Application Areas of Predictive Analytics in Healthcare

Full Paper

David Rueckel
Johannes Kepler University Linz
Department of Business Informatics
Information Engineering
david.rueckel@jku.at

Stefan Koch
Johannes Kepler University Linz
Department of Business Informatics
Information Engineering
stefan.koch@jku.at

Abstract

Applying predictive analytics on large datasets to predict future behaviour is common in finance and industry. This enables companies to act rather proactive than reactive. Adapting such procedures for healthcare systems can hold potential for both the effectiveness of clinical outcome and healthcare management. In order to identify suitable applications an empirical study following a qualitative approach was conducted. First abstract application areas were derived from literature (e.g. prediction of cancer). Then prospective scenarios for the use of predictive analytics were derived from qualitative data, as e.g. prediction of bowel cancer. These rather specific scenarios were clustered using the abstract application areas. An ABC-Analysis was conducted and application areas were then evaluated in-depth using appropriate criteria. The results indicate a wide rage of possible applications that emphasise the demand for further research.

Keywords: Big Data Analysis, Predictive Analysis, Healthcare, Proactive Medicine, Qualitative Study.

Introduction

Healthcare systems all over the world are facing great challenges. Changes in lifestyle and aging populations both change the demands concerning medical treatments and increase the pressure on the system's future sustainability (Kankanhalli et al. 2016, p. 233). Unfortunately healthcare systems and their core processes are frequently reported to lack efficiency and clinical outcomes, and although appearing to get better consequently, are still suboptimal. One major reason for these inefficiencies is the defective way of gathering, sharing and using healthcare data (Cortada et al. 2012, p. 2f). Ironically, never before the quantity of digitalised health and patient data has been bigger, especially as the realisation, application and integration of electronic health records are strategic goals in most healthcare systems' strategies. So, even though healthcare data is collected, stored and partially linked, the benefit is limited concerning cost and benefit. There is a substantial need for effective methods, tools and techniques to process and make use of this data (Tan et al. 2015, p. 546), respectively embed the usage into the operational processes. Making use of this data can also support medicine and healthcare in turning from a rather reactive to a more proactive field. Diseases shall not only be treated if they occur, but also predicted (Winters-Miner et al. 2015). If the occurrence or the course of a disease can be predicted, the effectiveness of both clinical outcome and healthcare management can be increased. Current adoption still often suffers from resistance from patients and / or healthcare providers, as well as broad strategies with unclear goals respectively benefits. All too often initiatives aim for an all-or-nothing approach, and lack clearly defined and evaluated scenarios

The aim of this study is help in this situation by identifying and evaluating scenarios for the application of predictive analytics (PA). PA comprises methods and techniques that allow the analysis of data in order to make predictions for the future, in healthcare. Therefore scenarios are derived both from literature and from qualitative interviews with medical doctors, nurses, hospital administration and hospital IT professionals. An ABC-analysis was conducted in order to classify the scenarios according to the value provided. The scenarios were finally evaluated using criteria again derived from literature.

The paper is organized as follows. Relevant literature concerning big data analytics in general and healthcare predictive analytics in particular is reviewed and research questions are presented. Next, the methodology is described and the key results are presented and discussed. Concluding, future avenues for research are identified.

Literature Review

PA methods are commonly used in finance and industry for years. PA is not a specific method, but can be seen as a set of statistical methods and techniques that combine and analyse transactional, current and historical data in order to make predictions using statistical patterns. As such patterns show both a current and historical significant relation between items, they are used to predict future probabilities of outcomes. Even though PA is not a new concept, it is an emerging discipline that brings together several scientific disciplines (Statistics, Applied Mathematics, Marketing, Business Strategy, Operations Research, Computer Science, Accounting, Finance). The current emerging interest in PA roots in the proposed potential seen in the phenomenon 'Big Data' and the proposed benefits in the application of PA on unstructured or non-metric datasets (Ogunleye 2014, p. 88f; Siegel 2016, p. 14f).

