Patient’s Postoperative Recovery Area Prediction

Deep Narendrabhai Mer (04)

M.Sc. Applied Statistics, Sardar Patel University, Anand

Acknowledgement

- In the accomplishment of completion of my project on Patient’s Postoperative Recovery Area Prediction, I would like to convey my special gratitude to Mr. Agniva Das Sir as well as Dr. Jyoti Divecha HOD, Department of Statistics.
- We are highly indebted to Mr. Agniva Das sir, for his guidance and continuous support in completing this project. It is because of the knowledge and skills acquired during the course work, along with his comprehensive style of teaching, that we are able understand the subject in a better way and are able to complete this modelling project successfully.

Introduction

Most deaths occurring due to a surgical intervention happen postoperatively rather than during surgery. The current standard of care in many hospitals cannot fully cope with detecting and addressing post-surgical deterioration in time. For millions of patients, this deterioration is left unnoticed, leading to increased mortality and morbidity. Postoperative deterioration detection currently relies on general scores that are not fully able to cater for the complex post-operative physiology of surgical patients. In the last decade however, advanced risk and warning scoring techniques have started to show encouraging results in terms of using the large amount of data available peri-operatively to improve postoperative deterioration detection. Relevant literature has been carefully surveyed to provide a summary of the most promising approaches as well as how they have been deployed in the perioperative domain. The integration of pre- and intra-operative data, e.g., comorbidities, vitals, lab data, and information about the procedure performed, in post-operative early warning algorithms would lead to more contextualized, personalized, and adaptive patient modelling. This, combined with careful integration in the clinical workflow, would result in improved clinical decision support and better post-surgical care outcomes.

Problem Statement

Attribute Information:

1) L-CORE (patient's internal temperature in C): high (> 37), mid (>= 36 and <= 37), low (< 36)

2) L-SURF (patient's surface temperature in C): high (> 36.5), mid (>= 36.5 and <= 35), low (< 35)

3) L-O2 (oxygen saturation in %): excellent (>= 98), good (>= 90 and < 98), fair (>= 80 and < 90), poor (< 80)

4) L-BP (last measurement of blood pressure): high (> 130/90), mid (<= 130/90 and >= 90/70), low (< 90/70)

5) SURF-STBL (stability of patient's surface temperature): stable, mod-stable, unstable

6) CORE-STBL (stability of patient's core temperature): stable, mod-stable, unstable

7) BP-STBL (stability of patient's blood pressure) stable, mod-stable, unstable

8) COMFORT (patient's perceived comfort at discharge, measured as an integer between 0 and 20)

The main objective is to build a predictive model, which could help them in predicting where patients in a postoperative recovery area should be sent to next. Because hypothermia is a significant concern after surgery.

9) decision ADM-DECS (discharge decision): I (patient sent to Intensive Care Unit),S (patient prepared to go home),

A (patient sent to general hospital floor)

ANALYSIS

- **Exploratory Data Analysis**

  After loading the dataset, we performed this method by comparing our target variable that is decision to be taken with other independent variables. This process helped us figuring out various aspects and relationships among the target and the independent variables. It gave us a better idea of which feature behaves in which manner compared to the target variable.

  - **Null values Treatment**

    Our dataset contains few null values in the form of special characters like '?' which might tend to disturb our accuracy hence we dropped them at the beginning of our project in order to get a better result.

  - **Encoding of categorical columns**

    We used One Hot Encoding to produce integers of 0, 1 and 2 to encode our categorical features because categorical features that are in string format cannot be understood by the machine and needs to be converted to numerical format.

  - **Standardization of features**

    Our main motive through this step was to scale our data into a uniform format that would allow us to utilize the data in a better way while performing fitting and applying different algorithms to it. The basic goal was to enforce a level of consistency or uniformity to certain practices or operations within the selected environment.

