

# Forecasting Return Visits of the Emergency Department

Natalia Mordvanyuk<sup>a</sup>, Ferran Torrent-Fontbona<sup>a</sup>, José-Maria Inoriza<sup>b</sup>, Beatriz López<sup>a</sup>,

<sup>a</sup>University of Girona, Campus Montilivi, building EPS4, 17071 Girona, Spain

<sup>b</sup>Serveis de Salut Integrats Baix Empordà. Grup de Recerca en Serveis de Salut i Resultats en Salut (GReSSiReS), 17320 Palamós

---

## Abstract

*Objective:* Emergency Department (ED) revisits are aggravating ED patient overcrowding. — ADAPT The use of artificial intelligence techniques to find out which Single Nucleotide Polymorphisms (SNPs) promote the development of a disease is one of the features of medical research, as such techniques may potentially aid early diagnosis and help in the prescription of preventive measures. In particular, the aim is to help physicians to identify the relevant SNPs related to Type 2 diabetes, and to build a decision-support tool for risk prediction.

*Methods:* In order to support ED manager, we propose the NearMiss algorithm in combination with an ANN approach to forecast ED revisits. – ADAPT We use the Random Forest (RF) technique in order to search for the most important attributes (SNPs) related to diabetes, giving a weight (degree of importance), ranging between 0 and 1, to each attribute. Support vector machines and logistic regression have also been used since they are two other machine learning techniques that are well-established in the health community. Their performance has been compared to that achieved by RF. Furthermore, the relevance of the attributes obtained through the use of RF has then been used to per-

---

\*Corresponding author

\*\*Corresponding author

Corresponding author

*Email addresses:* natalia.mordvanyuk@udg.edu (Natalia Mordvanyuk), ferran.torrent@udg.edu (Ferran Torrent-Fontbona), jminoriza@ssibe.cat (José-Maria Inoriza), beatriz.lopez@udg.edu (Beatriz López)

form predictions with  $k$  nearest neighbour method weighting attributes in the similarity measure according to the relevance of the attributes with RF.

*Results:* The method is applied to a dataset of 12 years and outperforms previous works in the literature with a prediction AUC of 89.6%. – ADAPT Testing is performed on a set of 677 subjects. RF is able to handle the complexity of features’ interactions, overfitting, and unknown attribute values, providing the SNPs’ relevance with an up to 0.89 area under the ROC curve in terms of risk prediction. RF outperforms all the other tested machine learning techniques in terms of prediction accuracy, and in terms of the stability of the estimated relevance of the attributes.

*Conclusions:* TODO - ADAPT The random forest is a useful method for learning predictive models and the relevance of SNPs without any underlying assumption.

*Keywords:* Emergency Department, Return visits, Predictive model, Artificial Neural Networks, Imbalanced learning problem

---

## 1. Introduction

Emergency departments (EDs) have drawn attention to patient overcrowding. To alleviate the problem, several tools are being studied for forecasting ED attendances [1, 2, 3], predicting hospital admissions [4], defining triage protocols  
5 [5], etc., with the final goal of aiding ED managers to improve the quality of the ED service. In that regard, ED revisits have been proposed as a quality assurance measure [6]. However, ED revisits are not so important as a reliable quality indicator, but because the conditions that initially were missed (e.g. appendicitis, acute myocardial infection, fracture, subarachnoid haemorrhages)  
10 and the health implications to patients [7].

The burden of ED revisits vary from one hospital population to another one. For example, [8] identifies 7.5% of returning visits in a time horizon of 3 days, and up to 22.4% within 30 days. Revisits are also related to hospital admissions. For example, [9] identifies 8.2% of returning visits in a time horizon of 3 days,

15 with 29% of which are hospital admissions, and up to 19.9% within 30 days, from which the 28% are admissions. The cost of revisits have been estimated in 30.3% of the primary events in a time horizon of 3 days, while the percentage increases up to 118% for a month period [9]. Therefore, it is necessary to review ED attendances protocols and tools to prevent revisits happening [7].

