Prediction of Propensity for Enterprise Cloud Computing Adoption

Niki Kyriakou
University of the Aegean
nkyr@aegean.gr

Manolis Maragoudakis
University of the Aegean
mmarag@aegean.gr

Euripides Loukis
University of the Aegean
eloukis@aegean.gr

Marinos Themistocleous
University of Piraeus
mthemist@unipi.gr

Abstract

Cloud computing (CC) can offer significant benefits to enterprises. However, it can pose some risks as well, and this has led to lower adoption than the initial expectations. For this reason, it would be very useful to develop 'predictive analytics' in this area, enabling us to predict which enterprises will exhibit a propensity for CC adoption. In this direction, we investigate the use of six well-established classifiers (fast large margin Support Vector Machine, Naive Bayes, Decision Tree, Random Forest, k-Nearest Neighbor, and Linear Regression) for the prediction of enterprise level propensity for CC adoption. Having as our theoretical foundation the Technology – Organization – Environment (TOE) framework, we are using for this prediction of set of technological (concerning existing enterprise information systems), organizational and environmental characteristics. Our first results, using a dataset collected from 676 manufacturing firms of the glass, ceramic and cement sectors from six European countries (Germany, France, Italy, Poland, Spain, and UK) through the e-Business W@tch Survey of the European Commission, are encouraging. It is concluded that among the examined characteristics the technological ones, concerning the existing enterprise systems, seem to be the most important predictors.

Keywords: ERP systems, Cloud Computing, Adoption

1. Introduction

Cloud Computing (CC) is a new paradigm of enterprise computing, which can provide significant benefits: reduction of the costs of ICT support of enterprise activities, conversion of related capital investments to operating costs, rapid and low-cost development of technological support required for process, product and service innovations, ubiquitous access, provision of flexible cost-effective computing capacity for supporting growth, rapid and low-cost access to new technologies (e.g. business analytics, mobile interactive applications) and scalability [2, 8, 19, 21, 27]. However, it is widely recognized that CC can pose some risks as well, such as performance related risks and data security risks (the latter associated mainly with firm's data integrity and confidentiality) [1, 4, 27]. For the above reasons, the adoption of CC by firms has been lower than the initial expectations [11, 12, 23, 25].

So it would be very useful to develop 'predictive analytics' [6] in this area, enabling us to predict which enterprises will exhibit a propensity for CC adoption. This is going to be highly beneficial for CC services providers, and all other CC stakeholders as well, as it will allow them to focus their marketing efforts, personnel and budgets on enterprises having high levels of propensity for CC adoption, instead of wasting valuable human and financial resources for unfocused marketing activities targeting too big numbers of firms. Furthermore, it will enable CC service providers to optimize their offerings in order to be more beneficial and less risky for the kind of enterprises predicted to be interested in CC adoption. Also, by examining the most important predictors of CC adoption propensity (such as various technological, organizational and environmental enterprise characteristics affecting it), we can get interesting insights as to the main CC adoption motivations and orientations of enterprises.

In this direction, we investigate the use of six well-established classifiers (fast large margin Support Vector Machine, Naive Bayes, Decision Tree, Random Forest, k-Nearest Neighbor and Linear Regression) for the prediction of enterprise level propensity for CC adoption. Having as our theoretical foundation the Technology – Organization – Environment (TOE) framework [3, 26], we are using for this purpose of set of technological (concerning existing enterprise information systems), organizational and environmental enterprise characteristics as possible predictors. The above characteristics of an enterprise affect the magnitude of both the benefits it can get from CC, and the risks that CC poses to it, so we expect that they will finally affect enterprise's propensity to adopt CC (as explained in more detail in section 3); some first empirical evidence for this is provided in [15]. The above classifiers have been built, tested and compared using a dataset collected from 676 manufacturing firms

HICSS

of the glass, ceramic and cement sectors from six European countries (Germany, France, Italy, Poland, Spain, and UK) through the e-Business W@tch Survey of the European Commission.

This paper is structured in five sections. In the following Section 2, the background of this study is outlined. Then in Section 3 our data and method are described, while the results are presented in Section 4. Finally, in Section 5, the conclusions are summarized.

