Received: 29 January 2019 | Revised: 2 February 2019 | Accepted: 5 February 2019

DOI: 10.1002/hpm.2763

RESEARCH ARTICLE

WILEY

Prediction of the increase in health services demand based on the analysis of reasons of calls received by a customer relationship management

Mª. Isabel Ramos[1] | Juan José Cubillas[2] | Juan Manuel Jurado[3] | Wilfredo Lopez[4] | Francisco R. Feito[5] | Manuel Quero[6] | Jose Maria Gonzalez[7]

[1]Department of Cartography, Geodesy and Photogrammetry Engineering, University of Jaen, Jaen, Spain

[2]Department of Health, Indra Soluciones Tecnologicas S.A., Madrid, Spain

[3]TIC-144 Andalusian Research Plan (PAI), Department of Computer Science, University of Jaen, Jaen, Spain

[4]Doctor of Salud Responde, Technology in Health, Jaen, Spain

[5]Department of Computer Science, University of Jaen, Jaen, Spain

[6]Nurse of Salud Responde, Technology in Health, Jaen, Spain

[7]Director of Salud Responde, Technology in Health, Jaen, Spain

Correspondence
Juan José Cubillas, Department of Health, Indra Soluciones Tecnologicas S.A., Madrid, Spain.
Email: jjcubillas@indra.es

Summary

Currently, customer relationship management (CRM) tools are very important in our society because they provide a comunication channel to the healthcare system for patients. Salud Responde is a CRM that provides many health services for the entire population of Andalusia, in southern Spain. The number and frequenzy of phone calls received change along the year. They depend on many factors, such as weekdays, seasons, vaccination campaigns, environmental factors, pandemic periods, etc. All these are the main reasons number of health calls changes along the year. This variability makes that the current management of resources for offering emergency services based on historical data is inefficient. The factors, which influence the phone calls along the year, are different from one period to another. Therefore, it is clear to demand an improved in the current management system. In this context, the main goal for this research is to develop an expert system able to identify and analyze, using different data mining algorithms, the most relevant factors to predict the variability of health service demand. Thus, here, it is proposed a methodology in which using reasons calls received in the CRM as input data, it is possible to predict in advance the healthcare resources demand.

KEYWORDS

big data analytics, CRM, healthcare, primary care, predictive system

1 | INTRODUCTION

Customer relationship management tools (CRMs)[1] are actually very useful for our society. These provide a unique access point of healthcare system for patients inside hospitals, health centers, or specific emergency centers. Salud Responde[2] is a CRM that has a large catalog of healthcare services such as medical appointment request, health information, tele-continuity care services, health advice based on medical decision protocols (flu, Ebola, heat wave, palliative care, …), appointment reminders, etc. This type of services provides great benefits to patients since they can get their medical doubts and administrative procedures solved without going to any health center. Also, it allows people to be attended through multiple channels such as telephone, email, mobile app, etc. In addition, it provides a great benefit to health systems since it allows them to manage their resources efficiently contributing to maintain a sustainable health system[2] with important economical savings.

The CRM Salud Responde works as follows: when a request is received (by telephone, email, or mobile app), it is attended, in a first level, by a specialized operator that solves the request acording to the type of it, for example, giving a medical appointment for its primary care center, providing medical information, solving questions about some administrative procedure, etc. In this sense, if the operator has any doubt about the answer to the request submitted, because it is a medical request, which requires to be solved by sanitary professional, then the operator transfers the request to the nurse department. Here, all the requests of this type received by the CRM are analyzed. Next, a nurse interviews the user who has called to know exactly its health request. Finally, the request of the user is solved by a nurse giving a specific health advice or deriving its query to another level of care (to a primary health care doctor or an emergencies doctor) (Figure 1).

Every year, Salud Responde performs more than 40 million procedures, 12 million of which are phone calls attended by operators and nurses. It supposes a 30% of the all procedures atended. Therefore, it becomes a clue having a correct sizing of the resources, in this case operators and nurses, every time to be able to care for patients. In this work, a study of the variables that influence the users to demand sanitary services is made. Thereby, the expert system developed is able to analyze such variables and make a prediction of the number of request calls that will be received in the service 7 days in advance.

