Hospital Length of Stay Prediction using Regression Models

Sneha Grampurohit
Electronics and Communication Engineering
K.L.E instiure of Technology
Hubli, India
snehagrampurohit@gmail.com

Sagar Sunkad
Electronics and Communication Engineering
K.L.E instiure of Technology
Hubli, India
sgsunkad842@gmail.com

***Abstract*: The increasing healthcare costs and the availability of large volume of medical data motivate the search for ways to increase the care efficiency. The Length of stay[LOS] of hospitalized patients is considered as an important factor for healthcare policy planning as it has a major impact on the technical and financial resources as well as facilities occupation. It also indicates the hospital service quality. The objective of the proposed work is to build a Decision support system that can assist the physicians to predict the inpatient hospital LOS using the regression models namely Linear, Ridge, Lasso and ElasticNet. The comparative results of these regressors is presented with Mean Absolute error as the evaluation metrics. A GUI has been built as the application of the proposed work.**

Keywords: Length of stay, Regression models, Data preprocessing, Regularization, penalty term.

I. INTRODUCTION

Length of stay is defined as the duration of hospitalization of a patient from the time of admission to the time of discharge. It significantly indicates the hospital service quality, resource utilization and inpatient hospitalization costs. The hospitals face limitations such as limited wards for patients , test labs, insufficient man power etc. These limitations lead to extended length of stay of hospitalized patients which increases the healthcare costs.

The healthcare system integrates many service units such as number of service providers, Number of diagnosis the patient encounters, number of laboratory tests available etc. Due to the advancement of digitalization, hospitals have managed to accumulate a large volume of administrative data of the patients[2]. With the help of machine learning we try to find the hidden pattern amidst the factors in the dataset and thus form an association rule[7]. The new entries of patients are then evaluated based on this association rule to predict the number of days the patent is likely to get admitted in the hospital. Regression analysis is a method that depicts the functional relationships among variables which are expressed in the form of equation or a model connecting the dependent and independent variables[3]. The regression models investigated are Linear, Ridge , Lasso and ElasticNet. Mean absolute error is used as evaluation metrics.

The proposed work considers the real world dataset from MIMIC2- Multiparameter intelligent monitoring in Intensive care. The dataset consists of 4927 patients records with 24 factors impacting the hospital LOS. Section II elaborates the overview of the proposed work. Section III elaborates the methodology used. Section IV presents the comparative results of the regressor models used and the application of GUI. Section V concludes the proposed work.

II. OVERVIEW

The brief explanation of each step is presented in this section. The General flow graph of the proposed approach is given as below

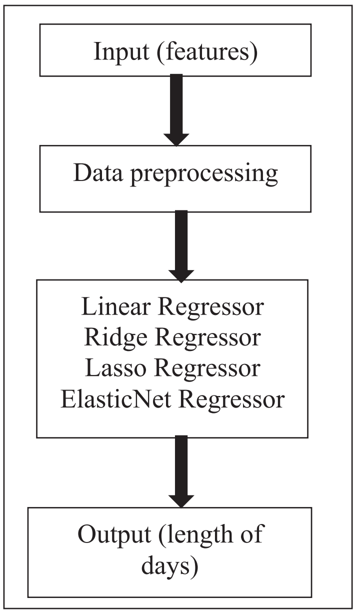

Fig. 1. Flowgraph of overview of proposed work

A. Input

The features such as inpatient age, number of diagnosis on admission, Number of caretakers available etc are given as the input to Decision support system.

The features included in the original dataset are:

TABLE I. FEATURES IN THE DATASET

Age	Number of Micro labs
Gender	IV meds
BMI	Non IV meds
Marital status	Imaging reports
Ethnicity	Notes
Admission type	Orders
Admission source	Care givers
Insurance	Care units
Religion	SOFA first score
Number of diagnoses on admission	Hospital admission ID
First diagnosis on admission	Hospital id
First procedure on admission	Number of lab tests

Authorized licensed use limited to: University of Michigan Library. Downloaded on November 21,2024 at 03:26:16 UTC from IEEE Xplore. Restrictions apply.

B. Data pre-processing

The data is converted and processed to an interpretable form using the following procedures:

- Columns such as hospital admission id do not contribute towards finding the LOS of an inpatient. Hence the column is dropped.
- We have then used one-hot -encoding method to convert categorical features such as gender, ethnicity, First diagnosis performed on admission, First procedure carried out, Admission type, admission source, marital status , insurance , religion to numerical Data.
- Normalize the data using StandardScaler class of sklearn.

