International Journal of Advanced Research and Review

www.ijarr.in

THE ROLE OF MACHINE LEARNING IN SERVICE ORIENTED ARCHITECTURE AND BUSINESS PROCESS IMPROVEMENT

Supun Dissanayake

School of Computing, University of Colombo, Sri Lanka.

ABSTRACT

Business organisations must adopt to new technologies in order to improve business processes due to the advancement of computer technologies. Machine learning and Big Data processing are widely used in the software engineering industry to provide intelligent/automatic data processing solutions by reducing manual workloads. Therefore, this paper identifies machine learning tools that can be used to benefit organisations with service oriented architectures. Moreover, it proposes ML metrics that can be applied through a maturity model to support business process improvements for organisations following SOA.

Keywords: Machine Learning, Service Oriented Architecture, Maturity Metrics, Business Process Improvement

I. INTRODUCTION

Service Oriented Architecture is a software development process where the development, asset structuring, component structuring and service structuring is designed using core principles of message passing [1]. Moreover, it allows the software system to use published/discoverable interfaces to provide facilities to services developed in a network. Big Data are datasets that are too big to be handled through a normal computing environment [2] and it has become an emerging phenomenon due to the increase of digital data processing in last few decades [3]. Thus, organisations following SOA should discover novel methodologies to efficiently process Big Data; this is where machine learning comes into effect. Machine learning is the usage of artificial intelligence techniques, which allows software to automatically learn without guidance [4]. Therefore, it can be proposed that organisations that use Big Data through Service Oriented Architectures should consider using machine learning tools to easily improve their business processes. Thus, this paper critically evaluates the usage of machine learning tools for SOA development and depicts their features, advantages and disadvantages. Moreover, it identifies the importance of Big Data for business process improvements and proposes a maturity metrics that should be adopted by organisations following SOA principles. Furthermore, it proposes methodologies to use machine learning as SOA maturity metrics in order to climb up through a maturity model such as CMMI. Therefore, organisations that are following SOA principles can follow this research methodology to improve their business processes easily and efficiently.

II. OVERVIEW OF MACHINE LEARNING TOOLS AND THEUR EFFECTS ON SOA

2.1: Machine Learning

Machine learning is a subsection of artificial intelligence that allows computers to learn without any internal programming guidance [5]. It transforms data into software, thus, it is being used to develop computer software that can make intelligent decisions when exposed to big data [5]. This implies that computer Systems that consist of machine learning tend to improve further when it is exposed to more data. Machine learning was initially developed for pattern recognition purposes [5]. It has the ability to detect and uncover patterns in order to make predictions for the future and make decisions for uncertain conditions [6]. This ideology evolved gradually and researchers tried to identify ways to analyse data, use artificial intelligence (AI) to learn from data and take intelligent decisions [5]. Thus, reliable, repeatable and robust decision-making abilities were given to computer programs through the usage of machine learning technologies. Furthermore, it is further evolved to use mathematical algorithms to process big data in a faster and efficient manner which has become extremely invaluable for modern day computing. Table 1 depicts few of the main reasons for the evolution of machine learning.

Reasons for the evolution of Machine Learning	**Explanation**
Exponential Data Growth	Due to the growth of the internet in last few decades, a gigantic amount of readily available and archived data is available for processing. Furthermore, devices that are connected with IoT (Internet of Things) have grown exponentially during past few decades. Therefore, Machine Learning techniques are applied to simulate data by accessing existing and historical data [4].
Cheap Global Digital Storage	There is an increasing availability of cheap, globally accessible data storage that can be accessed through the internet. These platforms provide services through cloud services and they can be accessed via multiple devices [4].
The Rise of the Big Data Analytics	Along with the growth of the internet, more and more organisations inclined towards the usage of big data analytics methodologies to improve their businesses [4].

Table 1: Reasons for evolution in machine learning

ML can be easily defined by comparing it with current programming paradigms. For example, traditional programming methodologies make the computer obtain data and the program and output the desired result. On the other hand, machine learning makes the computer to reverse-engineer data and the output and develop a new program [4].This is what gives the computer the ability to predict the desired output depending upon the input data. An example for this is the product recommendations from Amazon.com. When a user buys a product, Amazon uses ML and predictive analytics and displays all the products that the user might want to buy in the future. Thus, it can be regarded as a very impressive marketing strategy for large corporations, which allows them to improve their business.

2.2: Machine Learning tools for Service Oriented Architecture

There are a wide variety of machine learning tools available in the industry that can be used without having a deep knowledge of machine learning algorithms. These machine learning tools can be used to improve organisations that use SOA platforms. Following subsections depicts five main ML tools available in the industry and critically analysis of their features, advantages and disadvantages.

