ISSN: 2230-9926

Available online at http://www.journalijdr.com

IJDR

International Journal of Development Research
Vol. 11, Issue, 02, pp. 44785-44788, February, 2021
https://doi.org/10.37118/ijdr.21207.02.2021

RESEARCH ARTICLE OPEN ACCESS

MACHINE LEARNING FOR INCONSISTENCY DETECTION IN HOSPITAL CESAREAN BILLS

[1]Ana Carolina Dominicci de Souza, [2]Alexandre Gonçalves Sanches, [3]Gislaine Cristina Moisés, [4]Verônica Amparo QuispeChire, [5]José Luís Esteves Francisco, [6]Júlio César André and [7*]João Marcelo Rondina

[1]Undergraduate Student in Medicine - Medical School of Medicine of São José do Rio Preto - FAMERP, Brazil
[2]Health Care Corporation of São José do Rio Preto – SP – Brazil – Bensaúde
[3]Health Care Corporation of São José do Rio Preto – SP – Brazil – Bensaúde
[4]National Learning Service - Senac of São José do Rio Preto,
[5]Medical School of Medicine of São José do Rio Preto - FAMERP, Brazil,
[6]Center for the Study and Development of Health Education - CEDES, Medical School of Medicine of São José do Rio Preto - FAMERP, Brazil

ARTICLE INFO

Article History:
Received 20th December, 2020
Received in revised form
26th December, 2020
Accepted 11th January, 2021
Published online 28th February, 2021

Key Words:
Artificial Intelligence,
Machine Learning,
Auditorship.

****Corresponding author:***
João Marcelo Rondina

ABSTRACT

Aims: Evaluate the most adequate Machine Learning models to the analysis of inconsistencies and irregularities on the final values on the bills presented to the health care plan operator. **Methods:** 1,602 medical bills' receipts regarding caesarean hospitalizations, 50.20% of the receipts being inconsistent and 49.80% of the receipts not presenting inconsistencies. The selected documents are from the period between 2015 a 2019. Nine important variables on the charge receipts auditor ship were selected, logistic regression algorithm and K-Nearest Neighbors (KNN) algorithm were the classification ones and the observation set was divided in data to the training and the test, in order to verify if the model presented good predictive performance on both steps. Root Mean Squared Error (RMSE), confusion matrix, accuracy, sensibility and specificity were calculated and the Receiver Operating Characteristic curve was designed. **Results:** A 666.82 RMSE on the test phase, which is considered a expressive value, informing that the linear regression model didn't get a good predictive performance on the study. KNN algorithm with a 91.20% accuracy level and 91.52% accuracy on logistic regression, on a 0.63 threshold (cutoff), showing a good prediction performance to both models and a small significant difference between them. **Conclusions:** The results found on this study show that only the KNN models and logistic regression present themselves as a satisfactory tool on the classification of inconsistent receipts. However, the logistic regression model was better because the KNN model needs a superior computational capacity and, when it is applied in a real scenario with a bigger quantity of data, the processing time would be slow. In the future, adopting the classification models, the medical bills auditor's focus could be directed to the bills classified as inconsistent, dismissing the necessity of all the bills received, making the auditorship process assertive and agile.

Citation: ***Ana Carolina Dominicci de Souza, Alexandre Gonçalves Sanches, Gislaine Cristina Moisés, Verônica Amparo QuispeChire et al. 2021.*** **"**Machine Learning for inconsistency detection in hospital cesarean bills*", International Journal of Development Research*, 11, (02), 44785-44788.

INTRODUCTION

The auditorship is a tool that evaluates the management quality system's level, in order to avoid waste and irregularities on the invoices of services provided by entities. This tool has become very present on health institutions. The auditorship on the health system, according to Lima and Erdmann (2006), is a periodic evaluation method of institutional resources of each hospital, as a guarantee of quality on the assistance provided to the client.