C. Investigating Models

The Regressor models mentioned in the flow graph are investigated on the dataset. The hidden pattern between the features is recognized and an association rulesets are formed. Based on the ruleset formed the new entries are evaluated.

D. Output

Based on the ruleset formed the length of stay of a patient is estimated. The comparative results is produced. MAE is used as an evaluation metrics.

III. METHODOLOGY

A. Linear Regressor

Linear Regression is a machine learning algorithm that finds the linear relationship between the dependent and the independent variables. Linear regression functions with the following Assumptions

- The dependent variable should be a linear combination of independent variables.
  The LOS values should be a linear combination of the features present in the dataset.
  The hypothesis function is given by:

  $$Linear_{reg} = L_0 + Lp \tag{1}$$

  The linear equation shown assigns a single scale factor *L* to each feature p called the coefficient. L_0 is another coefficient added to give the line an additional degree of freedom.
- No autocorrelation in error terms.
  Presence of correlation in error terms penalize the model accuracy.
- Errors should have zero mean and be normally distributed.
  If error terms are not normally distributed, it implies confidence intervals will become too wide or narrow, which leads to difficulty in estimating coefficients based on minimization of least squares:
- No or little multi-collinearity.
  Multi-collinearity is the case in which the features are correlated with each other and this situation creates unstable models by inflating the magnitude of coefficients/estimates. It also becomes difficult to determine which variable is contributing to predict the response LOS.
- Error terms should be homoscedastic.
  Errors should have constant variance with respect to the features, which leads to impractically wide or narrow confidence intervals for estimates, which degrades the model's performance.

As we have more than one feature in the proposed work, Ordinary Least Squares is used to estimate the values of the coefficients. The Ordinary Least squares technique tends to minimize the sum of the squared residuals(errors). This means that given a regression line through the data we calculate the distance from each data point to the regression line, square it, and sum all of the squared errors(residuals) together. Mathematically, it is given by:

$$LinearReg_{lossfunction} = \frac{1}{n}\sum_{i=1}^{n}(y_{actual} - y_{predict})^2 \tag{2}$$

The cost function of linear function is the root mean square of the true value and the predicted value. The best line fitted to the model will have the least difference between the actual (yactual) value and the value that is predicted by the model (ypredict).

It can be noticed in the results section that Linear Regression failed to give an optimal result. While Linear regression will learn the model that best fits the training set, the training set probably isn't perfectly representative of the wider population. Therefore, it can be noticed that Linear regression has poorly performed on the test set leading to huge MAE value. Overfitting the training set is more likely to result in model parameters that are too large, Hence *Regularization technique* has been used to avoid overfitting.

B. Regularization

Regularization (also sometimes called *shrinkage*) is a technique that prevents the parameters of a model from becoming too large, and "shrinks" them towards zero[3]. The impact of regularization is that it results in models that, when making predictions on new data, have less variance. (i.e. they make less variable predictions on new data, as they are not as sensitive to the noise in the training set.). Regularization has the following functionalities:

- Regularization adds a penalty to the least squares that grows bigger with larger estimated model parameters. This process usually adds a little bias to the model, because we're intentionally underfitting the training set, but the reduction in model variance often results in a better model.
- Regularization can also help in situations that are *ill-posed[6]*. An ill-posed problem in mathematics is one that does not satisfy the three conditions of: having a solution, having a unique solution, and having a solution that depends on the initial conditions. In statistical modelling, a common ill-posed problem is when there is not one optimal parameter value, commonly encountered when the number of parameters is higher than the number of cases. In situations like this, regularization can make the estimating the parameters a more stable problem.

The principal job of regularization is to prevent algorithms from learning models that are overfit by discouraging complexity[6]. This is achieved by penalizing model parameters that are large, shrinking them towards zero.

Authorized licensed use limited to: University of Michigan Library. Downloaded on November 21,2024 at 03:26:16 UTC from IEEE Xplore. Restrictions apply.

There are two common penalties used:

- L1 norm
- L2 norm

Three particularly well-known and commonly used regularization techniques for linear models are:

- Ridge regression
- Least absolute shrinkage and selection operator (LASSO)
- Elastic net

These three techniques can be thought of as extensions to linear models that reduce overfitting. Because they shrink model parameters towards zero, they can also automatically perform feature selection by forcing predictors with little information to have no or negligible impact on predictions.