It is known that during some periods of time, there is an increase of sanitary demand, for example, when there is an outbreak of influenza,[3] when temperatures drop,[4] or during in the influenza vaccination campaign since users request an appointment to be vaccinated, in periods of allergies due to flowering of plants,[5] etc. These lead us to use as input variables of the predictive models the causes of calls received by the CRM of Salud Responde. The aim objective of the work is to implement a model capable of detecting in advance health problems that will carry demand for health services by the users, in this sense, a predictive model that calculates in advance how many users will call the CRM Salud Responde to request medical attention.

2 | METHODS

The causes and the number of the calls received are very variable depending on the day of the week, the season of the year, or if it coincides with a vaccination advertising campaign. Also, it influences environmental factors such as temperature, humidity, etc. Figure 2 shows the number of calls received during the year 2016.

As an example in Figure 2, it is shown that there are very significant differences in the number of calls received from 1 day to another and from one season to the next. Thus, that is the case of a Monday in March, March 28, when there were 53 501 calls compared with the Monday of the following week when it received 47 030 calls (April 04).

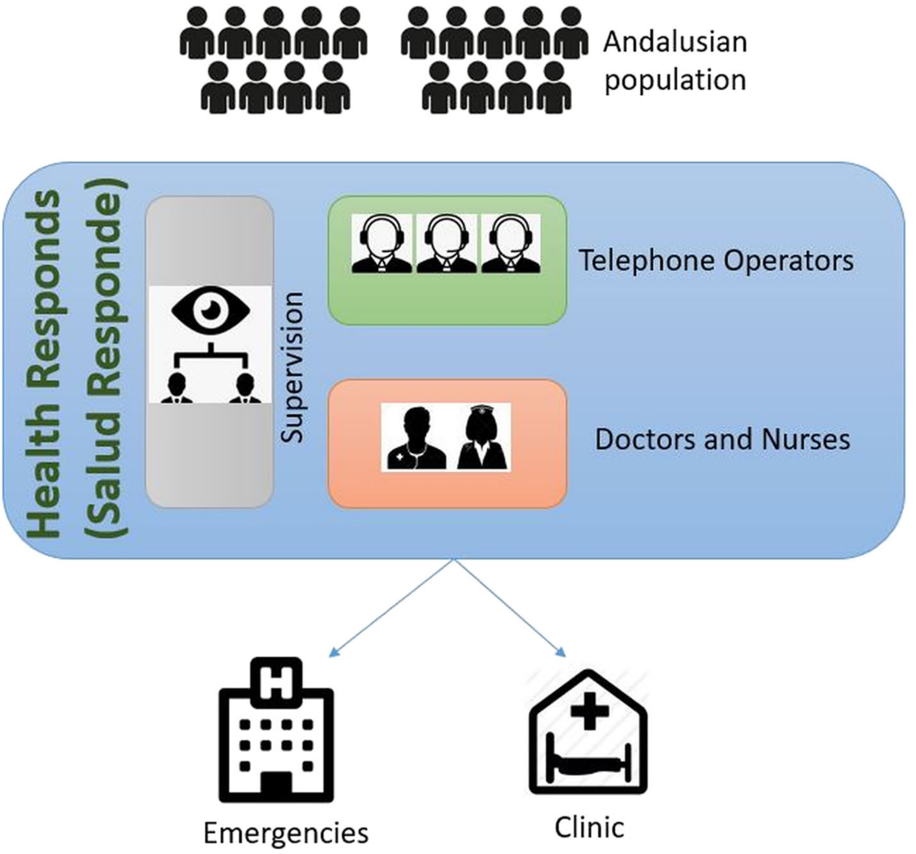

FIGURE 1 Work scheme of the health service respond

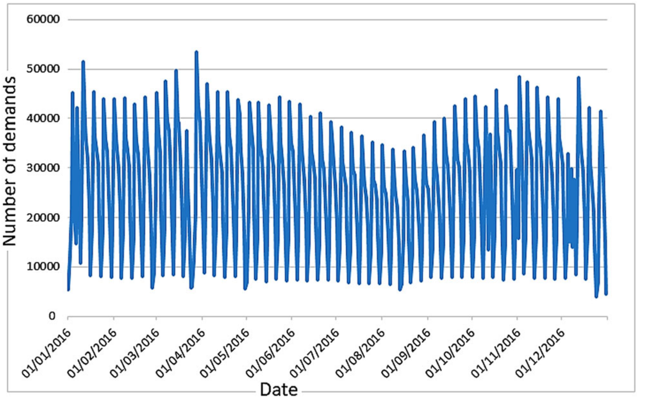

FIGURE 2 Calls received during year 2016

Hence, the complexity of the correct sizing of services that often causes certain errors when it is based only on experience and taking into account only the historical series. A bad sized of the service of this type implies giving a bad attention to the citizen, who cannot contact Salud Responde when the demand of the system is saturated.

For a better understanding how data vary from year to year, Figure 3 is shown. Here, the number of calls is grouped by months in 2015 and 2016. It can be seen how the largest accumulation of calls in 2016 was in November, which coincides with the period of influenza; however, in 2015, the highest demand because of an influenza outbreak was during the months of January and February.