From a disciplinary point of view PA can be seen as a subfield or subprocess of Data Mining (DM). In contrast to PA, DM focuses on the analysis of data by identifying hidden data on large datasets. Applications like "… customers who bought product A also bought product B …" provide considerable potential for companies to focus on their most important data to benefit from (economically). PA uses results of such DM analysis as its basis. The aim is to extend the basis with prediction using a predictor variable simulating the behaviour of an individual entity. By manipulating this variable, future behaviour, e.g. of a market actor, can be predicted. When it comes to the analysis and prediction of more complex systems multiple predictors are combined to set up a predictive model. The usage of such predictive models leads to knowledge based predictive decision-making (Meka et al. 2015, p. 2229f). Simulating the future behaviour by analysing and imitating past behaviour literally enables an information system to predict the future behaviour of individuals. Idealized each and every knowledge based decision relies on facts and evidence and provides outcomes in real time. The better the data is, the better the prediction will be. In this sense data quality is crucial for the quality of the prediction. Besides collecting data in huge quantities this is a firm call to care for high quality data (Ogunleye 2014, p. 89; Siegel 2016, p. 27f).

The application of DM is quite common in companies with a focus on financial performance, e.g. in order to optimise their sales management to be more efficient within their current processes. PA is the logical consequence to establish efficiency in the future, both in processes and decision-making. The aim to enable better decisions relying on historic data and evidence in future is a common demand both in healthcare in general and in hospital management in particular. The claim for automated tools in healthcare dates back to the 1960s. Triggered by the growth of medical knowledge and the shortage of physicians the demand of clinical decision support systems (CDSS) grew and still is growing. Not surprisingly the basic requirements of such systems have not changed. They are often summarized as, (1) to support data gathering processes and (2) to support decision makers in healthcare in the decision process (Goertzel et al. 1969, p. 689). As a consequence of the need for financial efficiency CDSS is also claimed to decrease and control the costs of the healthcare system by optimizing both managerial and clinical decisions. The costs of healthcare systems have reached an alarming level and are still increasing, due to rising average life expectancy and the change of lifestyle.

As most healthcare systems are partially or fully financed through public funds, an increase in efficiency seems to be the proper way to reduce the financial pressure on the systems. From a clinical point of view an increase in clinical efficiency is not only endeavoured to save money but also to increase the quality of a patient's treatment and the treatment process itself. In the course of time, healthcare changed from a personal relationship treatment to a process driven treatment. Process driven treatment is less personal, and designed to deliver treatment in a scalable manner. In reference to this fact, physicians know less about a certain patient in contrast to the past. There is a lack of knowledge about the patient, and the patient's clinical history (Dhar 2014, p. 115). Results from PA will not be able to inform the physician about a certain patient but about a patient alike the current. Kawamato et al. (2005, p. 1) summarise the preconditions for a successful application of CDSS in healthcare: (1) “automatic provision of decision support as part of clinicial workflow”, (2) “provision of recommendations rather than just assessments”, (3) “provision of decision support at the time and location of decision making” and (4) “computer based decision support”. Summarising this very compact literature review, authors agree that PA offers potential for the improvement of efficiency in healthcare, especially if predictions can prevent illnesses or optimise managerial/administrative processes. In order to evaluate the potential of the application of PA in healthcare, the following research questions are to be answered:

Research Question 1: Which scenarios for the application of predictive analytics within the healthcare system can be identified?

Research Question 2: What is the expected potential and acceptance of these identified scenarios?