  - **Fitting different models**

    For modelling we tried various classification algorithms like:

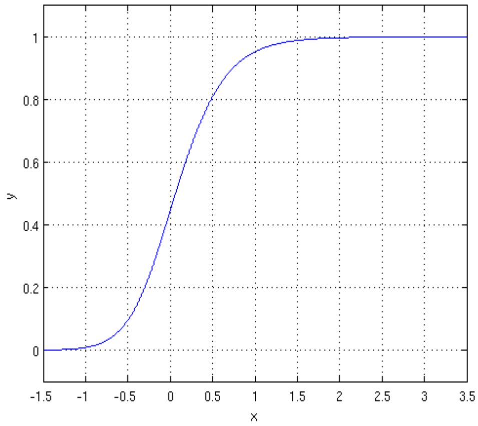
1. **Logistic Regression**
2. **Decision Tree Classifier**
3. **K Nearest Neighbour Classifier**
4. **SVM**

Algorithms

1.Logistic Regression:

Logistic Regression is actually a classification algorithm that was given the name regression due to the fact that the mathematical formulation is very similar to linear regression. The function used in Logistic Regression is sigmoid function or the logistic function given by:

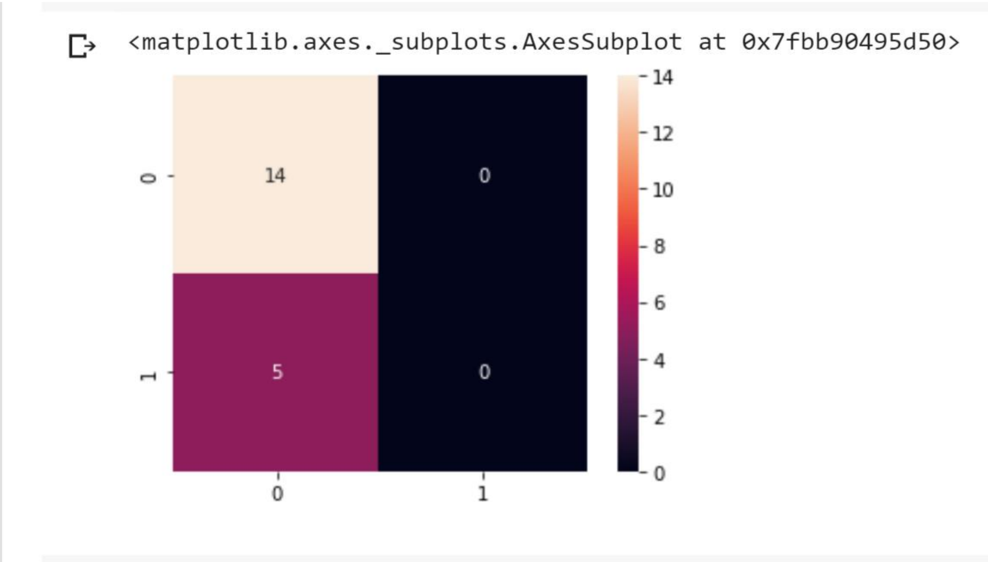
f(x)= 1/1+e ^(-x)



Model performance:

Model can be evaluated by metrics such as:

The confusion matrix is a table that summarizes how successful the classification model is at predicting examples belonging to various classes. One axis of the confusion matrix is the label that the model predicted, and the other axis is the actual label.