2. Background

The lower adoption of CC in comparison with the initial expectations [11, 12, 23, 25], which has been mentioned in the Introduction, has motivated considerable empirical research for the identification of factors affecting the adoption of CC by enterprises. Most of this research is based on the Technology, Organization and Environment (TOE) framework [3, 26], usually in combination with the Diffusion of Innovation (DOI) theory [24]. According to the TOE theory the adoption of technological innovations by enterprises is influenced by three groups of factors: technological (associated with perceived characteristics of the specific technological innovation, and also the technologies currently in use by the enterprise), organizational (characteristics of the enterprise) and environmental (characteristics of its external environment). The DOI theory, which is used for elaborating the technological factors, defines five critical characteristics of an innovation that determine the degree of its adoption: relative advantage, compatibility, complexity, trialability, and observability.

Based on the above theoretical foundations [17] examines the effect of a set of technological factors concerning the CC technology itself (relative advantage, complexity and compatibility), organizational factors (top management support, size and technology readiness) and environmental factors (competitive pressure and trading partner pressure) on CC adoption, using data from a sample of 111 enterprises in the high-tech industry in Taiwan. They conclude that perceived relative advantage, top management support, size, competitive pressure and trading partner pressure affect positively CC adoption. Another TOE-based study is presented in [11], examining the effect of perceived benefits and business concerns (technological factors), IT capability (organizational factor) and external pressure (environmental factor) on CC adoption intention, based on data from 200 Taiwanese firms. It concludes that the first three of these factors are significant determinants of CC adoption, while the fourth is not. [18] focuses on Software as a Service (SaaS), and examine the effects of a set of technological

factors concerning CC itself (relative advantage, compatibility, complexity, trialability, observability), organizational factors (organizational readiness, top management support) and environmental context (market pressure, market competition, vendor marketing, trust in vendor, government support) on SaaS adoption. Using data from 147 Indonesian enterprises, they estimate a logistic regression model, which leads to the conclusion that compatibility, observability, market competition and government support have a positive effect on SaaS adoption, while complexity has a negative effect. [23] examines the effects of three CC characteristics from a technological innovation perspective (relative advantage, complexity and compatibility), three organizational context characteristics (top management support, size, technological readiness) and two environmental context characteristics (competitive pressure, regulatory protection), using data from 369 Portuguese manufacturing and services enterprises. They found that relative advantage, technological readiness, top management support, and size had positive effects on CC adoption, while complexity has a negative effect. Another similar study has been conducted in [9], which examines the effects of a set of technological factors concerning the CC technology (relative advantage, complexity and compatibility), organizational factors (top management support, size, technological readiness) and environmental factors (competitive pressure, trading partners pressure) on CC adoption, using data collected from 257 mid-to-senior level decision-making business and ICT professionals from UK enterprises. They concluded that competitive pressure, complexity, technology readiness and trading partner pressure have a significant influence on the adoption of CC services.

Since TOE has been the main theoretical foundation of most previous empirical research on CC adoption factors, we have used it as our theoretical foundation for the development of prediction models of CC adoption propensity, having as predictors: a) technological characteristics (but not DOI-based perceived characteristics of the CC, as in the previous CC adoption research; we explore characteristics of the technologies currently in use by the enterprise, such as various types of enterprise systems, which affect CC perceived adoption benefits and propensity in general); b) organizational characteristics; c) environmental characteristics (but for both these categories exploring a wider set of characteristics than the limited ones investigated by previous CC adoption empirical literature).

3. Data and Method

3.1 Data

In this study for the development, test and comparison of CC adoption propensity prediction models we have used data collected through the "e-Business Survey", which has been conducted by the e-Business Market W@tch, under the auspices of the European Commission. In this survey have been collected various ICT and e-business use related data, as well as background information, from a sample of 676 enterprises of the glass, ceramic, and cement sectors, from six European countries (Germany, France, Italy, Poland, Spain, and UK). 53.8% of the sample firms were small (with 1-49 employees), 33.6% were medium (with 50-249 employees), and the remaining 12.6% were large firms (with more than 250 employees).