The heterogeneity in the distribution of the number of calls from 1 month to another makes very complex to dimension the health resources taking into account only the corresponding time periods. In this sense periods of influenza, allergies by pollen or vaccination campaigns for pandemics do not coincide from 1 year to the next. Therefore, it is a key to be able to extract variables that allow us to detect these periods in their initial phases in order to anticipate the increase in demand of health calls. To do this, in our predictive model, we introduce as input data the reasons for the calls received by the CRM about flu, allergy, fever, vaccination, etc.

Next, we describe the variables that we will use in our predictive model as follows:

- Call day (Monday, Tuesday, Wednesday, etc.)
- Month
- Type of day (working day or holiday)
- Forecast of temperatures
- Week of the month (since certain medications are prescribed during the first week of the month).
- Days after holidays. We indicate if the day is preceded by one or more holidays; it is known that after the holidays, there is more demand for health calls.
- Days before holidays. That is, we indicate how many days are left to get to a holiday, since it is also known that if the next day there is a holiday, more calls are received.
- Percentage of calls received calling about the flu. When the percentage of calls about the flu increases, it means that a period of flu may be approaching and people are beginning to suffer it.
- Percentage of calls received calling about vaccines. When these types of calls increase, it indicates that the influenza vaccination campaign is being initiated, which increases the number of calls to solve questions about vaccination.
- Percentage of calls received calling about fever. It is an indication of the level of health of patients; this percentage increases significantly in winter and coincides with the coldest temperatures and calls increase to request a primary care physician appointment.
- Percentage of calls about allergy and use of antihistamine. It is indicative that users with greater sensitivity to plant pollens begin to have the first symptoms, which is indicative that the allergy period is approaching.

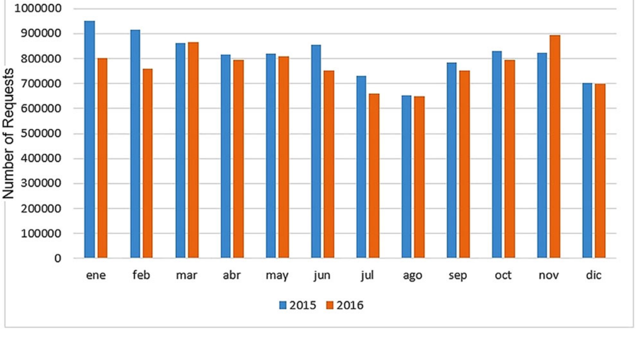

FIGURE 3 Number of calls received between the years 2015 and 2016

Before the study is begun, the influence of the different variables related to the number of calls to Salud Responde is calculated; for this, the minimum description length (MDL)[6] algorithm is used. This algorithm considers each attribute as a simple model and returns a value between −1 and 1. Value 1 means that the attribute has the highest relation to the target attribute, value 0 means that there is no relation, and a negative value means that the attribute is not related to the target attribute and therefore causes noise in the study.

