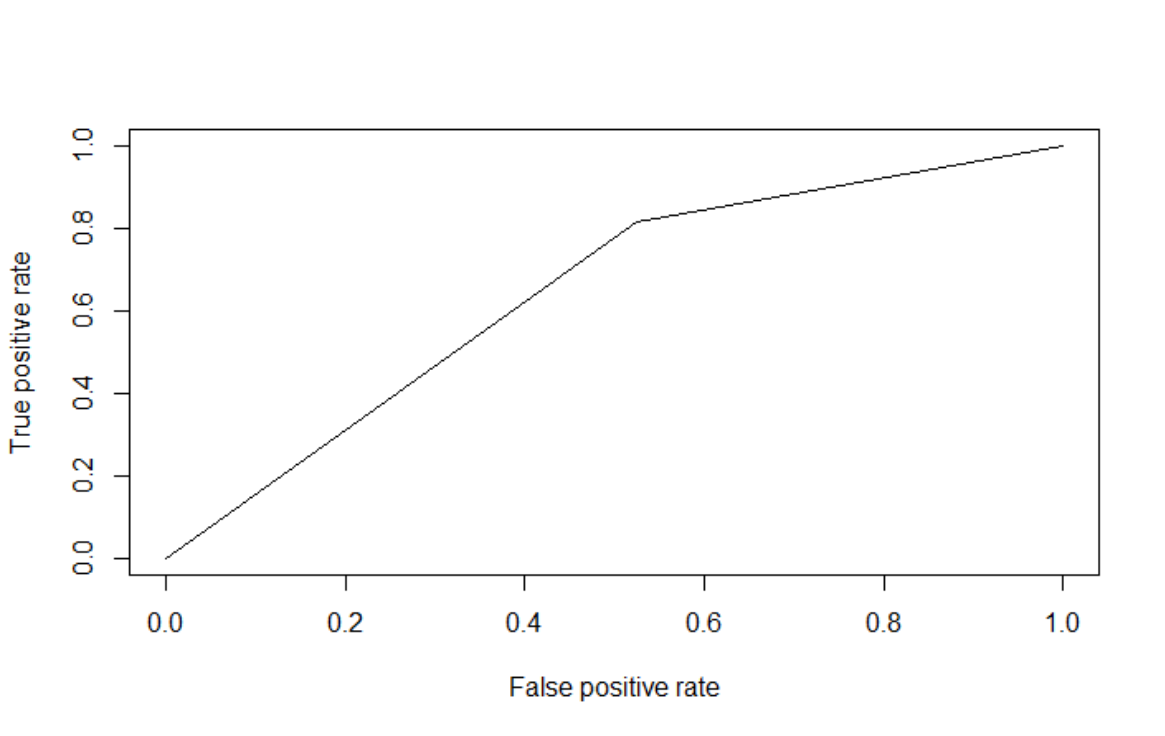
Results & Evaluation

```
> c50Tpred <- predict(c50Train, testing)
> confusionMatrix(c50Tpred, testing$time_in_hospital)
Confusion Matrix and Statistics

          Reference
Prediction long short
     long  2923  1762
     short 3223  7850

               Accuracy : 0.6837
                 95% CI : (0.6763, 0.6909)
    No Information Rate : 0.61
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.3054
 Mcnemar's Test P-Value : < 2.2e-16

            Sensitivity : 0.4756
            Specificity : 0.8167
         Pos Pred Value : 0.6239
         Neg Pred Value : 0.7089
             Prevalence : 0.3900
         Detection Rate : 0.1855
   Detection Prevalence : 0.2973
      Balanced Accuracy : 0.6461

       'Positive' Class : long

>
```

Confusion Matrix

1. *Accuracy: 68.4% - Fair result considering the fact our data is fairly balanced*
2. *Kappa Statistic: 31% - The level of agreement with our model - Fair result (>30)*
3. *Sensitivity: 47.6% of the positive cases were correctly classified by our model*
4. *Specificity: 81.7% of the negative cases were correctly classified*

Results and Evaluation




```
> auc
[1] 0.6461407
```

ROC curve and AUC

- Sensitivity - vs – Specificity graph
- The higher the graph and more inclined towards the true positive rate axis the better
- AUC is the area under ROC graph
- AUC = 65% - fair value 0 < AUC < 1 (1 is for a perfect model)
- Sensitivity is present here at the point the graph changes slope - 0.47

Conclusion

- Simple model produces same results
- Future work