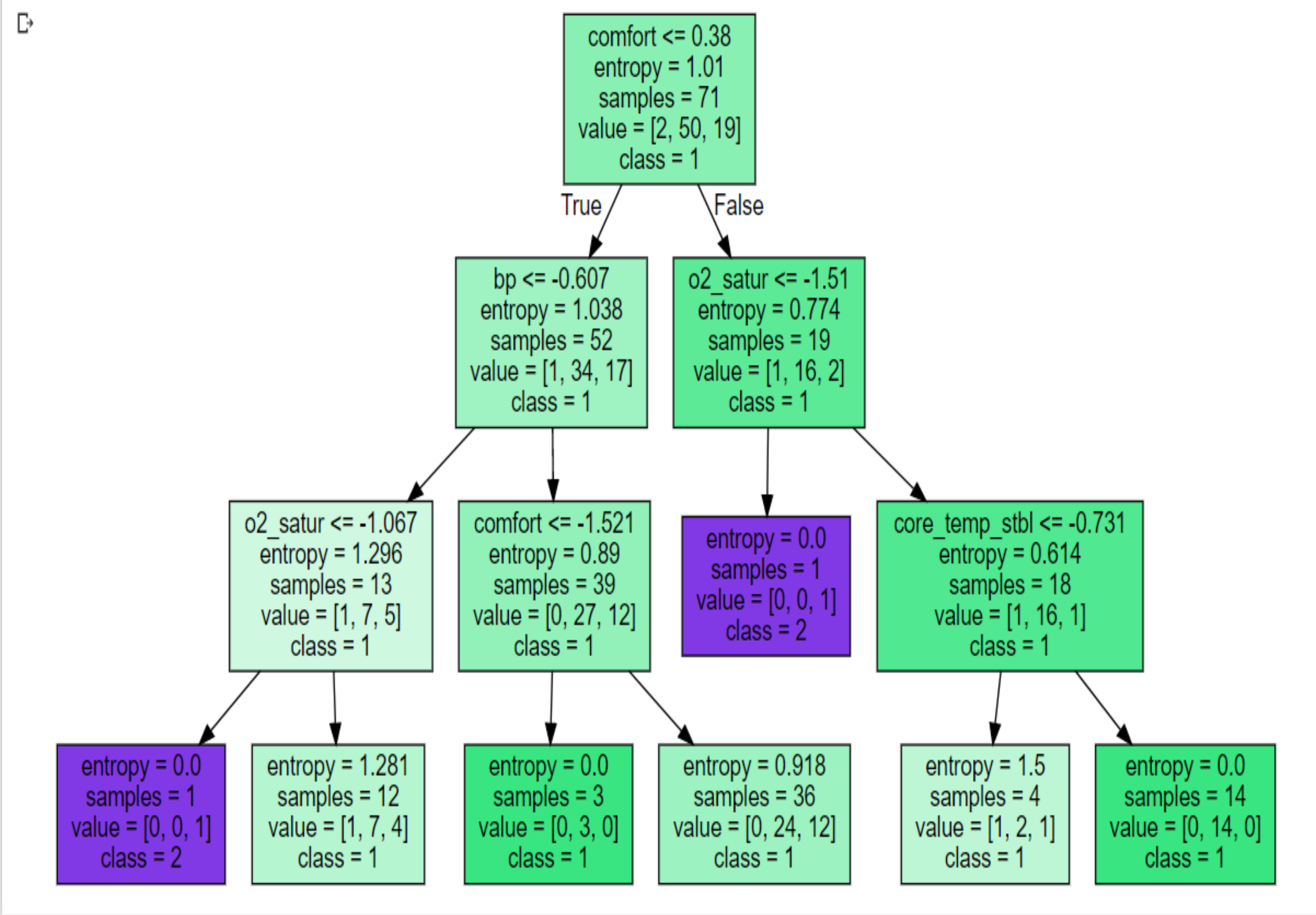


From above confusion matrix, we can conclude that among the test data, predicted 17 labels correctly as A, while 1 as A which was actually I and 5 were actual S.

The accuracy rate of the test data is **73.68**

2.Decision Tree:

Decision tree is type of supervised learning algorithm that is mostly used in classification problems. It works for both categorical and continuous input and output variables.



Model performance:

Accuracy:

Accuracy is given by the number of correctly classified examples divided by the total number of classified examples. In terms of the confusion matrix, it is given by: TP+TN/TP+TN+FP+FN

The accuracy rate of the test data is 73.68%.

3. k Nearest Neighbour

K-nearest neighbours (kNN) is a supervised machine learning algorithm that can be used to solve both classification and regression tasks. The value of a data point is determined by the data points around it. kNN classifier determines the class of a data point by majority voting principle. If k is set to 5, the classes of 5 closest points are checked. Prediction is done according to the majority class. Similarly, kNN regression takes the mean value of 5 closest points. The distance between data points is measured. There are many methods to measure the distance. Euclidean distance is one of most commonly used distance measurement.