The supplementary health operators generate a large quantity of data on a daily basis, which motivate the search for evaluation methods that can make possible the identification of irregularities and inconsistencies on bills in order to put them through auditorship, as a way to verify the characteristics of the medical solicitations to medicines and materials. In this context, the technology tools are important because they turn the auditorship process more efficient as analyse large quantities of information. Among this tools, we highlight the Artificial Intelligence (AI), which are used to improve and accelerate the decision making processes. A few examples of their use can be found on the work of Borges *et al.* (2020), where AI implications were identified on intern auditorship processes and on the work of Paulo *et al.* (2018), studying the use of Machine Learning (ML) on public bills' control practices. Machine Learning, an AI branch, is a learning method based on the idea that algorithms can embrace existing data structure and generate prediction rules to future decision making (LEI *et al.*, 2020; WARING, LINDVALL, UMETON, 2020). Between the leaning options, two techniques are highlighted: not supervised apprenticeship, when the algorithm is trained only with predictor variables, or supervised apprenticeship, when the algorithm is trained with predictor variables and the answer variable of interest known (ZHAIL *et al.*, 2020; SALLOUM *et al.*, 2020). Therefore, the supervised technique is the most adequate to be used in auditorships and makes possible the evaluation of bills that show inconsistencies and irregularities (PAULO *et al.* 2018). This study had as goal the evaluation of the most suitable ML methods to the inconsistencies and irregularities analysis on the final values of bills presented to the health care plan operator.

MATERIALS AND METHODS

The present work was an accuracy test that compared traditional auditorship methods with the existing Machine Learning ones. Initially it was determined which data would be used to do the test. The data were selected by the regularity and the completeness and the absence of failures of information, which would make necessary the additional research and/or polishing of these. We selected 1,602 medical bills' receipts regarding caesarean hospitalizations, from which 805 knowingly had some kind of inconsistency, because they have been through previous auditorship, showing that 50.20% presented inconsistencies and 49.80% didn't. In order to access this information, a data base from a health care plan from São Paulo state was used. The receipts were from a 5 year gap, a period between 2015 and 2019. For privacy precautions, the data base on this study remains anonymous, using the suppression technique (KOHLMAYER, PRASSER, KUHN, 2015), according to the Health Insurance Portability and Accountability Act (HIPAA). In the data base used, nine of the most relevant variables were selected on the auditorship of the charging bills, presented on Table 1.

Table 1. Selected variables and their descriptions

Variable	*Variable Description*
NHD	Number of hospitalization days
Proc1	Another procedure authorized alongside the caesarean
Proc2	Another procedure authorized alongside the caesarean
CPA	Caesarean's patient age
KA	Kind of accommodation (shared with two or three beds or private)
NBT	Number of births in time
NPB	Number of premature births
HBTV	Hospital bill total value
Separate	Categorical variable, which the value 1 indicated the existence of inconsistency and the value 0 indicated the absence of inconsistency.

Consequently, it was determined, according to the variables, which algorithms would be tested. The supervised model is divided in: classification algorithm (to answer variable qualitative, nominal or ordinal) and regression algorithm (to answer variable quantitative, discreet or continuous). Therefore, on the algorithm classification set, the following were chosen: logistic regression algorithm (LATTIN, CARROL, GREEN, 2015), for being one of the more traditional techniques and its linearity and high interpretation capacity of parameters and stability through time, as described by Olivera *et al.* (2018) and the K-Nearest Neighbors (KNN) algorithm, because it is a simple algorithm, with a guarantee of theoretical performance, according to Buza *et al.* (2015) e Raschka & Mirjalili (2018). Among the regression algorithms set, the linear regression algorithm was chosen (MORETTIN & BUSSAB, 2017), for originating a lot contemporary tools (SANTOS, 2018). On the model's adjust step, the set of observations was divided in data to the training and to the test, in order to verify if the model presented good predictive performance on both phases. The most used divisions are 60 % to 80% to the training and to the test, 20% to 40%. It's important to say that the proportion of the data to the training must be bigger than to the test (SANTOS *et al.*, 2019). Some algorithms are more complex and other less, the main problem with the ones with bigger complexity is the overfitting, which means, they model well the data on the steps of training and test, but don't present a good predictor performance to future data and this occur because the model memorize the data instead of understanding the pattern.