C. Ridge Regression

Ridge regression modifies the least squares loss function slightly, to include a term that makes the functions value larger, the larger the parameter estimates are. As a result, the algorithm now has to balance selecting the model parameters that minimize the sum of squares, and selecting parameters than minimize this new penalty. In ridge regression, this penalty is called the *L2 norm* , and Ridge regularization therefore penalizes models that are too complex (because they have too many predictors) [6].

The loss function of Ridge Regressor is given by:

$$Ridge_{lossfunction} = \sum_{i=1}^{n}(y_{act} - y_{pred}) + \lambda\sum_{q=1}^{r}\beta_q^2 \quad (3)$$

The calculation simply involves

- squaring of all LOS feature slope values and add them up (all except the intercept).
- Multiply the L2 norm by a value called *lambda* (*λ*). Lambda can be any value from zero to infinity and acts as a volume knob: large values of lambda strongly penalize model complexity, while small values of lambda weakly penalize model complexity. Lambda cannot be estimated from the data, and so is a hyperparameter than we need to tune to achieve the best performance by cross-validation. The Lambda value used in our proposed work is 0.1.
- Once we calculate the L2 norm and multiply it by lambda, we then add this product to the sum of squares to get our penalized least squares loss function.

The graphical explanation (considering only two factors) of L2 penalized function is presented below:

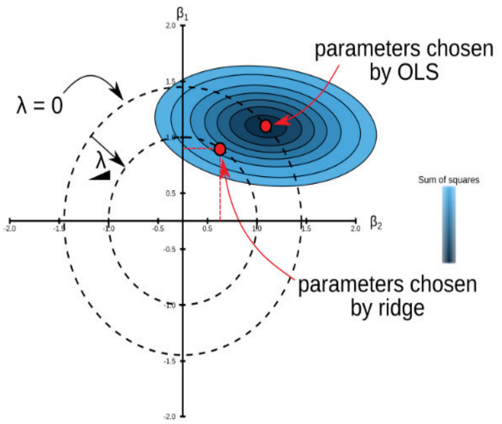

Fig. 2. Graphical representation of L2 penalized loss function [Source:6]

The x and y axes show values for two slope parameters, (β_1) and β_2). The shaded contour lines represent different sum of squares values for different combinations of the two parameters, where the combination resulting in the smallest sum of squares is at the centre of the contours. The dashed circles centred at zero represent the L2 norm multiplied by different values of lambda, for the combinations of β_1 and β_2 the dashed lines pass through. When lambda = 0, the circle passes through the combination of β_1 and β_2 that minimizes the sum of squares. When lambda is increased, the circle shrinks symmetrically towards zero. Now, the combination of parameters that minimizes the penalized loss function is the combination with the smallest sum of squares **that lies on the circle.**

D. Lasso Regressor

Lasso Regressor uses L1 norm to penalize. The L1 norm is only slightly different to the L2 norm. Instead of squaring the parameter values, we take their absolute value instead, and *then* sum them. The cost function and MAE of Lasso Regressor is given by

$$Lisso_{lossfunction} = \sum_{i=1}^{n}(y_{act} - y_{pred}) + \lambda\sum_{q=1}^{r}|\beta_q| \quad (4)$$

Ridge regression can shrink parameter estimates towards zero, but they will never actually be zero. LASSO is able to shrink small parameter values to zero, effectively removing that predictor from the model. The graphical representation of L1 penalized function is given by:

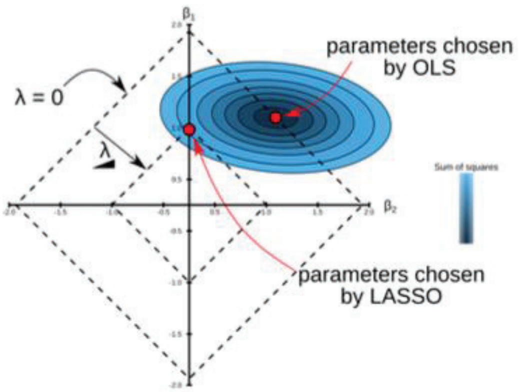

Fig. 3. Graphical representation of L2 penalized loss function [Source:6]

Authorized licensed use limited to: University of Michigan Library. Downloaded on November 21,2024 at 03:26:16 UTC from IEEE Xplore. Restrictions apply.