Methodology

As the literature shows, authors agree on the potential and the quick wins of PA in healthcare. What seems to be of further interest are the fields of application and the potential value within these fields of healthcare. As the fields of application of PA in healthcare in general and the association of applications to certain scenarios in particular are not yet clearly defined, a qualitative approach was chosen to answer the research questions. This paper is both exploratory and explanatory in nature. The research process was organised as follows:

(1) Application areas (AA) for PA in healthcare were derived from literature. Appropriate search terms were systematically used on the most important literature databases including google scholar. The gathered AA (e.g. “prediction of the risk of cancer”) were analysed and organized into a distinct framework inductively. The interim result was a set of 13 rather abstract and distinct AA for PA in healthcare (see Bates et al. 2014, Bradley and Kaplan 2010, Hoogendorn et al. 2014, , Kobayashi 2014, Maciejewski et al. 2011, Peck et al. 2014, Polykalas et al. 2013, Raghupathi and Raghupathi 2014, Raphael 2013, Sahasrabudhe et al. 2013, Simpao et al. 2014, Thornten et al. 2013, Ward et al. 2014, Winters-Miner 2015).

(2) Qualitative face-to-face interviews with 11 individuals representing physicians (5), nursing staff (3), hospital administration (2) and hospital IT (1) were conducted. The participants were selected due to their technical affinity and IT knowledge. The interviewers used a standardised interview guideline with open-ended questions. After a short introduction into PA in healthcare to achieve a common sense of the topic, the participants were asked for specific and tangible scenarios (e.g. “prediction of the risk of bowel cancer”) for PA. The interviews were transcribed and a qualitative content analysis was applied. As a result of that, 77 tangible scenarios were identified.

(3) Each of the 77 tangible scenarios gathered by conducting the interviews was associated with the AA identified in step (1). If a mentioned scenario was too abstract to be associated with a particular AA, the scenario was added to the AA scheme. As a result of that, the number of the AA was raised to 14 at the end.

(4) An ABC-analysis was applied to classify the AA according to the number of scenarios associated with them in the interviews. The frequency of scenarios within each specific AA served as the basis for this analysis.

(5) An in-depth analysis of the top-rated AA was conducted, using multiple criteria (again derived from literature) to evaluate their potential.

Results and Discussion

As a result of step (1) and complemented through step (3) of the research process the following 14 AA for PA in healthcare are identified. The following enumeration sums up the interim results and illustrates the rather abstract AA along with some example scenarios to make them more tangible and to make the distinct classes transparent.

(1) The most intuitive AA is the optimisation of clinical treatments by analysing historical data and predicting future behaviour, e.g. long term effects of chemotherapy.

(2) The course of a disease can be predicted using PA varying different predictors representing variables that influence the course, e.g. the use of geographical information systems.

(3) The effects of a medication can be predicted, especially in case of interdependencies or allergies, e.g. in case of new medicines.

(4) The optimisation of hospital staff's scheduling, e.g. in timely planning the case of a wave of influenza.

(5) Research in medicine as a multidimensional AA for PA.

(6) The risk of cancer can be predicted along special indicators, such as gen mutations in case of bowel cancer.

(7) The optimisation process of bed occupancy is one of the major benchmarks when evaluating a hospital's efficiency.

(8) Valid and evident (clinical) indicators support the quality of hospital's business intelligence (BI), e.g. for the use in (managerial) scorecards.

(9) The spread of an epidemic can be predicted by simulating different risks of infection or incubation times.

(10) The process and efficiency of a hospital's medication ordering can be optimised by predicting varying demands.

(11) A reduction of costs can be achieved by predicting sickness rates and thereby reducing insurance costs.

(12) The demand (management) of organ donation can be supported by PA in order to reduce cases of transplant rejection.

(13) The misuse/abuse of the healthcare system can be partially prevented by simulating different factors that led to misuse in the past.

(14) The load of hospital IT, e.g. the hospital information system, under specific conditions can be predicted.