On the Machine Learning supervised algorithms, the n observations of the set are represent by xi, j, i= 1, 2,…, n e j =1, 2,…, p, called explanatory variables and yi is the answer variable of interest, The most important characteristics of the chosen models are described below:

Linear Regression Algorithm: The linear regression algorithm is used to regression problems, which objective is to predict continuous answers. One of the purposes of the model is the interpretation of the predictor variable's interpretation xi and the answer variable yi. This relation is represented by the following equation:$y_i = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \cdots + \beta_p x_{i,p}$, where β_0, β_j, j = 1, 2, 3, …, p, are unknown parameters. In this scenario, the objective is to estimate the equation parameters that better describe the relation between the explanatory variables and the answer variable, for the purpose of using on the prediction of the answer in a new data set. From the training data, the least squares method will be used to estimate the parameters values which minimize the Residual Sum of Squares (RSS), given by the following equation: RSS = $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$. To each parameter value we'll have a result to the sum of squares and the solution is choosing the values that will make this sum the least possible (MORETTIN & BUSSAB, 2017).

Logistic Regression Algorithm: The logistic regression is a model used to problem classification, aiming to predict the probability of each xi observation belonging to one of the answer categories. To this model, the answer variable is binary, which means, yi takes on two values, 0 or 1, representing "failure" or "success", respectively, highlighting that "success" is the event of interest. That way, the simple logistic regression is expressed by: $P(y = y_i/x_i) = \frac{e^{\beta_0 + \sum_{j=1}^{p} \beta_j x_{i.j}}}{1 - e^{\beta_0 + \sum_{j=1}^{p} \beta_j x_{i.j}}}$, where β_0, β^j, j = 1, 2, 3, …, p, are unknown parameters. One of the estimation methods frequently used to this model is the maximum likelihood method, which expression. Is given by: $L(\beta) = \prod_{i=1}^{n} P(Y = y_i/x)^{y_i}\left(1 - P(Y = y_i / x)\right)^{1-y_i}$. So, in order to obtain the parameters estimators, the L(β) function must be used (LATTIN, CARROL, GREEN, 2015). In this model, the important is choosing the cutoff to the estimate probabilities $P(Y=y_i /x_i)$ above 0.5 belonging to the category with the event of. This is also known as threshold.

K-Nearest Neighbor Algorithm (KNN): The KNN algorithm is one of the most known. The goal is to estimate the category of a new sample and to look for the K-nearest neighbors to it, in a set of known data, according to a defined distance measure. On the training phase of the closest K-neighbors model there are two important choices: the k number of nearest neighbors to the new sample and the distance measure responsible for choosing the k neighbors (RASCHKA & MIRJALILI, 2018). The k value is defined as a quality representation of nearest neighbors which will be used to check which category the new sample belongs to.

The model uses the proximity between the new sample (x_j^*, j = 1, 2,…, p) and the explanatory variables pf the training set to define a neighborhood. In this context, a very used model is the Eucledian distance, which measures the proximity and it's written like: $d(x^*, x_i) = \sqrt{\sum_j^p (x_j^* - x_i)^2}$. On the test phase, after defining the number of neighbors, the distance measure to be used is calculated. The most recurring category observed on the x_j^* neighborhood will represent the new sample (RASCHKA & MIRJALILI, 2018). During the selected models training and test phases, the following metrics (BATISTA & FILHO, 2019), summarized on Table 2 were used to evaluate the performance of the supervised learning predictor models.