3.2 Predictors

Using these data, we attempted to build prediction models for the enterprise level propensity to adopt CC. It has been initially measured in a three levels scale (corresponding to enterprise perception of the CC as very relevant, partly relevant or not relevant), but has been recoded as a binary attribute, with very relevant or partly relevant coded as 0, and not relevant coded as 1. We have examined 28 enterprise characteristics as features - possible predictors of CC adoption propensity (details about the corresponding questions of the "e-Business Survey" questionnaire, which have used in this study as the features, are shown in Appendix I):

i) Twelve technological characteristics, which concern the use of ten important types of enterprise systems (ERP, SCM, CRM, SRM, CAD and CAM, electronic customers ordering, electronic invoicing, electronic collaboration with other firms for product/service or process innovations, telework), and also the interest in two new emerging technologies (data warehouses/data mining, and mobile services). According to the previous literature [19, 21, 27], CC can be a low-cost way for enterprises not currently using the above important types of enterprise systems to introduce them without having to make big investments; so based on this line of arguments we would expect these variables to be predictors of CC adoption propensity having negative impact on it. However, in the same literature there are also arguments in the opposite direction: enterprises having extensive ICT infrastructure might exhibit high CC adoption propensity in order to reduce their high relevant costs (e.g. operations, support, maintenance and upgrade costs), e.g. by using Infrastructure as a Service (IaaS)

and Platform as a Service (PaaS) services for hosting some of their applications, or even by using Software as a Service (SaaS) services for replacing some older and/or bespoke applications with more modern standard software packages. Based on this latter line of arguments we would expect these variables to be predictors of CC adoption propensity, but having a positive impact on it. With respect to the two variables concerning firm's interest in data warehouses/data mining and mobile technologies, we expect them to be important predictors of CC adoption propensity, having positive impact on it; previous relevant literature emphasizes as one of the main benefits of CC is the opportunities it provides for the experimentation with and adoption of new emerging ICT, at low cost, and without having to make big and risky investments for them [19, 21, 22, 27, 28].

ii) Eleven organizational characteristics, which concern the sector and size of the enterprise, its innovation activity (innovations in its products/services, and also processes, and external innovation collaboration), its operational complexity (having multiple production locations, and also wide geographical scope of sales and procurement); furthermore, some ICT management characteristics (employment of ICT personnel, ICT outsourcing, ICT investment reduction due to the economic crisis). We expect enterprise innovation activity, and also external innovation collaboration, to be important predictors, affecting positively the propensity for CC adoption, as previous literature has emphasized the huge capabilities provided by CC for the low-cost electronic support of them [16, 27, 21]. Also, we expect operational complexity, as it increases ICT support needs, to be a predictor as well, affecting positively CC adoption propensity. The same holds for ICT investment reduction strategy, as it reduces financial resources available for enhancing existing enterprise systems infrastructure, making CC an attractive alternative. Finally, ICT personnel and previous ICT out-sourcing experience are expected to be positive predictors as well, as they increase the relevant knowledge base of the enterprise, which allows better analysis of existing CC offerings, selection of the most appropriate ones for the enterprise, and also management of relationship(s) with selected CC provider(s).

iii) Five environmental factors, which concern various types of competition the enterprise experiences (with respect to price, product quality, customer service, image and product design competition, and also 'e-competition' generated through the Internet e-business channel). All these types of competition put strong pressures on the enterprise, and increase the needs for ICT support, but without much increase of operating costs, making CC quite attractive; so we expect them to be predictors of CC adoption propensity, having a

positive impact on it.

3.3 Prediction Models

In order to build prediction models for CC adoption propensity we have used a variety of Machine Learning (ML) methods [10, 14, 20, 29]. In this era of data explosion, in which enterprises collect and disseminate vast volumes of data, ML plays an important role for highly sophisticated analysis of such data, aiming to extract useful knowledge or patterns from it, and also use them for making predictions; the latter has given rise to the development of predictive analytics [6]. In particular, an important stream of ML is dealing with classification, namely the process of modeling annotated data, in which a feature is selected to represent the class variable of the problem (dependent variable), and a set of other features are used as possible predictors of it (independent variables); the extracted model can be used in order to accurately predict the class of a new, previously unseen instance.

For constructing the particular method, we used for building our CC adoption propensity prediction models we had to address an important problem: the unbalanced nature of our dataset with respect to CC adoption propensity, as the enterprises having no propensity to adopt CC were about seven times more frequent than the enterprises having such a propensity. In such cases, all classification algorithms tend to favor the most frequent class. A good solution is the creation of more balanced subsets of the original dataset, where examples are selected so that we have equal classes (or at least less unbalanced), thus helping classification algorithms to avoid over-fitting towards the majority class.