ABC Analysis

Table 1 shows the results of the ABC-analysis classifying the AA regarding their frequency of scenarios turning up in the qualitative interviews. As Table 1 shows, almost 60 percent (57.14) of the scenarios belong to the three AA "the optimisation of treatments", "the course of a disease" and "the effects of medication" that form class A. The vast majority (31.17 percent) of scenarios mentioned are in AA "the optimisation of treatments". On the one hand this seems hardly surprising due to the fact that the aim of medicine is curing illnesses or injuries. On the other hand it is worth mentioning because literature draws a different picture by insisting on the increase of efficiency as the aim of PA. Both "the course of a disease" and "the effects of medication" are represented by 10 scenarios (12.99 percent). These AA also focus on the patients and their treatment. Class A, which represents all AA that consist of ten or more scenarios, therefore refers merely to healthcare core processes.

Number	Application Areas	Absolute Frequency	Relative Frequency	ABC
1	Optimisation of treatments	24	31.17	A
2	Course of a disease	10	12.99	A
3	Effects of a medication	10	12.99	A
4	Optimisation of hospital staff	8	10.39	B
5	Research in medicine	6	7.79	B

6	Risk of cancer	5	6.49	B
7	Optimisation of bed occupancy	4	5.19	B
8	Optimisation of hospitals' BI	2	2.6	C
9	Spread of an epidemic	2	2.6	C
10	Optimisation of medication ordering	2	2.6	C
11	Reduction of costs	1	1.3	C
12	Demand of organ donation	1	1.3	C
13	Misuse of the healthcare system	1	1.3	C
14	Load of hospital IT	1	1.3	C

Table 1. Results of the ABC Analysis

Class B, which summarizes all AA that feature more than two scenarios, consists of another four AA that total 30 percent (29.87). Both AA "the optimisation of hospital staff" including 8 scenarios (10.39 percent) and AA "the optimisation of bed occupancy" (4 scenarios, 5.19 percent) focus on rather managerial or administrative tasks and efficiency. AA "research in medicine" covers six scenarios (7.79 percent), which all deal with rather clinical than managerial research attempts. AA "the risk of cancer" includes five scenarios and stresses the importance of focussing on that single disease. Class C covering all AA that consist of two or less scenarios comprises 7 AA presenting 10 scenarios. Seeing Table 1 from a different angle by dividing all scenarios into two classes, one representing rather clinical tasks another one representing rather managerial / administrative tasks, shows an unambiguous result: 75.32 percent of all scenarios are clinical in nature and just 24.68 percent focus on the efficiency of non-clinical processes.

In-Depth Analysis and Evaluation

AA building classes A and B are further analysed in-depth by checking them against evaluation criteria derived from literature (Amarasingham 2014, Rama et al. 2015, Ogunleye 2014, Winters-Miner 2015). The criteria are,

(1) benefit,

(2) frequency,

(3) feasibility,

(4) quality of data and

(5) trust.

Benefit describes the advantages or value that can be gathered and frequency describes how frequent a certain scenario would take place. Feasibility means the capability (probability) of a certain scenario being realised. The criterion "quality of data" refers to the relevance of data quality for the successful application of PA within the certain scenario. The criterion "trust" describes the intensity of trusting in predictions in this scenario. All participants of the study were asked to evaluate the scenarios they mentioned on a four-point scale. Figure 1 shows the results of the evaluation of A and B classified AA. Feasibility, benefit and frequency of all AA are considered high (> 3), which confirms the results of the ABC-analysis. The impact of data quality on the AA varies more and rates very high in case of research and optimising bed occupancy.