Table 2. Description of metrics used to evaluate the algorithms' performance

Algorithm	*Metrics used to evaluate the performance*	*Description*
Regression	Root Mean Squared Error - RMSE.	The Root Mean Squared Error (RMSE) is given by: $RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}}$. The RMSE value will be small, if the predictor values by the model are close to the values observed. This indicates a good algorithm performance (VANDEPUT & FORECAST, 2020).
Classification	The confusion matrix, illustrated on Table 3, describes a crossed tab between the real and the predicted category, where the main diagonal represents the correct classifications (TN e TP) and the other diagonal indicates the classification errors. (FN e FP). Based on the matrix confusion, the accuracy, sensibility and specificity are calculated from the threshold* chosen.	Accuracy $\frac{TP+TN}{(TN+FP+FN+TP)}$ is the success proportion. Sensibility $\frac{TP}{(FN+TP)}$, is the proportion of true positives correctly identified. Specificity $\frac{TN}{(TN+FP)}$, is the proportion of true negatives correctly identified.
	ROC Curve (Receiver Operating Characteristic).	It shows how the sensibility varies with the specificity to different thresholds*, so that the algorithm performance is evaluated by the area below the curve area under the ROC curve (AUC). The bigger the AUC, better the algorithm performance.

TN - true negative; TP - true positive; FN - false negative; FP - false positive
*threshold: the cut separating categories

Table 3. Confusion Matrix to binary classification

	Predicted category	
True class	Absent	Present
Absent	True Negative (TN)	False Positive (FP)
Present	False Negative (FN)	True Positive (TP)

All the steps to models prediction were built on the Python language, version 3.6, using the Anaconda 2020.02 ambient, with Jupyter Notebook - Python IDE, alongside scikit-learn libraries, version 0.22.1, pandas, seaborn and Matplotlib version 3.2.1

RESULTS AND DISCUSSION

On Table 4 are presented the results of the training and test phases of the selected predictor models.

Table 4. Algorithms training and test

To the linear regression model, only 797 receipts with no inconsistencies were used, from the 1,602, aiming to predict the correct values. Using separated receipts to the test phase, the RMSE value of 666.82 was calculated, which is this algorithm's metric, presented on Table 2.

Table 4. Algorithms training and test

Algorithms	*Linear regression*	*KNN*	*Logistic regression*
Training			
Sample size	597 75% in 797 No inconsistency	1,202 75% in 1,602	1,202 75% in 1,602
Predicted category	NHD	NHD	NHD
	Proc1	Proc1	Proc1
	Proc2	Proc2	Proc2
	CPA	CPA	CPA
	KA	KA	KA
	NBT	NBT	NBT
	NPB	NPB	NPB
		HBTV	HBTV
Response variable	HBTV	Separate	Separate
Test			
RMSE	666.82		
Threshold			0,63
Accuracy		91.2%	91,52%
K		9	
TP (Sensibility)			216 (0.98)
FP			31
TN (Specificity)			151 (0,82)
FN			3

This value is considered large, informing that the linear regression model didn't have a good prediction performances to the study. In hopes of predicting the classification of future inconsistencies, the logistic regression models and the KNN were used. To both models the 1,602 receipts were used. The performance of the models was obtained on the test phase, calculating the accuracy, which is the metric utilized to these algorithms, presented on Table 2. To the KNN, the accuracy level was 91.20% with k = 9 (number of closest neighbors) and, to the logistic regression, the accuracy level was 91.52%, to a 0.63 threshold (cutoff). Regarding the accuracy level, it was observed a good prediction performance to both models and a slightly significant difference between them. However, it was took under consideration the KNN model computational cost compared to the logistic model, which presents a bigger processing time (BATISTA & FILHO, 2019). Aiming to show the assertiveness and the performance of the logistic regression model, it was calculated the sensibility (0,98), where the TP (True Positive) value was 216 receipts and the FP (False Positive) value was 31 receipts, meaning that only a small part of documents was analyzed unnecessarily and the specificity (0,82) was calculated, where the TN (True Negative) value was 151 receipts and the FN (False Negative) was 3 receipts, meaning that only a small part of receipts were paid when there was inconsistency.