The solid, concentric circles represent the sum of squares value for different combinations of the parameters. The dashed diamonds represent the L2 norm multiplied by lambda. The combination of parameters with the smallest sum of squares that touches the diamond is one where parameter β_2 is zero. This means that the predictor represented by this parameter has been removed from the model.

E. ElasticNet Regressor

ElasticNet is an extension of linear modelling that includes both L2 *and* L1 regularization in its loss function. It finds a combination of parameter estimates somewhere between those found by ridge regression and by LASSO. We're also able to control just how much importance we place on the L2 vs the L1 norms using the hyperparameter, alpha. The loss function and MAE of Elastic Net is given by

$$ElasticNet_{lossfunction} = \sum_{i=1}^{n}\left(y_{act} - y_{pred}\right) + \left((1-\alpha)\sum_{q=1}^{r}\beta_q^2 + \right. \quad (5)$$

when alpha is zero, the L1 norm becomes zero, and we get ridge regression when alpha is one, the L2 norm becomes zero, and we get LASSO when alpha is between zero and one, we get a mixture of ridge regression and LASSO. We tune it as a hyperparameter and let cross-validation choose the best-performing value for us.

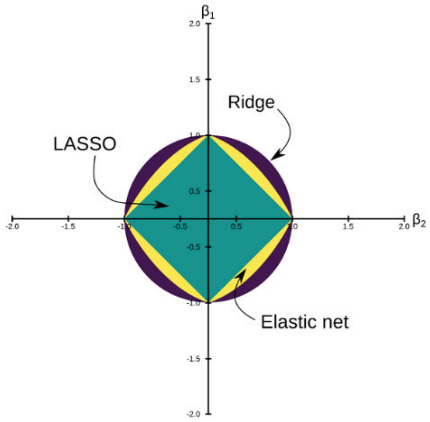

Fig. 4. Graphical representation of ElasticNet in comparison with ridge and Lasso regression [Source:6]

Figure 3 compares the shapes of the ridge, LASSO and elastic net penalties. As the elastic net penalty is somewhere between the ridge and LASSO penalties, it looks like a square with rounded sides. ElasticNet can shrink parameter estimates to zero, allowing it to perform feature selection like LASSO. But it also circumvents LASSO's limitation of not being able to select more variables than there are cases. Another limitation of LASSO is that if there is a group of predictors that are correlated with each other, LASSO will only select one of the predictors. Elastic net on the other hand is able to retain the group of predictors.

IV. RESULTS

The evaluation metrics used in the proposed work is the Mean Absolute error.

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|y_{iact} - y_{ipred}\right| \quad (6)$$

Mean absolute value measures the average magnitude of the errors in a set of predictions made by the model without taking into account their direction. It's simply the average over the test samples of the absolute difference between the actual value and the predicted value where all the features have equal weight.

The MAE values for each regressor investigated on the dataset is given by:

TABLE II. MAE RESULTS

Regressor	MAE value
Linear Regressor	198379877732011.9
Ridge Regressor	0.82131
Lasso Regressor	0.96865
ElasticNet Regressor	0.95121

As seen above, Linear Regression has a high MAE value which indicates that it has large difference between the actual and the predicted values when tested over the sample test cases. Hence, it can be inferred that linear Regressor did not give the satisfactory results. Furthermore, it can be inferred that Regularization methods i.e. Lasso, Ridge and ElasticNet has considerable performed well on the considered database. ElasticNet has given the most optimal results.

A sample graphical user interface has been created using tkinter in python. In the GUI we have considered 5 features namely Number of diagnosis performed on the patients at the time of admission, The imaging reports, Number of care takers available, Age of the patient and gender.