2.2.1: Azure ML

Azure ML is a machine learning service developed by Microsoft that allows its users to develop predictive analytic models, data exploration/visualisation, machine learning algorithms etc. for cloud services [7]. Moreover, it allows developers to produce experiments, evaluations and "fail fast" to develop practical prediction models [7]. Thus, users can understand and prove whether the business processes will have a successful outcome rather than leaving the final result as a random guess. Diagram 3 depicts three main Azure ML algorithms.

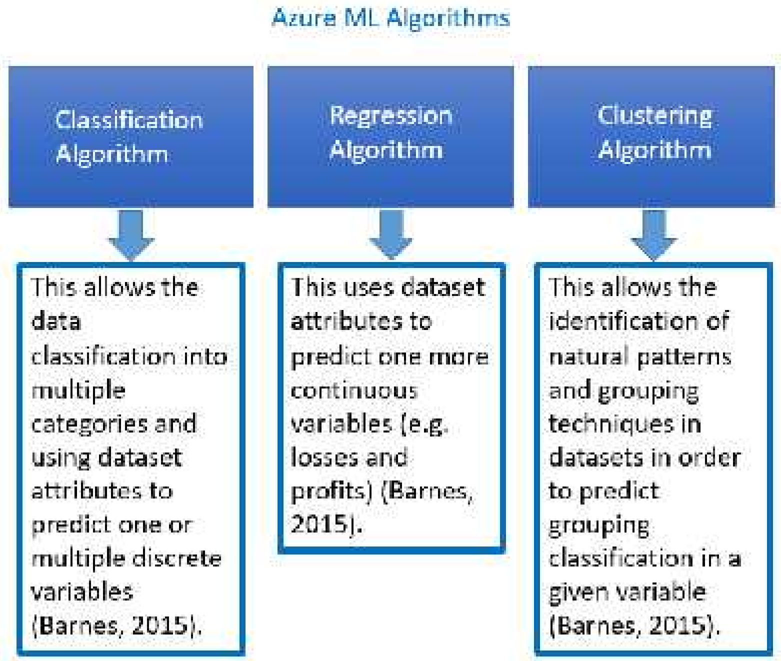

Fig 1: Azure ML Algorithms

Strengths

Predictive analytics methodologies of Azure ML are extremely beneficial to aid business decisions since it increases the chance of obtaining a positive outcome through the usage of AI [7]. For example, it can be used to depict the effectiveness of sales and marketing strategies of organisations by processing customer responses through data patterns involving factors such as demographics, prices, discounts, seasons and social media effect by identifying data patterns [4]. Therefore, these patterns can be used to take educated marketing decisions for the wellbeing of the business. Moreover, it allows to identify and track the past and present behaviours of the target market to refine marketing strategy and improve business processes.

Weaknesses

The performance of Azure can be quite slow [8]. Since it provides R studio, it doesn't have an add-on for Excel. Thus, novel users who prefer Excel will have to adapt to a completely new software. It stores all the data in the cloud, thus it needs to be continuously managed and maintained [9]. Furthermore, it requires a good knowledge about Azure platform to ensure all the functions are working efficiently and according to standards. This implies that if Azure platform is not monitored, it can cost an organisation large amount of money in the longer term.

2.2.2: Cloud ML

Cloud machine learning is developed by Google and it allows its developers to produce machine learning models to process data of any size.

Strengths

It provides "TenserFlow", which is a framework that allows the development of machine learning models for SOA architectures such as Google Cloud Speech [10]. It consists of a scalable infrastructure; thus, the size of the model can be scaled to the desired level [10]. Moreover, the integration of service with the Cloud Dataflow allows pre-processing of data and data access to Google Cloud Storage, Google BigQuery etc. [10]. This implies that Cloud ML is extremely beneficial for SOA since it provides scalable architecture depending on the business needs. For example, the organisation will only require small cloud space when the business initiates, however it can be scaled up when necessary without transferring into an alternative service.

Weaknesses

Due to the architecture of the platform, it is difficult to transition away from the Google Cloud [11]. This implies that the developers must stick to Google rather than have any backup option. Moreover, compared to Amazon, there is a limitation in the choice of programming languages [11], hence less diversity for development processes. Furthermore, most of the components available will only work for Google branded platforms [11]. This depicts that organisations will not have much control over VMs. Table 3 depicts few of the key features used by Cloud ML.

2.2.3: Amazon ML

Amazon ML platform is a cloud-based service that permits its developers to efficiently use machine learning methodologies to optimise their products. It achieves this feat through the provision of visualisation tools and wizards that allow the development of machine learning models for SOA [12].