In particular, the method we used for building our CC adoption propensity prediction models consists of seven phases, which are shown in Figure 1. In the first phase, the original dataset is being read and loaded. In phase two we cleanse our dataset from records that do not represent a valid entry. More specifically, we remove records that either do not include the majority of the independent variable (features) or do not have a class label (value of dependent variable, concerning the existence or not of CC adoption propensity). However, in this case, none of our data points fell in any of the aforementioned categories and thus we did not need to reduce the original dataset. During the third phase we normalize the dataset by replacing nil or empty values (some of the questions may not be answered) with zeroes, for compatibility with our classifiers.

In the fourth phase we calculate the information gains of the predictors [10, 14, 20], which are used in order to generate their weights and rank the features in our feature vectors. The feature selection takes place in the fifth phase. We chose a weight threshold so that each attribute with a weight less than the threshold is dropped from the dataset, and only the attributes with a weight above the threshold are considered for the next phase. In other words, we calculate the weight of attributes with respect to our class attribute. The higher the weight of an attribute, the more relevant it is considered. The feature ranking details can be found in Appendix II.
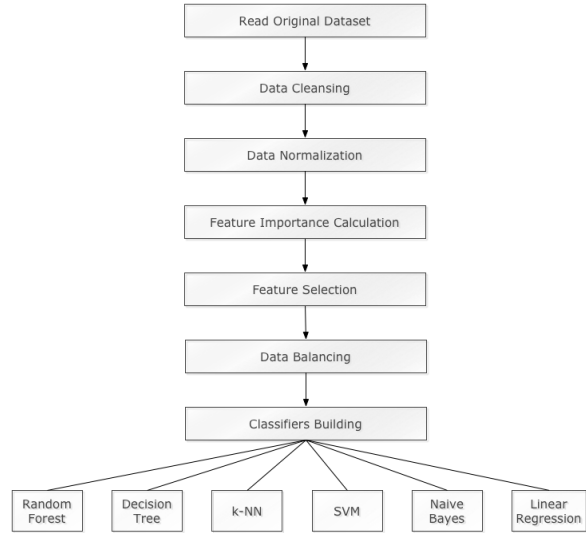

Figure 1: Phases of Models Building Method

Since our dataset, as mentioned above, is unbalanced with respect to the propensity to adopt CC in the sixth phase we had to balance it. Only 84 out of the 676 sample enterprises (12.5%) have a propensity to adopt CC, while the remaining 592 do not (87.5%). Thus, in the sixth phase (Data Balancing) we created a sub-sample of our initial sample, by selecting part of the records of the larger class (not having propensity to adopt CC) randomly: so this sub-sample included all the enterprises that have propensity to adopt CC and only 30% of those that do not. Finally, in the seventh phase from the above data we built, tested and compared six different classifiers: a Fast Large Margin SVM, a Naive Bayes, a Decision Tree, a Random Forest, a k-NN and a Linear Regression [10, 14].

4. Results

4.1. Feature Selection

As mentioned in Section 3 before building and comparing the different classifiers, we select the most significant features based on their information gains. The feature weights are shown in the Table of the Appendix II. By examining them we can gain useful

insights as to the CC adoption motivations and orientations of the enterprises of these three manufacturing sectors (glass, ceramic and cement). We can see that the feature with the maximum weight was the interest in data warehouses/data mining (with a weight of 18.40%), followed by the interest in mobile services (9.20%), and the use of SCM systems (6.53%). These results indicate that the three most important predictors of enterprises CC adoption propensity in these sectors are technological, being associated with existing enterprise systems (use of SCM), and most importantly with future plans for enhancing them by incorporating new emerging technologies (such as data warehouses/data mining and mobile services). Therefore the main motivation of the enterprises of these sector for adopting CC are: i) The experimentation and introduction of these two emerging ICT, at a low cost, without having to make big investments, and also with low levels of risk (if these technologies are not appropriate and beneficial for the enterprise the corresponding CC services can be terminated, without having any loss of investment); ii) the reduction of the operating, support, maintenance and upgrade costs of enterprise systems (such as SCM).