20 Our research is concerned with forecasting ED revisits. To that end, we explore the use of Artificial Neural Networks (ANN) on ED data of the Hospital of Palamós, which is located in a tourist region. We used 12 years of data from the ED hospital records, from which the prediction models have been built, tested and analysed. Data has been complemented with calendar, weather, and  
25 the phase of the moon variables in order to improve the accuracy of ED revisits forecasts.

The paper is organised as follows: First, in Section ??, we expose the related work. Afterwards, in Section ??, we describe and justify the methodology and the criteria used in data cleaning and pre-processing. Next, we explain the  
30 experimentation set-up and the results achieved in Section ?. In Section ?, we conclude the explanation of this work and point out some future work lines.

RELATED WORK INCLUIR.HO AQUI o CONMPROVAR QUE NO SIGUI  
= DISCUSSIO

Forecasting ED visits has been approached with ensemble decision trees in  
35 [10], achieving results up to 73% AUC with one-year data (2012). [11] proposed a decision-tree approach with a similar dataset achieving 70% AUC. We outperform these results with the use of ANN with an accurate preprocessing stage and a greater dataset (13 years).

Other works such as [12] perform meta-analysis in order to obtain ratios  
40 to then do predictions. The authors consider three scenarios according to the factors that causes the revisit: patient demographics, prior year utilizations (i.e., ED revisit within 72 h (yes/no), the total cost, and so on), and co-morbidities. The results achieved, however, are poor, reaching in the best scenario 74% AUC. Moreover, the dataset analysed in [12] is biased towards elderly people, while  
45 our study covers all population ages.

[13] developed a clinical tool to identify discriminatory characteristics that can predict paediatrics patients who will return within 72 hours to the Pediatric emergency department. Their method consists of applying particle Swarm Optimization (PSO) techniques for feature selection coupled with an optimization-based discriminant analysis model (DAMIP) for identifying a classification rule with relatively small subsets of discriminatory factors that can be used to predict the return visits. We have also considered the situation of paediatrics to compare our methodology with the work just mentioned, obtaining also better results.

[14] analyses monthly revisits rates grouping them according to three factors: illness related (up to 75.6%), patient related (13.3%) and doctor related (11.1%). A similar study is reported in [15] with a different outcome: illness related (up to 35%), patient related (35%) and doctor related (50%). Although, [14] highlights that the key issue is the good patient-physician relationship instead of a single factor, e.g. anxious patients versus medical errors or diseases versus optimal therapy. This should be further studied but it is out of the scope of our work.

## 2. Material and methods

ADAPT - The Biomedical Research Institute of Girona has been gathering information about the SNPs of subjects with their corresponding diagnoses (T2D, glucose intolerance), as well as that of healthy subjects. Based on the available data, a T2D risk prediction model has been obtained with the use of the RF technique, from which the relevance of each feature is obtained by using the Gini importance [? ].

### 2.1. The problem

We formulate the problem as a classification problem: whether an ED attendance will become a revisit or not. Then, we solve it by means of an ANN approach, in agreement with some past successful applications of this technique for ED attendances forecasting [16, 17].

The input of the method,  $\vec{X}$ , consists of ED attendances data enriched with  
 75 exogenous variables. ED attendances are described according to the following  
 variables: Arrival timestamp, patient ID, age, patient origin, service origin,  
 state, circuit, triage value (ICD9 code, ICD9-aggregated code, ICD9-CMD),  
 departure timestamp, destination place, destination service, diagnosis (up to 15  
 detailing variables), sex, birthday, foreigner (out of the hospital population),  
 80 patient residence region data (up to 11 variables), nationality, and exitus.

On the other hand, exogenous data include: calendar information (month  
 of the year, workable day), weather data (average, maximum and minimum  
 temperature of the day; maximum, average and relative humidity of the day;  
 and 24h solar radiation), and moon data (moon illumination portion).

85 Summing up, the input vector  $\vec{X}$  has a total of 75 variables. Each ED input  
 is accompanied with a the revisit label  $y$ ,  $y = 0$  meaning no-revisit, or  $y = 1$   
 meaning that the person comes again to the ED in the next  $h$  days.

## 2.2. Pre-processing

A pre-process has been applied in order to select attributes, remove spuri-  
 90 ous variables (such as patient id) and variables with the same meaning or too  
 correlated, transform categorical variables to numerical, scale variables within  
 $[0, 1]$  and deal with missing values.