Strengths and ML model features

Amazon allows its model deployment as an SOA through an API thus the management of infrastructure and code generation is easily taken care of by the platform [12]. It uses mathematical models that generate predictions by efficiently identifying data patterns, which allows organisations to simplify complex business situations [13]. It is built around three machine learning models

- Binary Classification: The process of using machine learning to predict values that can only be in two states (e.g.: true or false) [13].

- Multiclass Classification: The process of using machine learning to predict values that are limited, set of permissible values and pre-defined [13].

- Regression: The process of using machine learning to predict numerical values [13].

Weaknesses

It is prone to multi-zone failures [14]. This can be an issue for organisations that need to spread its services to multiple zones. Furthermore, it can cause issues with its shutting down process [14]. For example, users are only notified a few days before it is being shut down, which can be a very risky situation for organisations if they do not have a viable backup option.

2.2.4: WEKA 3

WEKA 3 is a machine learning workbench that allows the development of applications that consist of machine learning functionalities to deal with various real word problems [15]. It provides a platform that is easy to understand without being an expert in machine learning methodologies [15]. Thus, developers do not require higher knowledge in mathematical algorithms to understand how to use machine learning tools of Weka 3. Fig 2 depicts some of the projects developed using WEKA 3 work bench.

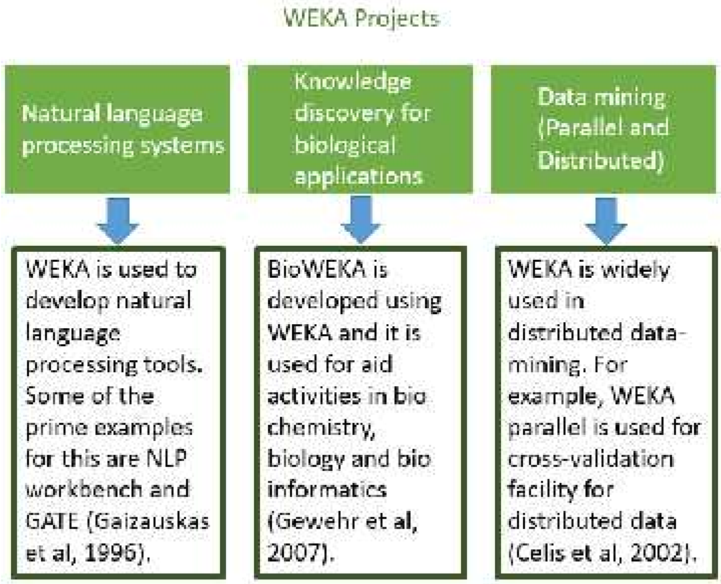

Fig 2: WEKA Projects

Strengths

WEKA was mainly developed to provide its users with ML algorithms and pre-processing tools for data analytics. It allows the comparative analysis of machine learning and dataanalytics tools in order for users to understand the best possible methodology to use in their projects. Moreover, its extensible and modular architecture provide the ability to develop data mining tools with a wide range of base learning algorithms. Due to the provision of WEKA API, toolkit extension is a much easier process and also plugins are available to automate the connection of further algorithms through WEKA GUI. Algorithms for clustering, regression, attribute rule selection, classification etc. are available through the workbench [15]. Furthermore, data visualisation methodologies, pre-processing tools, statistical evaluations and results evaluation supports data mining processes such as CRISP-DM [16].

Weaknesses

It only has the ability handle smaller datasets [17]; thus, when larger datasets are introduced it gives “Out of Memory Errors”. Hence, this could be problematic for an organisation that is processing a large amount of data constantly. Moreover, sometimes it gives exception errors when comparing produced clusters and also it gives errors for non-global clusters [17]. Furthermore, it does not scale appropriately with increasing datasets, which might cause organisations to look for alternative machine learning techniques when the organisation grows.

2.2.5: PROMISE Open Data Set

PROMISE (Predict Or Models In Software Engineering) is a collection of open source software engineering tools and datasets that are available for research purposes and development of predictive software models [18]. PROMISE was initially developed by NASA and the University of Ottawa. It was inspired by UCL Machine Learning Repository, which is a widely accepted and recognised repository used by experts in the industry [19].

PROMISE arranges its data sets into 5 categories and table 6 depicts these categories and what they contain.

Strengths and Functionalities

Data Set	**Description**
Defect Prediction	This contains the most number of datasets with 54 datasets and each of these datasets is connected to one singular purpose with local attributes and number of instances [19].
Effort Prediction	Effort Prediction consists of 12 datasets, which are coming from multiple sources that have different purposes with various attributes, numbers of instances and definitions [19].
Text Mining	Text Mining consists of 8 datasets and most of these were collected between 2005 and 2008 from NASA [20].
Model-Based Software Engineering	This contains 3 datasets which were available since 2009 for multiple purposes [19]. For example, CM1-bn dataset contain data of the quality-measures from 6 attributes with different instances [21].
General	This contains 7 datasets which can be used for purposes which are different from previous sections. For example, it consists of Reuse dataset which contains a set of candidate reusability factors [22].