Once the valid attributes are known, they are applicable to the generation of predictive models. For this study, generalized linear models (GLMs)[7] and support vector machines (SVMs) (linear and Gaussian kernel)[8-10] algorithms are used. The algorithms have been developed with Oracle data mining tool.[11] The GLM algorithm is used because linear models make a set of restrictive assumptions, most importantly because the target (dependent variable) is normally distributed conditioned by the value of predictors, with a constant variance regardless of the predicted response value. The advantages of linear model and their restrictions include computational simplicity, an interpretable model form, and the ability to compute certain diagnostic information about the quality of the fit. GLMs relax these restrictions, which are often violated in practice. For example, binary (yes/no or 0/1) responses do not have the same variance across classes. Furthermore, the sum of terms in a linear model typically can have very large ranges encompassing very large negative and positive values. For the binary response example, we would like the response to be a probability in the range between 0 and 1. Also, GLMs accommodate responses that violate the linear model assumptions through two mechanisms: a link function and a variance function. The link function transforms the target range potentially from minus infinity to plus infinity so that the simple form of linear models can be maintained. We configured the GLM algorithm with a confidence level[12] of 95%.

On the other hand, SVMs are a powerful algorithm based on the statistical learning theory. The main advantage of the SVM algorithm is that models have similar functional forms to neural networks and radial basis functions, both popular data mining techniques. However, neither of these algorithms have the well-founded theoretical approach to regularization that forms the basis of SVM. The quality of generalization and ease of training of SVM is far beyond the capacities of these more traditional methods.[13]

In this study, the linear kernel function reduces the original attributes in the training data to a linear equation. A linear kernel works well when there are many attributes in the training data. The Gaussian kernel transforms each case in the training data to a point in an n-dimensional space, where n is the number of cases. The algorithm attempts to separate the points into subsets with homogeneous target values. The Gaussian kernel uses nonlinear separators, but within the kernel space, it constructs a linear equation.[13] Specifically, with the SVM model, the parameters that have been used are as follows: tolerance: 0.001, kernel cache size: 50.000.000, and complexity factor: 0.76.

3 | RESULTS

The results of the first phase of the study correspond to the implementation of the MDL algorithm to identify which attributes are those that most influence the target attributes; in this case, they have been calculated for each of the two cohorts, Table 1.

The regression models are generated from the algorithms already described in the previous section. The results obtained are shown in Table 2 where the precision of models generated from each algorithm is shown. These results concluded that the algorithm theoretically best modeled to our issue is the SVM with linear kernel, with a predictive confidence of 65.79% and an absolute error of 1.824 over an average of 27.895 calls.

Finally, the model is applied, and its effectiveness is verified from January 01, 2017, to 31/08/2017. During this more than half year that the model has been implemented in production, it can be observed that the average absolute error is 2.89%. Figure 4 shows the comparison of real calls and those obtained from the predictive model.

The model has been able to anticipate increases in demand of health calls about influenza in January and about pollen in March. In the following, Figure 5, it is shown as an example the actual calls received and the prediction made by the system for the following week.

TABLE 1 Influence the target attributes

Attribute Name	Important Attribute
Day, week	0,98
Type of day	0,38
Month	0,21
Temperature	0,18
Days before holidays	0,17
Days after holidays	0,15
Percentage of calls referring to influenza	0,14
Percentage of calls regarding vaccine	0,13
Percentage of calls referring to allergy	0,12
Percentage of calls regarding fever	0,12
Week of the month	0,03

TABLE 2 Precision of models

	Predictive Trust	Quadratic Error	Absolute Error
GLM	66,97%	4.139	2.430
SVM con kernel Gausiano	66,94%	4.127	2.307
SVM con kernel lineal	68,79%	3.911	1.824

Abbreviations: GLM, generalized linear model; SVM, support vector machine.