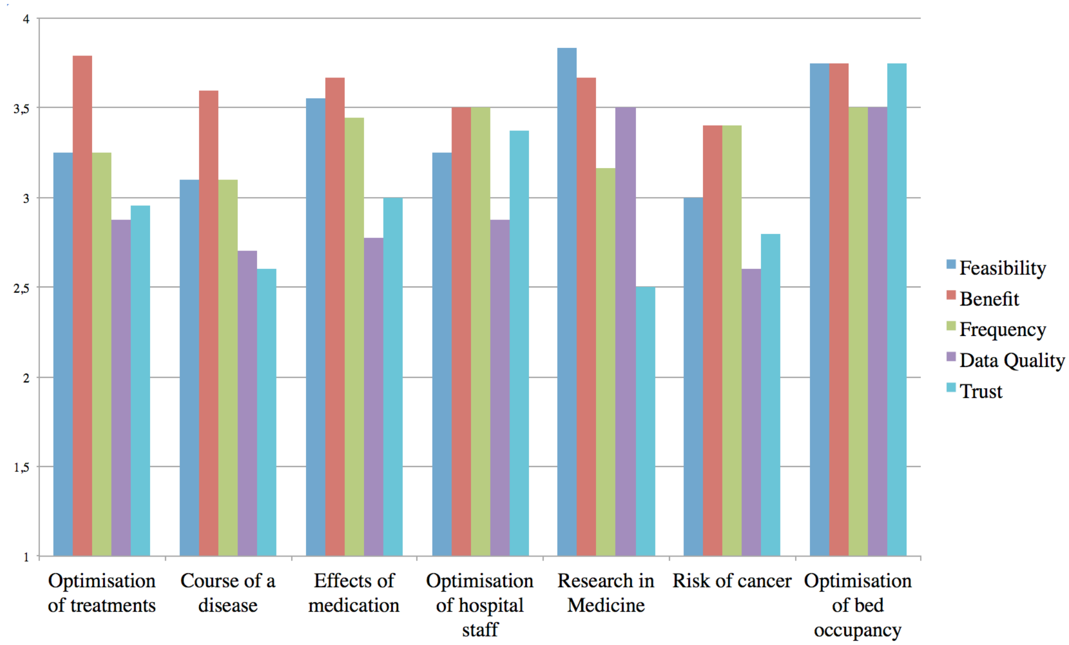

Figure 1. Results of the Evaluation

This supports the notion that the quality of data is crucial in precise applications. Interestingly the importance of data quality is considered pretty low in the prevention of a course of a disease or cancer risk. According to the qualitative data this can be explained with a complex cause-effect chain. The criterion trust ranges the most. While trust is pretty low in research and the prediction of a course of a disease, which again relies on complex cause-effect chains, it is considerable high in AA that are manageable more easily, such as optimizing bed occupancies.