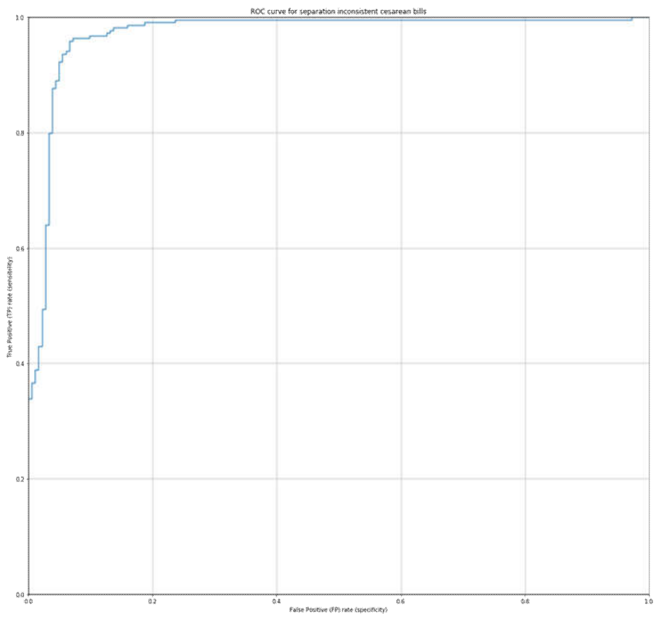

Figura 1. The logistic regression model's ROC Curve

To validate the performance of the model, the metrics value was also calculated (Table 2) AUC ROC (0.971), according to Figure 1, which proves the good performance. Therefore, the KNN models and logistic regression present the best performance on the receipts with inconsistencies classification, according to the metrics values calculated. Although, the logistic regression model has a better performance compared to the KNN model due to the smaller processing time, given that processing speed is necessary on daily work. (BATISTA & FILHO, 2019).

CONCLUSION

The supplementary health care operators generate a large number of data, motivating the search of agile methods on the identification of inconsistencies on bills that go through auditorship. In this context, three supervised models were analyzed to 1.602 medical bills' receipts. Even if the sample is small, it represents data that will be further observed because they will follow the same pattern from where the sample was taken off. The results found in this study allow the observation that only the KNN models and logistic regression present themselves as satisfactory tools on the classification of receipts with inconsistencies. Despite that, the logistic regression model displayed itself as better, because the KNN model needs a bigger computational capacity and, when applied to a real scenario with a larger number of data, the processing time would be slow. In time to come, when adopting classification models, the medical bills' auditor focus can be dragged to the bills classified as inconsistent, dismissing the necessity the total analysis of received bills, making the auditorship agile and assertive.

REFERENCES

Batista, A. F. M., Chiavegatto Filho, A. D. P. 2019 Machine Learning aplicado à Saúde. Available online at https://sol.sbc.org.br/livros/index.php/sbc/catalog/download/29/95/245-1?inline=1

Borges, W. G, Leroy, R. S. D., Carvalho, L. F., & Oliveira, J. M. 2020 Implications of Artificial Intelligence in Internal Auditing in Brazil: Analysis under Professionals Perception. Rev UFRJ. 151. Available online at https://revistas.ufrj.br/index.php/scg/article/view/25284/pdf

Buza, K., Nanopoulos, A., & Nagy, G. 2015 Nearest neighbor regression in the presence of bad hubs. Knowledge-Based Syst. 86, pp. 250-260. https://doi.org/10.1016/j.knosys.2015.06.010

Carmo, P. N. S., Souza, B. F., Reis, M. V. G., & Vieira, J. C. 2018 Aprendizado de Máquina em Ações de Controle no Tribunal de Contas do Estado do Maranhão. Available online at http://sistemas.deinf.ufma.br/anaisjim/artigos/2018/201809.pdf

Kohlmayer, F., Prasser, F., & Kuhn, K. A. 2015The cost of quality: Implementing generalization and suppression for anonymizing biomedical data with minimal information loss. J Biomed Inform. 58, pp. 37-48. http://dx.doi.org/10.1016/j.jbi.2015.09.007.