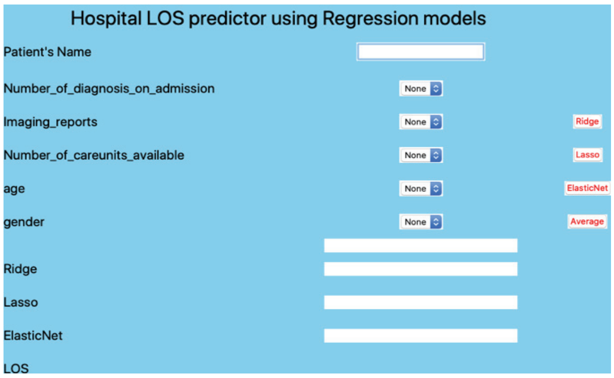

Fig. 5. An Empty GUI

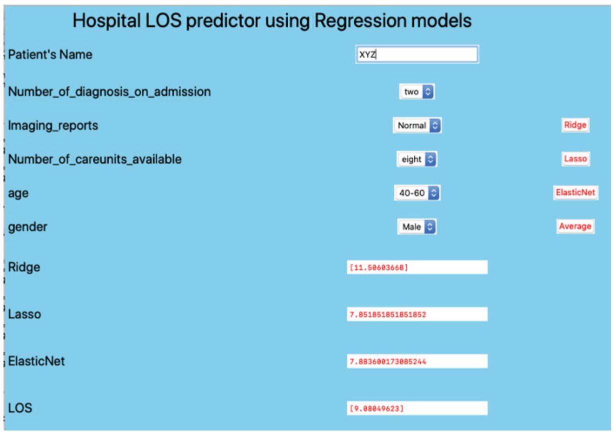

Fig. 6. SAMPLE GUI

Authorized licensed use limited to: University of Michigan Library. Downloaded on November 21,2024 at 03:26:16 UTC from IEEE Xplore. Restrictions apply.

The loss functions of the regressors for the above case is given as

TABLE III. LOSS FUNCTION COMPARISON RESULTS

Regressor	Loss function value
Ridge Regressor	0.66550
Lasso Regressor	0.9125
ElasticNet Regressor	0.9040

The prediction of **Length of stay** by different algorithms based on the inputs given by the patients are:

TABLE IV. LOS COMPARISON RESULTS

Regressor	Length of stay
Ridge Regressor	11
Lasso Regressor	8
ElasticNet Regressor	8

That is, it can be depicted that a patient with name XYZ who has encountered **2 diagnosis** at the time of his admission and whose **imaging reports are found to be normal**, with **8 care takers** currently available for him[**Male**] and his age is in between **40 to 60** years is more likely to be for **9 days** in the hospital.

V. CONCLUSION

The proposed work has made the comparative analysis of Linear, Ridge, Lasso and ElasticNet regressor. Linear regression did not give the satisfactory results due to the overfitting over the training dataset it couldn't give the optimal results on the test set. Hence the model was extended using the regularization techniques i.e. Lasso, Ridge and ElasticNet which considerably performed well on the dataset.

The proposed work provides assistance to the physicians. The physicians can know what features lead to long length of stay of the patients in the hospital and can concentrate on these features and improvise it. This in-turn improves the hospital quality and saves healthcare costs. The future work can be extended on collecting more factors leading to longer stay of patients in the hospital.

REFERENCE

[1] Livieris, Dimopoulos, et.al. "Predicting length of stay in hospitalized patients using SSL algorithms", *2018 Data science and Ai Association (DSAI),* 2018.

[2] Peng Liu, Lei Lei, et.al. "Healthcare Data Mining: Prediction of Inpatient Length of Stay", *2006 3rd International Conference on Intelligent Systems.* IEEE, 2006.

[3] Catherine Combes, Farid Kadri, Sondès Chaabane. "Predicting Hospital Length of Stay Using Regression Models: Application To Emergency Department". HAL,2014.

[4] Tanuja S, Dr. U. Dinesh Acharya, et.al. "Comparison of different data mining techniques to predict hospital length of stay" , *2017 Journal of Pharmaceutical and biomedical Sciences.*2017.

[5] Lior Turgeman , Jerrold H. May, et.al. "Insights from a machine learning model for predicting the hospital Length of Stay (LOS) at the time of admission", Elsevier.2017.

[6] Hefin Rhys " Machine Learning with R,tidyverse, and mlr", in *Manning Publications,* 2019,ch. 12,pp. 262-295.

[7] B. Thompson, K. O. Elish and R. Steele, "Machine Learning-Based Prediction of Prolonged Length of Stay in Newborns," *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, Orlando, FL, 2018, pp. 1454-1459, doi: 10.1109/ICMLA.2018.00236.

[8] E. Carter and H. Potts, "Predicting length of stay from an electronic patient record system: a primary total knee replacement example", *BMC Medical Informatics and Decision Making*, vol. 14, no. 1, 2014.

[9] M.-T. Chuang, Y.-h. Hu and C.-L. Lo, "Predicting the Prolonged Length of Stay Of General Surgery Patients: a Supervised Learning Approach", *International Transactions in Operational Research*, pp. 75-90, 2016.

Authorized licensed use limited to: University of Michigan Library. Downloaded on November 21,2024 at 03:26:16 UTC from IEEE Xplore. Restrictions apply.