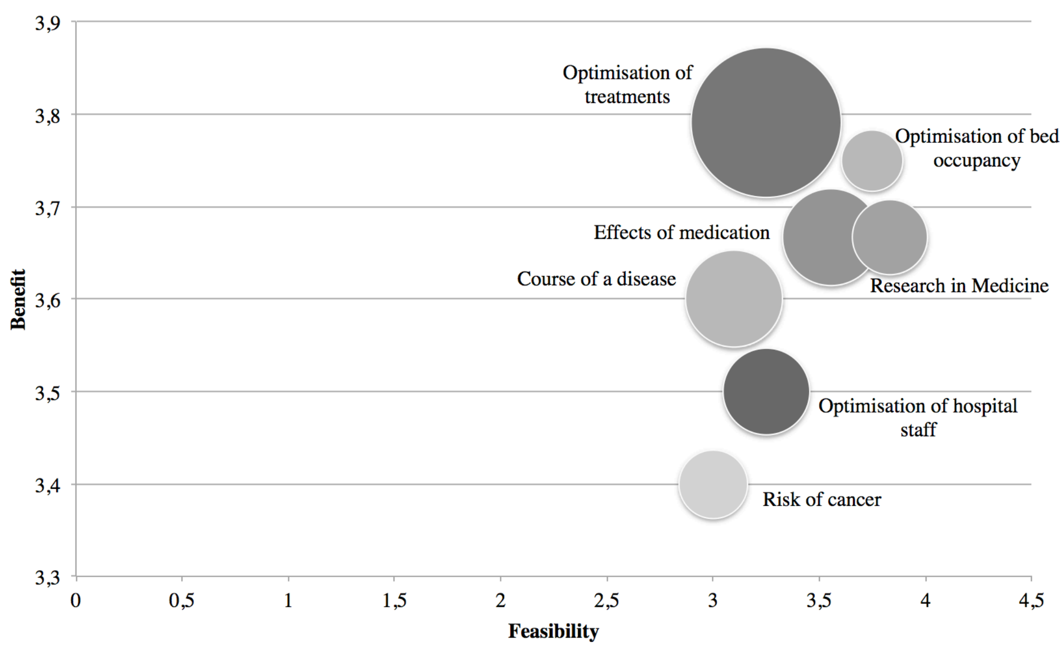

Figure 2. Results of the Portfolio Analysis

Figure 2 combines the results of the ABC-analysis and the evaluation. It shows “feasibility” and “benefit” as the axes of the system, as they seem more decisive from a managerial point of view. The size of the bubbles represents the number of scenarios within the AA. The optimisation of treatments would doubtlessly cause the largest benefit, both for clinical and managerial efficiency. This AA also features the most scenarios by far. Interestingly the feasibility is considered comparably low in contrast to research in medicine or the optimisation of bed occupancy. This again can be explained by the complex structure of the scenarios. The optimisation of bed occupancy shows comparable benefits but seems more feasible. This may result from the clearly defined scenarios in this AA. Both research in medicine and the effects of medication are feasible and can provide benefits. The number of scenarios (and trust) is astonishingly low in research, which might have its cause in the case based history of medical research.

Conclusion and Future Research

As these preliminary results show, PA doubtlessly is perceived as able to provide benefits in healthcare systems. Both in clinical and in managerial or administrative AA a considerable number of scenarios were identified, even though the data pool was limited due to the qualitative research design employed. The number of scenarios associated with clinical tasks was significantly higher than the ones dealing with managerial / administrative tasks. This can be seen as a surprising outcome of this study, and a clear focus on core processes rather than supporting processes in applying new methods. Whether this holds valid on other contexts beyond healthcare would be interesting to study. The AA of classes A and B were all considered feasible, beneficial and frequent concerning their application. When it comes to the impact of data quality, this study shows inconclusive results, but trust seems to clearly be correlated with the complexity of the scenario. This assumed correlation could be investigated using a larger data set gathered from different, maybe quantitative research designs applied on the same type of populations. Additionally, the criteria used in the evaluation could be further refined, e.g. by including complexity. Furthermore the patients’ point of view could be included in future research. It would be of great interest which scenarios are considered beneficial by the people constituting the target of the application and – in any manner - paying for the application of PA in healthcare. A possible limitation of the study finally results from the country the study was conducted in. Healthcare systems with a different type of financing could lead to divergent results, as e.g. efficiency could be interpreted differently by participants.

REFERENCES

Amarasingham, R., Patzer, R. E., Huesch, M., Nguyen, N. Q., Xie, B. 2014. “Implementing Electronic Health Care Predictive Analytics: Considerations And Challenges,” *Health Affairs* (33:7), pp. 1148-1154.

Bates, D. W., Saria, S., Ohno-Machado, L., Shah, A., Escobar, G. 2014. "Big data in health care: using analytics to identify and manage high-risk and high-cost patients." *Health Affairs* (33:7) , pp. 1123-1131.

Bradley, P., Kaplan, J. 2010. "Turning hospital data into dollars: healthcare financial executives can use predictive analytics to enhance their ability to capture charges and identify underpayments." Healthcare Financial Management (64.2) ,pp. 64-69.

Cortada, J. W., Gordon, D., Lenihan, B. 2012. *The value of analytics in healthcare: From insights to outcomes*, IBM Global Business Services, Executive Report.

Dhar, V. 2014. “Big Data and Predictive Analytics in Health Care,” *Big Data* (2:3), pp. 113-116.

G. Goertzel, 1969. “Clinical decision support system,” *Annals of the New York Academy of Sciences* (161:2), pp. 689–693.