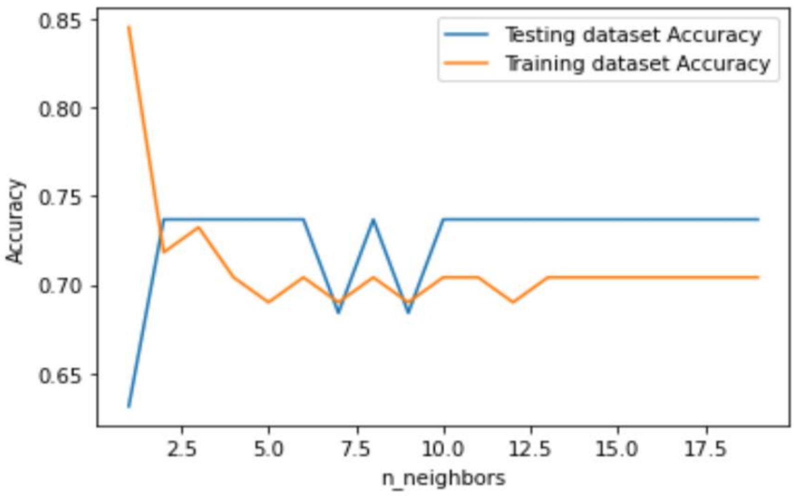
Model Performance:

Confusion matric and Classification report:

```
 [ 5  0]]
              precision    recall  f1-score   support

           1       0.74      1.00      0.85        14
           2       0.00      0.00      0.00         5

    accuracy                           0.74        19
   macro avg       0.37      0.50      0.42        19
weighted avg       0.54      0.74      0.63        19
```



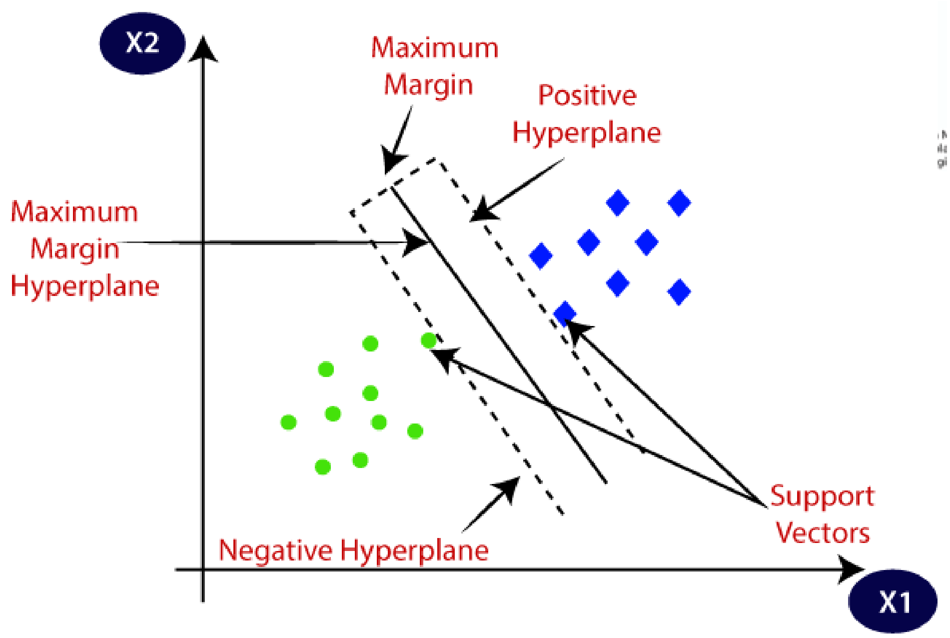
The accuracy rate of the test data is 73.68% for k = 13 that is 13 nearest neighbours.

4) SVM

Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning.

The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane.

SVM chooses the extreme points/vectors that help in creating the hyperplane. These extreme cases are called as support vectors, and hence algorithm is termed as Support Vector Machine. Consider the below diagram in which there are two different categories that are classified using a decision boundary or hyperplane:



The accuracy rate of the test data is 73.68% for SVM.

Conclusion:

The accuracy rate for all classification technique of the test data is 73.68%.

Github repository

https://github.com/TommyShelby05/ML_Assignments/blob/c8dc96b97a7a75b4599fa96a13697e18c09e1dbb/Patients'-Postoperative_Recovery_Area_Prediction_.ipynb