Regarding the pre-processing of categorical variables, we need to distinguish  
 between variables with a wide range of values (such as diagnose codes) from  
 95 variables with few values. Therefore, first we apply a regular simplex method  
 [18] to convert a categorical variable  $v^i$  with  $N$  values to  $N$  variables  $v_1^i, \dots, v_N^i$ ,  
 $v_j^i \in \{0, 1\}$ . Next, variables  $v_1^i, \dots, v_N^i$  with a big  $N$  are coded as a 0-1 string  
 in a single variable  $v'^i$ . After the categorical variable encoding, a correlation  
 analysis performs the dropping off of some binary variables. The dimension of  
 100 the input vector has been reduced from 77 to 24 after selecting variables, while  
 increasing again up to 66 due to the categorical variable encoding.

In order to deal with missing values, we use an imputation method that  
 assigns a value to missing features ('0').

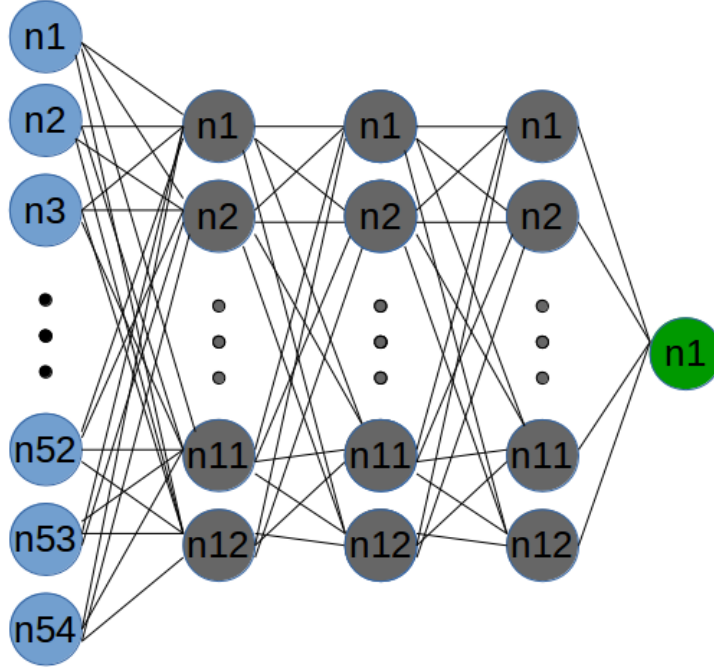


Figure 1: Structure of the Multi-layer Perceptron

### 2.3. Artificial Neural Network Model

105 The network topology is Multi-layer Perceptron classifier with an input of 54 nodes, and, for the sake of simplicity, 3 hidden layers with  $\text{trunc}(n_{features}/2)$  nodes each one. We set up up to 500 iterations during the training of the network. The activation function used is *relu*, and the used solver is *adam*. Figure 1 shows the structure of the Multi-layer Perceptron.

110 [8] uses logistic regression to study the prediction horizon and determines that the best cut-off is 9 days. However, most of the previous studies use a 3 day horizon [19, 20, 13]. Therefore we have set the time horizon to 3 days. Figure 2 shows the distribution of patients according to the time horizon  $h$ , and in the Results section the impact of the  $h$  value is analysed.

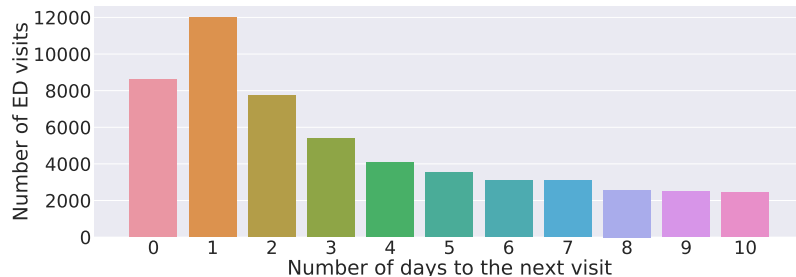


Figure 2: Number of patients (y-axis) regarding the time horizon  $h$  (x-axis).