Hoogendorn, M., Moons, L. M. G., Numans, M. E., Sips, R.-J. 2014. “Utilizing Data Mining for Predictive Modeling of Colorectal Cancer Using Electronic Medical Records,” in *Proceedings of the Brain Infomatics and Health, International Conference*, Slezak, D., Tan, A.-H., Peters, J. F., Schwabe, L. (eds.), Warshaw, Poland, pp. 132-141.

Kankanhalli, A., Hahn, J., Tan, S., Gao, G. 2016. “Big data and analytics in healthcare: Introduction to the special section,” *Information Systems Frontiers* (18:2), pp. 233-235.

Kawamoto, K., Houlihan, C. A., Balas, E. A., Lobach, D. F. 2005. "Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success," *BMJ* 330(7494): 765.

Maciejewski, R., Hafen, R., Rudolph, S., Lares, S. G., Mitchell, M. A., Cleveland, W. S., Ebert, D.S. 2011 "Forecasting hotspots—A predictive analytics approach." *IEEE Transactions on Visualization and Computer Graphics*.(17.4), pp. 440-453.

Meka, R. C. R., Shariff, N., Patil, A. 2015. "Performing Predictive Data Analytics in Data Mining Using Various Tools," *International Journal of Innovative Technology and Research*(3:4), pp. 2229-2233.

Ogunleye, J. 2014. "The Concepts of Predictive Analytics," *International Journal of Developments in Big Data and Analytics* (1:1), pp 86-94.

Peck, J. S., Benneyan, J. C., Nightingale, D. J., Gaehde, S. A. 2014. "Characterizing the value of predictive analytics in facilitating hospital patient flow," *IIE Transactions on Healthcare Systems Engineering* (4), pp. 135-143.

Polykalas S. E., Prezerakos G. N., Konidaris A. 2013 „An algorithm based on Google Trends' data for future prediction. Case study: German elections.," In *Proceedings of IEEE International Symposium on Signal Processing and Information Technology*, pp. 69-73.

Raghupathi, W., Raghupathi, V. 2014. "Big data analytics in healthcare: promise and potential." *Health Information Science and Systems* (2:1), pp. 1-10.

Raphael, H. 2013. *Business Intelligence im Krankenhaus*. Berlin: Springer.

Sahasrabudhe, M., Panwar, M., Chaudhari, S. 2013. "Application performance monitoring and prediction," *Proceedings of the IEEE International Conference on Signal Processing, Computing and Control*, Solan, pp. 1-6.

Siegel, E. 2016. Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die. Hoboken, NJ: Wiley.

Simpao, A. F., Ahumada, L. M., Galvez, J. A., Rehman, M. A. 2014. "A review of analytics and clinical informatics in health care." *Journal of Medical Systems* (38.4), pp. 1-7.

Tan, S. S.-L., Gao, G., Koch, S. 2015. "Big Data and Analytics in Healthcare," *Methods of Information in Medicine* (54:6), pp. 546-547.

Thornton, D., Mueller, R. M., Schoutsen, P. "Predicting Healthcare Fraud in Medicaid: A Multidimensional Data Model and Analysis Techniques for Fraud Detection," *Proceedings of the CENTERIS Conference on ENTERprise Information Systems / ProjMAN 2013 - International Conference on Project MANagement/ HCIST 2013 - International Conference on Health and Social Care Information Systems and Technologies*. Cruz-Cunha, M. M., Varajão, J., Krcmar, H., Martinho, R. (eds.), pp. 1252 – 1264.

Ward, M. J., Marsolo, K. A., Froehle, C. M. 2014. "Applications of business analytics in healthcare." *Business horizons* (57:5), pp. 571-582.

Winters-Miner, L. A., Bolding, P. S., Hilbe, J. M., Goldstein, M., Hill, T., Nisbet, R., Walton, N., Miner, G. D. 2015. *Practical Predictive Analysis and Decisioning Systems for Medicine*. Amsterdam: Elsevier.

Yang, C.-G., Lee, H.-J. 2016. "A study on the antecedents of healthcare information protection intention," *Information Systems Frontiers* (18:2), pp. 253-263.