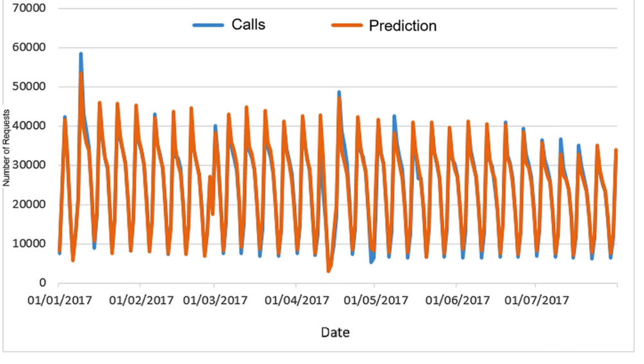

FIGURE 4 Comparison between predictive model and real case

4 | DISCUSSION

In this paper, it is observed that the most important data for the predictive model are seasonal variables, such as the day of the week, month, type of day, etc. But what makes the model better suited to reality is the use of the reasons for calls of the users received in the CRM, Salud Responde. This is because of the fact that these calls are indicative of a health problem that is beginning to manifest and consequently, in the coming weeks, it will lead to an increase health services demand. For example, the percentage of patients who have called for health advice on flu or fever is a clear indicator that the flu begins to manifest in the population, in its initial phase, which allows health services to

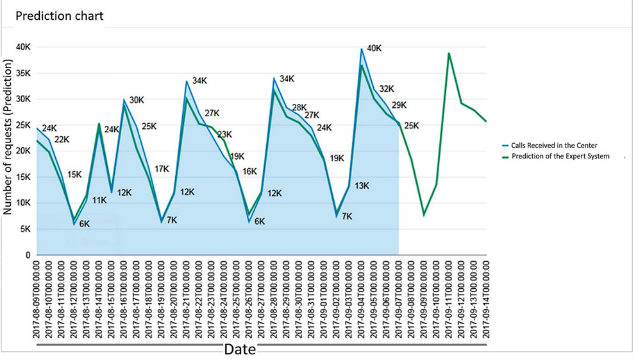

FIGURE 5 Graphic visualization of the model prediction

anticipate the increase in demand. Similarly, the percentage of calls made by users about vaccines or allergies influences the predictive model.

Another attribute with a lot of weight on the predictive study is the temperature changes. These are related to certain pathologies, which means that patients have to go to their primary care doctor.[14] Nowadays, these data temperature variations are very easy to obtain through numerous websites with weather forecast, which provide with a very reliable prediction for the next 10 days.

In short, this type of health demand services is very sensitive to temporality, and if we compare, the series from 1 year to another is very similar, but what really provides us an improvement in predictive models is the use of external variables that give us indications that a health problem may be presented in its initial phase and this will allow health service managers to anticipate the correct dimensioning of human and technical resources.

5 | CONCLUSIONS

The expert system is able to predict the number of patients who will request healthcare to the CRM Salud Responde. This can be a key tool for health center resource managers to correctly design and sizing their services. The use of this tool can help to improve various aspects of healthcare management. First, from an economic point of view, knowing the demand in advance allows managers to prevent having too many staff on duty. Another important aspect that can be achieved with the use of the expert system is improvement in patient satisfaction, as the manager can detect in advance when there will be higher health services demand, thus increasing the quality of healthcare and providing on-time service to users when needed. In short, this tool can help resource managers to design more efficient and sustainable primary care services.

Concluding, in this system, it is a key to use call reasons of the users as input variables in the predictive model. These allow to detect the health services demand in very early phases, when they begin to manifest, allowing the model to take it into account and anticipate the increase in demand with at least a week in advance, which is a key to correctly size the services.