115 *2.4. Dataset and undersampling*

The methodology has been implemented and tested in a PC with a processor Intel Core i7-4790 CPU @ 3.60GHz x 4, with a 15.6GB of Memory, and a NVIDIA GT218 [GEMForce 210] graphics card. The experiments have been carried out with the ED attendance data of the Hospital of Palamós. The implementation has been carried out with the support of scikit-learn [21], pandas 120 [22], and PyAstronomy [23].

The Hospital of Palamós data consists of the ED visits from January 1st 2002 to December 31st 2014 (13 years). There are a total of 705075 records; the number of revisits in all of this period is 33690 (for  $h \leq 3$ ). Since it is an 125 imbalanced dataset, we used the 1st version of NearMiss undersampling method [24] to balance it. This algorithm selects samples from the majority class for which the average distance of the  $k$  nearest samples of the minority class is the smallest (see Figure 3)

*2.5. Experimental set-up*

130 A 10-cross validation method has been used for experimentation purposes. We decided to apply 10-folds cross-validation methodology following the example of [13] and because the dataset has been considerably reduced after the balancing step.

We also implemented logistic regression (LR) and random forests (RF) methods 135 to compare its performance with our methodology. These methods have

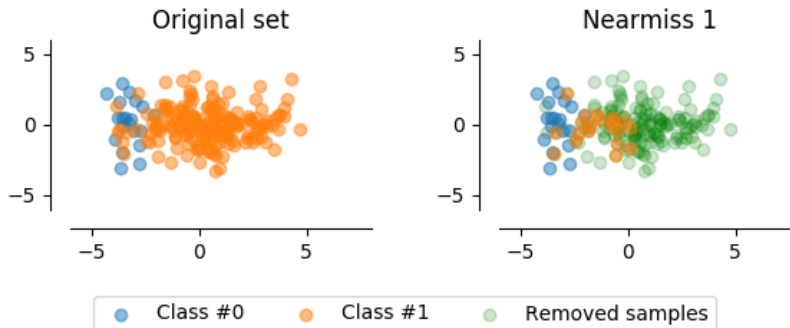


Figure 3: An illustration of the NearMiss 1 undersampling method. Source:[25]

Table 1: Comparison of DAMIP/PSO results against our methodology.

Method	Accuracy	AUC
<b>DAMIP/PSO results</b>	83.1	-
<b>ANN</b>	<b>85.99</b> (std:2.92)	93.12(std:2.39)

been chosen because they have been used in other previous works [8, 20, 13].

### 3. Results

We have also extracted the pediatric data from our dataset in order to compare our method with Lee’s work [13]. Table 3 shows that our method also  
 140 achieves better results than [13].

The impact of the  $h$  value is provided in Figure 5. We achieve an AUC score around 90% when  $h \leq 0$  (revisits happens the same day), but it decreases as  $h$  increases, reaching the lowest value (around 88.00) when  $h$  is 10.

#### 3.1. Initial results

Results are provided in Table 2. The proposed methodology achieves an  
 145 AUC up to 89.7% with the ANN model, outperforming previous related works. Even with RF we obtain better results (88.25%) than in previous works, since the highest AUC value observed with decision trees is around 73% (see [10]).



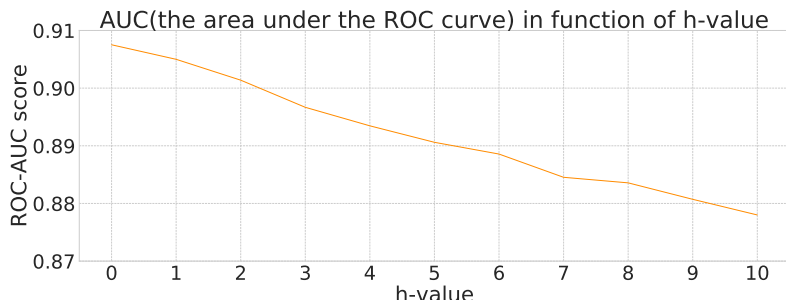


Figure 4: AUC of the ANN regarding to revisit horizon  $h$ .