Taking into account the weights shown in the Appendix II these technological predictors are followed by two organizational ones, which however concern ICT as well: negative impact of the economic crisis on ICT investment plans (6.23%), and existence of ICT personnel (5.04), which seem to be important predictors of CC adoption propensity. The above indicate that enterprises of these manufacturing sectors that reduce their ICT investment due to the economic crisis find CC as an attractive alternative for satisfying needs for enterprise systems enhancements (e.g. for increasing the computing power and the functionality of their ICT infrastructures), which are required for improving the electronic support of their processes and activities, without having to make additional ICT investments. Furthermore, the above results indicate that the existence of ICT personnel is an important predictor of CC adoption propensity affecting it positively; this probably reflects the important role that ICT personnel have in CC adoption: in finding high quality CC providers and services, examining their offerings, and their advantages and disadvantages, and finally selecting the most appropriate ones for satisfying the particular needs of the enterprise.

From the environmental characteristics, the most important predictor is the e-competition in the sector, which is generated through the Internet e-business channel (5.04%). This indicates that enterprises experiencing this kind of competition in order to survive need to enhance their ICT infrastructures, and also incorporate in them new emerging technologies, and CC is an attractive way for achieving these.

4.2. Modeling Results

In Table 1 we can see the overall accuracies of the six alternative classifiers we have used. We have performed a k-fold cross-validation of them, and set our k equal to 10, which according to Kohavi et al. (1995) is a generally accepted classifiers' performance testing method. Accuracy has been calculated by taking the percentage of correct predictions over the total number of examples. We can see the Random Forest classifier appears to have the best performance, exhibiting the highest accuracy among all the examined classifiers, at the level of 86.63%.

Table 1: **Classifiers' accurac**ies

	Overall Accuracy
Random Forest	86.63%
Decision Tree	78.99%
Naive Bayes	73.82%
k-NN	81.66%
SVM	79.88%
Linear Regression	81.21%

In Appendix III we can see the Receiving Operating Characteristic (ROC) curves for the above six classifiers. The ROC curve is a graphical plot, which shows the true positive rate (y-axis) versus the false positive rate (x-axis) of a classifier, and illustrates the performance of it [5]. A good ROC curve minimizes the false positive rate (x-axis) while maximizing the true positive rate (y-axis). We see that from these ROC curves it is confirmed that the Random Forest classifier performs the best, but also the Decision Tree classifier also performs well.

Next we examined the performance of the Random Forest classifier per class, which is shown in Table 2. The results indicate that it performs very well in classifying correctly the enterprises not having propensity to adopt CC: for 98.14% of them a correct classification is achieved; also, looking horizontally in Table 2, 88.16% of the enterprises classified as not having propensity to adopt CC are correct predictions.

However, things are different for the other class of the enterprises having propensity to adopt CC: only for 7.14% of them a correct classification is achieved; looking horizontally in Table 2, only 35.29% of the enterprises classified as having propensity to adopt CC are correct predictions. This problem is due to the abovementioned unbalanced nature of our dataset, which includes a much larger number of enterprises not having CC adoption propensity, and a smaller number of enterprises having a propensity to adopt CC (with this dataset reflecting the existing situation in the examined manufacturing sectors); so our classifiers are better trained for correctly classifying the former than the latter. Better results can probably be achieved if the same method/classifiers are applied in a larger and more balanced dataset with sufficient numbers of examples from both classes. However, even using such an unbalanced dataset, the Random Forest classifier can provide a good 'filtering out' of enterprises having no propensity to adopt CC, and this can be quite beneficial for CC services providers, as it allows them to focus their marketing activities, personnel and budgets on the remaining enterprises, which have some propensity and therefore higher likelihood of adopting CC.