Lattin, J., Carrol, J. D., & Green, P. 2011 Análise de dados multivariados. Cengage Learning, São Paulo, Brasil.

Lei, Y., Yang, B., Jiang, X., Jia, F., Li, N., &Nandi , A. K. 2020 Applications of machine learning to machine fault diagnosis: A review and roadmap. Mechanical Syst Signal Process. 138, pp. 106587. https://doi.org/10.1016/j.ymssp.2019.106587

Lima, S. B. S., & Erdmann, A. L. 2006 Nursing role during accreditation process of an emergency service. Acta Paul Enferm. 193, pp. 271-278. https://doi.org/10.1590/S0103-21002006000300003

Morettin, P. A., & Bussab, W. O. 2017 EstatísticaBásica. Saraiva, São Paulo, Brasil.

Olivera, A. R. 2016 Comparison of machine learning algorithms to build predictive models of undiagnosed diabetes. Ph. D. Thesis in Computer Sciences. Universidade Federal do Rio Grande do Sul Instituto de Informática, Porto Alegrte RS Brasil.

Olivera, A. R., Roesler, V., Iochpe, C., Schmidt, M. I., Vigo, A., Barreto, S. M., Duncan, B. B. 2017 Comparison of machine-learning algorithms to build a predictive model for detecting undiagnosed diabetes - ELSA-Brasil: accuracy study. Sao Paulo Med J. 1353, pp. 234-246. doi: 10.1590/1516-3180.2016.0309010217

Raschka, S., & Mirjalili, V. 2017 Python Machine Learning 2nd ed.. Packt Publishing Ltd, Birmingham, England.

Salloum S. A., Alshurideh M., Elnagar A., &Shaalan K. 2020 Machine Learning and Deep Learning Techniques for Cybersecurity: A Review. In A. E. Hassanien, A Azar, T. Gaber, D Oliva, & F. Tolba Eds.. Proceedings of the International Conference on Artificial Intelligence and Computer Vision AICV2020. AICV 2020. AdvIntellSystComput. 1153, pp. 50-57. https://doi.org/10.1007/978-3-030-44289-7_5

Santos, H. G. 2018 Comparação da performance de algoritmos de machinelearning para a análise preditiva em saúde pública e medicina. Ph. D. Thesis in Epidemiology. Universidade de São Paulo, Faculdade de Saúde Pública, São Paulo SP Brasil. Available online at https://teses.usp.br/teses/disponiveis/6/6141/tde-09102018-132826/publico/HellenGeremiasdosSantos_DR_ORIGINAL.pdf

Santos, H. G., Nascimento, C. F., Izbicki, R., Duarte ,Y. A. O., &Filho, A. D. P. C. 2019 Machine learning for predictive analyses in health: An example of an application to predict death in the elderly in São Paulo, Brazil. Cad Saúde Pública. 357, pp. 00050818. https://doi.org/10.1590/0102-311x00050818

Vandeput, N. 2019 Forecast KPI: RMSE, MAE, MAPE & Bias. Available online at https://towardsdatascience.com/forecast-kpi-rmse-mae-mape-bias-cdc5703d242d

Waring, J., Lindvall, C., &Umeton, R. 2020Automated machine learning: Review of the state-of-the-art and opportunities for healthcare. ArtifIntell Med. 104, pp. 101822. https://doi.org/10.1016/j.artmed.2020.101822

Zhai , X., Yin, Y. , Pellegrino, J. W., Haudek, K. C., & Shi, L. 2020 Applying machine learning in science assessment: a systematic review. Stud Sci Educ. 561, pp. 111-151. https://doi.org/10.1080/03057267.2020.1735757

******