Table 2: Results

Method	AUC (std)	Train t (s)	Test t (s)
<b>LR</b>	87.8461 (5.246)	0.9586	0.0010
<b>RF</b>	88.2596 (4.542)	6.3249	0.1425
<b>ANN</b>	<b>89.7082</b> (4.263)	11.0139	0.0031

This is probably due to the exhaustive dataset we have (up to 13 years), and  
 150 the preprocessing step we proposed. In fact, if we use the data as provided,  
 the ANN model only achieves a 83.69% AUC. Significant differences can also  
 be seen in terms of training computational time, in favour of LR. However the  
 testing computational time of LR and ANN is of the same magnitude.

### 3.2. Sensitive analysis on $h$

155 The impact of the  $h$  value is provided in Figure 5. We achieve an AUC score  
 around 90% when  $h \leq 0$  (revisits happens the same day), but it decreases as  $h$   
 increases, reaching the lowest value (around 88.00) when  $h$  is 10.

### 3.3. Comparative analysis with Pediatric

We have also extracted the pediatric data from our dataset in order to com-  
 160 pare our method with Lee’s work [13]. Table 3 shows that our method also  
 achieves better results than [13].

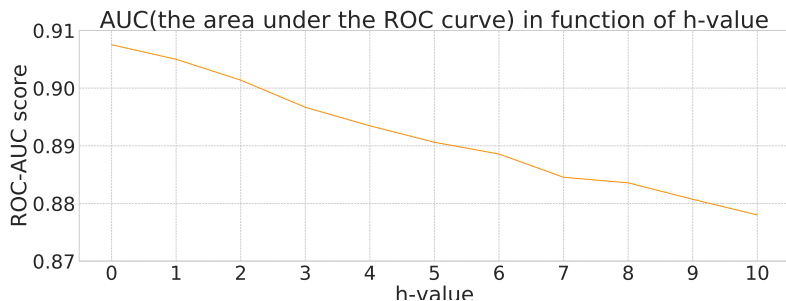


Figure 5: AUC of the ANN regarding to revisit horizon  $h$ .

Table 3: Comparison of DAMIP/PSO results against our methodology.

Method	Accuracy	AUC
<b>DAMIP/PSO results</b>	83.1	-
<b>ANN</b>	<b>85.99</b> (std:2.92)	93.12(std:2.39)

#### 4. Discussion

We have performed predictions in a general basis, without focussing on specific factors as other works do. For example, [14] analyses monthly revisits rates grouping them according to three factors: illness related (up to 75.6%), patient related (13.3%) and doctor related (11.1%). A similar study is reported in [15] with very different outcome: illness related (up to 35%), patient related (35%) and doctor related (50%), the later case due to a high percentage of suboptimal management than due to misdiagnosis. Although, [14] highlights that the key issue is the good patient-physician relationship more than focussing on a single factor (i.e. anxious patients versus medical errors, diseases versus optimal therapy), that should be further studied.

On the other hand, we have considered all patients exhibiting the same behaviour, while other works, as [20] focus on the identification of patients with frequent revisits, and they alert that there could be patients that accumulate up to 47% of revisits (regarding their total amount of EDs attendances). Therefore, our forecasting model could be probably improved by analysing such specific

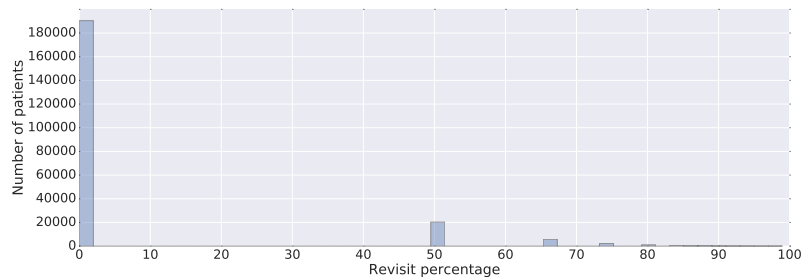


Figure 6: Distribution of the percentage of revisits

situations.