Table 2: Random Forest Performance

	True no propensity to adopt CC	True propensity to adopt CC	Class Precision
Predicted no propensity to adopt CC	581	78	88.16%
Predicted propensity to adopt CC	11	6	35.29%
Class Recall	98.14%	7.14%	

5. Conclusions

As the adoption of CC has been so far much lower than the initial high expectations, it would be useful to develop 'predictive analytics' [6] in this area, enabling us to predict which enterprises will exhibit a propensity for CC adoption. Our study makes a contribution in this direction, investigating the use of six well established classifiers (Fast Large Margin Support Vector Machine, Naive Bayes, Decision Tree, Random Forest, k-Nearest Neighbor, and Linear Regression) for predicting enterprise level propensity to adopt CC, based on technological, organizational and environmental

characteristics. The results indicate that the use of a Random Forest classifier enables a highly reliable prediction of enterprises not having propensity to adopt CC, which allows 'filtering out' them, and this can be highly beneficial for CC services providers and other CC stakeholders. It is followed by two 'simpler' classifiers: the k-Nearest Neighbor and the Linear Regression.

Also we have found that the most important predictors of CC adoption propensity are related with technology. Five out of the ten most important predictors of the CC adoption propensity shown in Appendix II concern the use of some types of enterprise systems: SCM, ERP, CAM, electronic collaboration systems and electronic invoicing ones. Also, among the top ten predictors are the interest in incorporating two new emerging technologies in existing enterprise systems: data warehousing/data mining and mobile services. Finally, we have another technology related predictor among the ten top ones, which concerns the adoption of an ICT reduction strategy, resulting in cancellation or significant downsizing of ICT or e-business projects, due to existing economic crisis. Therefore, eight out of the ten top predictors are related with technology. The above findings indicate that enterprises of the examined three manufacturing sectors view CC mainly as a way of reducing the operating, support, maintenance and upgrade costs of their enterprise systems (e.g. by using IaaS and PaaS services for hosting some of their applications, or even by using SaaS services for replacing some older and/or bespoke applications with more modern standard software packages). Furthermore, they view CC as quite useful for the experimentation with and the introduction of new emerging ICT, at a low cost, without having to make big investments, and also with low levels of risk.

This study has interesting implications for research and practice in the area of CC. It opens up interesting research directions for the development of predictive analytics in the area of CC adoption, by exploiting the advances in the area of Machine Learning. Also the proposed prediction methods are highly useful for CC practice. We expect they will be beneficial for CC services providers, and all other CC stakeholders as well, enabling them to focus their marketing efforts, personnel and budgets on enterprises having high levels of propensity for CC adoption, and avoid wasting valuable human and financial resources for unfocused marketing activities. Furthermore, the proposed prediction methods will enable CC service providers to optimize their offerings in order to be more beneficial and less risky for the kind of enterprises predicted to be interested in CC adoption.

The main limitation of this study is that it has been based on data from only three European manufacturing

sectors, which are rather conservative in terms of adoption of new ICT, and innovative business practices in general, and therefore representative of 'traditional' manufacturing [7]. So its findings may have been influenced to some extent by this particular sectoral and national context. Also, it does not distinguish between different categories of CC services (IaaS, PaaS, SaaS). So further research is required concerning the prediction of the propensity to adopt different types of CC services, in various sectoral and national contexts, based on larger and more balanced datasets.

6. References

[1] Ackermann, T., T. Widjaja, A. Benlian, and P. Buxmann, "Perceived IT Security Risks of Cloud Computing: Conceptualization and Scale Development", Proceedings of 33rd International Conference on Information Systems (ICIS), St. Louis, USA, 2012.

[2] Armbrust, M., A. Fox, R. Griffith, A. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica and M. Zaharia, "A View of Cloud Computing", Communications of the ACM 53(4), 2010, pp. 50- 58.

[3] Baker, J., The technology-organization-environment framework, In Dwivedi, Y., M. Wade, and S. Schneberger, (Eds), Information Systems Theory: Explaining and Predicting Our Digital Society, Springer, New York, 2011, pp. 231-246.

[4] Benlian, A. and T. Hess, "Opportunities and Risks of Software-as-a-Service: Findings from a Survey of IT Executives", Decision Support Systems, 52(1), 2011, pp. 232-246.

[5] Brown, C. D. and H. T. Davis, "Receiver operating characteristics curves and related decision measures: A tutorial", Chemometrics and Intelligent Laboratory Systems, 80(1), 2006, pp. 24-38.

[6] Davenport, T. H., Enterprise Analytics. Pearson Education – International Institute of Analytics, USA, 2013.