ACKNOWLEDGEMENTS

We thank to Salud Responde and to the project TIN2017-84968-R for the support of the research.

COMPLIANCE WITH ETHICAL STANDARDS

- The study does not have any type of financing

- In this work, there is no conflict of interests. Author J.J. Cubillas declares that he has no conflict of interest. Author M.I. Ramos declares that she has no conflict of interest. Author J.M. Jurado declares that he has no conflict of interest. Author F.R. Feito declares that he has no conflict of interest. Author W. Lopez declares that he has no conflict of interest. Author M. Quero declares that he has no conflict of interest.
- Ethical approval: This article does not contain any studies with human participants performed by any of the authors.

ORCID

Mª. Isabel Ramos https://orcid.org/0000-0002-3006-4166
Juan José Cubillas https://orcid.org/0000-0002-5138-8872
Juan Manuel Jurado https://orcid.org/0000-0002-8009-9033
Francisco R. Feito https://orcid.org/0000-0001-8230-6529

REFERENCES

1. Cubillas JJ, Ramos MI, Feito FR, González JM, Gersol R, Ramos MB. Importancia de los customer relationship management (CRM) sanitarios en las pandemias y alertas sanitarias. *Aten Primaria*. 2015;47(5):267-272.
2. Quero M, Ramos MB, López W, Cubillas JJ, González JM, Castillo JL. Uso de customer relationship management para mejorar la atención sanitaria de la ciudadanía. Servicio Salud Andalucía 24 horas. Salud Responde. *Gac Sanit*. 2016;30(5):397-400.
3. Menon DK, Taylor BL, Ridley SA. Modelling the impact of an influenza pandemic on critical care services in England. *Anaesthesia*. 2005;60(10):952-954.
4. Basu R, Samet JM. Relation between elevated ambient temperature and mortality: a review of the epidemiologic evidence. *Epidemiol Rev*. 2002;24(2):190-202.
5. Vourdas D, Syrigou E, Potamianou P, et al.. Double-blind, placebo-controlled evaluation of sublingual immunotherapy with standardized olive pollen extract in pediatric patients with allergic rhinoconjunctivitis and mild asthma due to olive pollen sensitization. *Allergy*. 1998;53(7):662-672.
6. Davies RH, Twining CJ, Cootes TF, Waterton JC, Taylor CJ. A minimum description length approach to statistical shape modeling. *IEEE Trans Med Imaging*. 2002;21(5):525-537.
7. Nelder JA, Wedderburn RW. Generalized linear models. *Journal of the Royal Statistical Society: Series A (General)*. 1972;135(3):370-384.
8. Burges CJ. A tutorial on support vector machines for pattern recognition. *Data Min Knowl Disc*. 1998;2(2):121-167.
9. Scholkopf B, Sung KK, Burges CJ, et al. Comparing support vector machines with Gaussian kernels to radial basis function classifiers. *IEEE Trans Signal Process*. 1997;45(11):2758-2765.
10. Keerthi SS, Lin CJ. Asymptotic behaviors of support vector machines with Gaussian kernel. *Neural Comput*. 2003;15(7):1667-1689.
11. https://www.oracle.com/technetwork/database/options/advanced-analytics/odm/overview/index.html
12. https://docs.oracle.com/cd/E18283_01/datamine.112/e16808/algo_glm.htm
13. https://docs.oracle.com/cd/E11882_01/datamine.112/e16808/algo_svm.htm#DMCON025
14. López-Vélez R, Molina Moreno R. Cambio climático en España y riesgo de enfermedades infecciosas y parasitarias transmitidas por artrópodos y roedores. *Rev Esp Salud Publica*. 2005;79(2):177-190.

How to cite this article: Ramos MI, Cubillas JJ, Jurado JM, et al. Prediction of the increase in health services demand based on the analysis of reasons of calls received by a customer relationship management. *Int J Health Plann Mgmt*. 2019;34:e1215–e1222. https://doi.org/10.1002/hpm.2763