En imatges he posat la corba roc respecte h (pero queda lletja)

180 I tambe totes les ROC juntes respecte h=1,3,7,10

Per aixó he decidit deixar aquesta

ULL:La grafica Revisites correspon a h = 10. A imatges trobareu 2 grafiques sobre revisites, una quan h es 3 ReturnPercentatge3days.pdf i una altre quan h es 10 ReturnPercentatge10days.pdf

185 Figure 6 shows that there are 20280 ULL: amb h3 es 16577 patients that revisit ED 50 percent of times, 5787 ULL: amb h3 es 3775 that revisit it 66 percent of times, 4923 ULL: amb h3 es 2322 revisit ED more then 70 percent of times.

## 5. Conclusions

190 ED revisits is increasing the ED overload while this is a hospital service that is characterised by a high overcrowding. This paper presents an ANN model for ED revisits forecasting in order to provide to ED managers tools to alleviate the problem. Experimentation is carried out with a dataset of 13 years of data from the Hospital of Palamós. After an accurate pre-processing stage of the data,  
 195 our ANN model reaches an AUC of 89.6032, outperforming all of the previous works. Nevertheless, there is room for improvement.

First, we need to identify the discharge situation that causes revisits in order

to support the development of the appropriate guidelines to prevent them. Second, revisits could be analysed regarding different stratification of the hospital population, such as identifying frequent revisits [20], visits from people suffering a chronic diseases [26] or elderly people [27]. Finally, our study has been done with data from a single hospital. Therefore, we need to extend this study with the collaboration with other hospitals in order to address the problem at the adequate healthcare level [28].

## Acknowledgements

Funding: This work was supported by the University of Girona [grant number MPCUdG2016]; and the research group SITES awarded with distinction by the Generalitat de Catalunya (2017 SGR 1551).

This work was supported by Fondo Europeo de Desarrollo Regional (FEDER) funds.

## References

- [1] Y. e. a. Sun, Forecasting daily attendances at an emergency department to aid resource planning, *BMC Emergency Medicine* 9 (1) (2009) 1. doi:10.1186/1471-227X-9-1.  
URL <http://bmccemergmed.biomedcentral.com/articles/10.1186/1471-227X-9-1>
- [2] L. M. Schweigler, J. S. Desmond, M. L. McCarthy, K. J. Bukowski, E. L. Ionides, J. G. Younger, Forecasting models of emergency department crowding., *Academic emergency medicine : official journal of the Society for Academic Emergency Medicine* 16 (4) (2009) 301–8. doi:10.1111/j.1553-2712.2009.00356.x.  
URL <http://www.ncbi.nlm.nih.gov/pubmed/19210488>
- [3] S. Poole, S. Grannis, N. H. Shah, Predicting Emergency Department Visits., *AMIA Joint Summits on Translational Science proceedings. AMIA*

- 225 Joint Summits on Translational Science 2016 (2016) 438–45.  
URL <http://www.ncbi.nlm.nih.gov/pubmed/27570684><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5001776>
- [4] D. Golmohammadi, Predicting hospital admissions to reduce emergency department boarding, *International Journal of Production Economics* 182. doi:10.1016/j.ijpe.2016.09.020.  
230
- [5] M. R. Vaghasiya, M. Murphy, D. O’Flynn, A. Shetty, The emergency department prediction of disposition (EPOD) study, *Australasian Emergency Nursing Journal* 17 (4) (2014) 161–166. doi:10.1016/j.aenj.2014.07.003.  
235 URL <http://www.ncbi.nlm.nih.gov/pubmed/25112947><http://linkinghub.elsevier.com/retrieve/pii/S1574626714000639>
- [6] K. D. Keith, J. J. Bocka, M. S. Kobernick, R. L. Krome, M. A. Ross, Emergency department revisits., *Annals of emergency medicine* 18 (9) (1989) 964–8.  
240 URL <http://www.ncbi.nlm.nih.gov/pubmed/2764329>
- [7] A. C. Meltzer, J. M. Pines, What Big Data Can and Cannot Tell Us About Emergency Department Quality for Urolithiasis, *Academic Emergency Medicine* 22 (4) (2015) 481–482. doi:10.1111/acem.12639.  
URL <http://doi.wiley.com/10.1111/acem.12639>
- 245 [8] K. L. Rising, T. W. Victor, J. E. Hollander, B. G. Carr, Patient Returns to the Emergency Department: The Time-to-return Curve, *Academic Emergency Medicine* 21 (8) (2014) 864–871. doi:10.1111/acem.12442.  
URL <http://www.ncbi.nlm.nih.gov/pubmed/25154879><http://doi.wiley.com/10.1111/acem.12442>
- 250 [9] R. Duseja, N. S. Bardach, G. A. Lin, J. Yazdany, M. L. Dean, T. H. Clay, W. J. Boscardin, R. A. Dudley, Revisit rates and associated costs after an emergency department encounter: a multistate analysis., *Annals of internal*

medicine 162 (11) (2015) 750–6. doi:10.7326/M14-1616.