[7] Empirica GmbH, "ICT and e-Business Impact in the Glass, Ceramics & Cement Industry". European Commission, DG Enterprise & Industry, Bohn/Brussels, 2009.

[8] Etro, F. "The Economic Impact of Cloud Computing on Business Creation, Employment and Output in Europe", Review of Business and Economics, 54(2), 2009, pp. 179-208.

[9] Gutierrez, A., E. Boukrami, and R. Lumsden, "Technological, Organizational and Environmental Factors Influencing Managers' Decision to Adopt Cloud Computing in the UK", Journal of Enterprise Information Management, 28(6), 2015, pp. 788 – 807.

[10] Hastie, T., R. Tibshirani, and J. Friedman, "The Elements of Statistical Learning", Springer Verlag, 2009.

[11] Hsu, P., F. Ray, and S. Li-Hsieh, "Examining Cloud Computing Adoption Intention, Pricing Mechanism and Deployment Model", International Journal of Information Management, 34(4), 2014, pp. 474- 488.

[12] Kung, L., C. G. Cegielski, and H. J. Kung, "An integrated environmental perspective on software as a service adoption in manufacturing and retail firms", Journal of Information Technology, 30(4), 2015, pp. 352-363.

[13] Kohavi, R. et al., "A study of cross-validation and bootstrap for accuracy estimation and model selection", International Joint Conference on Artificial Intelligence, 14, 1995, pp. 1137–1145.

[14] Kotsiantis, S.B., Supervised Machine Learning: A Review of Classification Techniques. In I. Maglogiannis (ed). Emerging Artificial Intelligence Applications in Computer Engineering: Real Word AI Systems with Applications in eHealth, HCI, Information Retrieval and Pervasive Technologies, Ios Press, 2007.

[15] Loukis, E. and N. Kyriakou, Organizational Factors Affecting Propensity to Adopt Cloud Computing, Hawai, January 2015, Grand Hyatt Kauai.

[16] Loukis, E., Kyriakou, N. and Pazalos, K. (2015). Operational and Innovation Collaboration and Cloud Computing, European, Mediterranean & Middle Eastern Conference on Information Systems 2015 (EMCIS2015), Athens, Greece, June 2015.

[17] Low, C. and Y. Chen, "Understanding the determinants of cloud computing adoption", Industrial Management & Data Systems, 111(7), 2011, pp. 1006-1023.

[18] Mangula, I.S., I. Weerd, and S. Brinkkemper, "The Adoption of Software-as-Service: An Indonesian Case Study", Proceedings in Pacific Asia Conference on Information Systems (PACIS), 2014.

[19] Marston, S., Z. Li, S. Brandyopadyay, J. Zhang, and A. Ghalsasi, "Cloud Computing – The Business Perspective", Decision Support Systems, 51(1), 2011, pp. 176-189.

[20] Mitchell, T., "Machine Learning. McGraw-Hill, 1997.

[21] Müller, S. D., S. R. Holm, and J. Søndergaard, "Benefits of Cloud Computing: Literature Review in a Maturity Model Perspective", Communications of the Association for Information Systems, 37, 2015, pp. 851 – 878.

[22] Nimje, A. R., "Data Analytics as a Service (DAaaS): An Arriving Technology in Cloud Computing", International Journal of Emerging Trend in Engineering and Basic Sciences, 2(1), 2015, pp. 181-186.

[23] Oliveira, T., M. Thomas, and M. Espadanal, "Assessing the determinants of cloud computing adoption: An analysis of

the manufacturing and services sectors", Information & Management, 51(5), 2014, pp. 497-510.

[24] Rogers, E., Diffusion of Innovations - Fifth Edition, The Free Press, New York, USA, 2003.

[25] Siepermann, M. A. Sutaj., P. Lübbecke, and R. Lackes, "Refuse to Walk on Clouds - Why Firms still do not Use Cloud Computing. An Empirical Study of Barriers and Enhancers", Proceedings of the 49th Annual Hawaii International Conference on System Sciences (HICSS-49), Kauai, Hawai, USA, 2016.

[26] Tornatzky, L.G. and M. Fleischer, The Processes of Technological Innovation, Lexington Books, Lexington, MA, 1990.

[27] Venters, W. and E. Whitley, "A Critical Review of Cloud Computing: Researching Desires and Reality", Journal of Information Technology, 27(3), 2012, pp. 179-197.