URL <http://annals.org/article.aspx?doi=10.7326/M14-1616>  
255 <http://www.ncbi.nlm.nih.gov/pubmed/26030633>

- [10] B. Jin, Y. Zhao, S. Hao, A. Y. Shin, Y. Wang, C. Zhu, Z. Hu, C. Fu, J. Ji, Y. Wang, Y. Zhao, Y. Jiang, D. Dai, D. S. Culver, S. T. Alfreds, T. Rogow, F. Stearns, K. G. Sylvester, E. Widen, X. B. Ling, Prospective stratification of patients at risk for emergency department revisit: resource  
260 utilization and population management strategy implications, BMC Emergency Medicine 16 (1) (2016) 10. doi:10.1186/s12873-016-0074-5.

URL <http://bmcemergmed.biomedcentral.com/articles/10.1186/s12873-016-0074-5>

- [11] S. Hao, B. Jin, A. Y. Shin, Y. Zhao, C. Zhu, Z. Li, Z. Hu, C. Fu, J. Ji, Y. Wang, Y. Zhao, D. Dai, D. S. Culver, S. T. Alfreds, T. Rogow, F. Stearns, K. G. Sylvester, E. Widen, X. B. Ling, Risk Prediction of Emergency Department Revisit 30 Days Post Discharge: A Prospective Study,  
265 PLoS ONE 9 (11) (2014) e112944. doi:10.1371/journal.pone.0112944.

URL <http://dx.plos.org/10.1371/journal.pone.0112944>

- 270 [12] G. Pellerin, K. Gao, L. Kaminsky, Predicting 72-hour emergency department revisits, The American Journal of Emergency Medicine doi:10.1016/j.ajem.2017.08.049.

URL <http://www.ncbi.nlm.nih.gov/pubmed/28855065>  
<http://linkinghub.elsevier.com/retrieve/pii/S0735675717306988>

- 275 [13] E. K. Lee, F. Yuan, D. A. Hirsh, M. D. Mallory, H. K. Simon, A clinical decision tool for predicting patient care characteristics: patients returning within 72 hours in the emergency department, in: AMIA Annual Symposium Proceedings, Vol. 2012, American Medical Informatics Association, 2012, p. 495.

- 280 [14] C.-L. Wu, F.-T. Wang, Y.-C. Chiang, Y.-F. Chiu, T.-G. Lin, L.-F. Fu, T.-L. Tsai, Unplanned Emergency Department Revisits within 24 Hours in a

Referral Hospital, *J Emerg Crit Care Med* 19 (4) (2008) 146–154.  
URL [http://www.seccm.org.tw/magazine/pdf/200819\\_{\\_}0402.pdf](http://www.seccm.org.tw/magazine/pdf/200819_{_}0402.pdf)

285 [15] J. Sri-On, A. Nithimathachoke, G. P. Tirrell, S. Surawongwattana,  
S. W. Liu, Revisits within 48 Hours to a Thai Emergency De-  
partment., *Emergency medicine international* 2016 (2016) 8983573.  
doi:10.1155/2016/8983573.

URL <http://www.ncbi.nlm.nih.gov/pubmed/27478642><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4961813>

290 [16] M. P. Joy, S. Jones, Predicting Bed Demand in a Hospital using Neural  
Networks and ARIMA models: a Hybrid Approach, in: *ESANN proceed-*  
*ings*, Bruges, 2005, pp. 127–132.

URL [http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.443.2003\\_{&}rep=rep1\\_{&}type=pdf](http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.443.2003_{&}rep=rep1_{&}type=pdf)

295 [17] S. S. Jones, A. Thomas, R. S. Evans, S. J. Welch, P. J. Haug,  
G. L. Snow, Forecasting Daily Patient Volumes in the Emergency  
Department, *Academic Emergency Medicine* 15 (2) (2008) 159–170.  
doi:10.1111/j.1553-2712.2007.00032.x.

300 URL <http://www.ncbi.nlm.nih.gov/pubmed/18275446><http://doi.wiley.com/10.1111/j.1553-2712.2007.00032.x>

[18] H. Niitsuma, T. Okada, *Covariance and PCA for Categorical Variables*,  
Springer, Berlin, Heidelberg, 2005, pp. 523–528. doi:10.1007/11430919\_  
61.

URL [http://link.springer.com/10.1007/11430919\\_{\\_}61](http://link.springer.com/10.1007/11430919_{_}61)

305 [19] B. Lerman, M. S. Kobernick, Return visits to the emergency department,  
*The Journal of emergency medicine* 5 (5) (1987) 359–362.

[20] J. Ryan, J. Hendler, K. P. Bennett, Understanding Emergency Department  
72-Hour Revisits Among Medicaid Patients Using Electronic Healthcare  
Records, *Big Data* 3 (4) (2015) 238–248. doi:10.1089/big.2015.0038.

310 URL <http://online.liebertpub.com/doi/10.1089/big.2015.0038>

- [21] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: Machine learning in python, *J. Mach. Learn. Res.* 12 (2011) 2825–2830.  
315 URL <http://dl.acm.org/citation.cfm?id=1953048.2078195>
- [22] pandas, <https://pandas.pydata.org>, online; accessed 25 November 2017.
- [23] PyAstronomy, <http://www.hs.uni-hamburg.de/DE/Ins/Per/Czesla/PyA/PyA/index.html>, online; accessed 25 November 2017.
- [24] I. Mani, I. Zhang, knn approach to unbalanced data distributions: a case  
320 study involving information extraction, in: *Proceedings of workshop on learning from imbalanced datasets*, Vol. 126, 2003.
- [25] NearMiss, [http://contrib.scikit-learn.org/imbalanced-learn/stable/auto\\_examples/under-sampling/plot\\_nearmiss.html](http://contrib.scikit-learn.org/imbalanced-learn/stable/auto_examples/under-sampling/plot_nearmiss.html), online; accessed 1st February 2018.
- 325 [26] S. Adeyemi, E. Demir, T. Chausalet, Towards an evidence-based decision making healthcare system management: Modelling patient pathways to improve clinical outcomes, *Decision Support Systems* 55 (1) (2013) 117–125. doi:10.1016/J.DSS.2012.12.039.  
URL [http://www.sciencedirect.com/science/article/pii/S0167923613000092?{}\\_rdoc=1&{}\\_fmt=high&{}\\_origin=gateway&{}\\_docanchor=&{}\\_md5=b8429449ccfc9c30159a5f9aeaa92ffb&{}\\_dgcid=raven{}\\_sd{}\\_recommender{}\\_email](http://www.sciencedirect.com/science/article/pii/S0167923613000092?{}_rdoc=1&{}_fmt=high&{}_origin=gateway&{}_docanchor=&{}_md5=b8429449ccfc9c30159a5f9aeaa92ffb&{}_dgcid=raven{}_sd{}_recommender{}_email)
- 330 [27] L. P. Latham, S. Ackroyd-Stolarz, Emergency department utilization by older adults: a descriptive study., *Canadian geriatrics journal : CGJ* 17 (4) (2014) 118–25. doi:10.5770/cgj.17.108.  
URL <http://www.ncbi.nlm.nih.gov/pubmed/25452824http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4244125>



[28] B. D. Shy, G. T. Loo, T. Lowry, E. Y. Kim, U. Hwang, L. D. Richard-  
340 son, J. S. Shapiro, Bouncing Back Elsewhere: Multilevel Analysis of  
Return Visits to the Same or a Different Hospital After Initial Emer-  
gency Department Presentation, *Annals of Emergency Medicine* (In press).  
doi:10.1016/j.annemergmed.2017.08.023.  
URL [http://www.annemergmed.com/article/S0196-0644\(17\)31488-9/](http://www.annemergmed.com/article/S0196-0644(17)31488-9/)  
345 pdf