[28] Verma, H., "Data-warehousing on Cloud Computing", International Journal of Advanced Research in Computer Engineering & Technology, 2(2), 2013, pp. 411 – 416.

[29] Witten, H., E. Frank, and M. A. Hall, Data mining practical machine learning tools and techniques, 3rd ed., Elsevier, San Francisco, CA, 2011.

Appendix

I. Questions used as initial features

Variable	Definition
Label (Dependent Variable)	
Propensity for cloud computing adoption	How relevant is cloud computing for your company ?
Features (Independent Variables)	
Technological Characteristics	
Telework	Can employees of your company access your computer system remotely from outside the company, for instance from home, from field operation or while travelling?
Use of ERP systems	Do you use an ERP system, that is Enterprise Resource Planning?
Use of SCM systems	Do you use an SCM system, that is Supply Chain Management?
Use of CRM systems	Do you use a CRM system, that is Customer Relationship Management?
Use of SRM systems	Do you use an SRM system, that is Supplier Relationship Management?
Use of CAD systems	Do you use a CAD system, that is Computer Aided Design?
Use of CAM systems	Do you use a CAM system, that is Computer Aided Manufacturing?
Electronic Orders from Customers	Can customers order goods or services from your company online on the internet or through other computer networks, not counting manually typed e-mails?
Electronic invoicing	Does your company use e-invoicing, that is sending or receiving invoices electronically?
Use of software applications to collaborate with other firms for product/service or process innovations	Does your company use online software applications other than e-mail to collaborate with business partners in the development of new products, services or processes?
Interest in Data warehouses/data mining	Do you consider the topic of Data warehouses and data mining to be very relevant, partly relevant, or not relevant for your company?
Interest in Mobile services	Do you consider the topic of Mobile Services such as mobile commerce and remote access technologies to be very relevant, partly relevant, or not relevant for your company?
Organizational Characteristics	

Sector	Survey Sector Number/Name
Size	How many employees does your company have in total, including you?
Product/Service Innovation	During the past 12 months, has your company launched any new or substantially improved products or services?
Process Innovation	During the past 12 months, has your company introduced any new of significantly improved internal processes, for example in production of logistics?
Collaboration with other firms in product/service innovations	Were external experts or business partners involved in developing the new processes?
Employment of ICT Personnel	Does your company currently employ ICT practitioners?
ICT Outsourcing	In the past 12 months, has your company outsourced any ICT services to external service providers which were previously conducted in-house?
ICT investment reduction	Have you cancelled or significantly downsized any ICT or e-business projects due to the economic crisis?
Production locations	In how many locations in total, does your company operate production plants?
Geographical scope of sales	What is your company's most significant sales market?
Geographical scope of procurement	Do you procure primarily from suppliers in your region or from an international supplier base?
Environmental Characteristics	
e-Competition	Do you think that e-business has an influence on competition in your sector?
Price competition	Is the price of products very important, quite important or not so important?
Product quality competition	Is the product quality very important, quite important or not so important?
Image and product design competition	Is image and design of the products or respectively company very important, quite important or not so important?
Customer service competition	Is customer service very important, quite important or not so important?

II. Feature Weights

Feature	Weight (%)
Interest in data warehouses/data mining	18.40
Interest in mobile services	9.20
Use of SCM systems	6.53
ICT investment reduction	6.23
Employment of ICT Personnel	5.04
e-Competition	5.04
Use of ERP systems	4.15
Use of software applications to collaborate with other firms for product/service or process innovations	3.56
Use of CAM systems	3.56
Electronic invoicing	3.26

Size	2.97
ICT Outsourcing	2.97
Product/Service Innovation	2.67
Use of SRM systems	2.67
Process Innovation	2.08
Involvement of other firms in product/service innovations	2.08
Product Quality competition	2.08
Use of CRM systems	1.48
Sector	1.48
Geographical Scope of Procurement	1.19
Image and product design competition	1.19
Telework	1.19
Customer service competition	1.19
Price competition	1.19
Use of CAD systems	0.89
Electronic Orders from Customers	0.59
Production Locations	0.59
Geographical Scope of Sales	0.30

III. Receiving Operating Characteristic (ROC) Curves of Classifiers