Table 2: Promise Datasets

Weaknesses

It can cause overlapping of classes where some data contain samples from different class values when it is dealing with classifications [23]. This was evident in the PROMISE repository that was used by NASA where they found inconsistent and contradictory data [23]. This implies it can cause difficulties when developing predictive models for SOA due to this error. Moreover, it can cause problems with data shifting when there are differences in test data distribution and the training distribution [23]. Furthermore, Rodriguez et al [23] implicate that PROMISE can sometimes cause data imbalance. This means that it can generate models with hidden factors and misleading predictions.

Overall, it is evident that above tools can be used to enhance business processes of organisations following SOA principles. Hence, to conclude, table 3 is developed to depict some of the functionalities that would be required for an organisation following SOA principles and the availability of these functionalities in above machine learning tools.

	Azure ML	Cloud ML	Amazon ML	Weka 3	PROMISE
Data Pattern Identification	Yes	Yes	Yes	Yes	Yes
Model creation	Yes	Yes	Yes	Yes	Yes
Scalability	Yes	Yes	No	Yes	Yes
Fast Performance	No	Yes	Yes	Yes	Yes
Diverse program language dependency	No	No	Yes	Yes	Yes
Data Pre-Processing	Yes	Yes	Yes	Yes	Yes
Data prediction	Yes	Yes	Yes	Yes	Yes
Natural Language Processing	Yes	Yes	Yes	Yes	Yes
Processing Large Data sets	Yes	Yes	Yes	No	Yes
Multi Zone Shifting/Availability	Yes	Yes	No	Yes	Yes

Table 3: ML functions for SOA and their availability in ML tools

III. APPLICATION OF MACHINE LEARNING FOR BUSINESS PROCESS IMPROVEMENT

3.1: Big Data for BPI (Business Process Improvement)

Big Data can be defined as datasets that are too big to be handled through a normal computing environment [2]. Big data has become an emerging phenomenon due to the increase of digital data processing in last few decades [3]. Thus, researchers in the information science industry have taken a more serious approach to identify methodologies that allow big data processing methodologies to improve businesses [24].

According to Kraska [25], Big Data contains two main problems, which are “big throughput” and “big analytics”. Big throughput has issues associated with data storage, manipulation and transformation of data into valid knowledge [25]. Conversely, data analytics is the process of using algorithms which are built around machine learning and clustering to identify valid content from Big Data [3]. It is important to denote that there are multiple analytics tools available in the industry which are specifically designed to carry out data analytics processes using artificial intelligence methodologies. According to LaValle et al [26], organisations with great performance use advanced big data processing techniques compared to lower performing organisations. Moreover, it is evident that these Big Data analytics techniques are efficiently used in industries such as medicine, agriculture, academia etc. to improve their organisation performances [27]. This proves the fact that the efficient processing of Big Data can be used for business process improvement. Therefore, the challenge is to identify how to control big data efficiently for BPI. As a result, table 4is developed to initially identify four main processing methods that businesses should use to handle big data.

Big Data Processing Methods for Business Process Improvements	**Description**
Data Mining	Data mining is analysing large sets of data to identify unpredicted associations and make data obtained more understandable for the data analyst [28]. The input data is added as tables; however, output data will be in the form of patterns, graphs, clusters, tree structures etc. [28].
Machine Learning	As it was depicted in part 1, machine learning is using artificial intelligence techniques such as neural networks to process data without being explicitly being programmed to understand data. Thus, it makes the software to learn, adapt and evolve in order to improve business processes.
Process Mining	Process mining allows combining process perspective with ML and data mining. For example, event data can be used to identify models and they are replayed to analyse performance and compliance [28].
Visualisation and visual analytics	This is used when there are too many unknowns that cannot be resolved through data and process mining. It involves combining automated analysis with interactive visualisation to improve Big Data identification in order to have a smooth decision-making process [29].

Table 4: Big Data Analytics for BPI

In order to apply above methods for an organisation following SOA principles, the business must identify their maturity levels to plan and control their business processes. This allows them to compare and analyse the maturity status of the organisation. Traditionally, maturity models for SOA are developed by taking wide variety of factors into consideration such as business size, environment, market, business goals, requirements etc. Therefore, maturity level metrics depicted in section 3.2 is developed for Service Oriented architecture by closely following research ideologies depicted by Welke et al [30], Harris [31], Inaganti et al [32] and Waehner [33]. These metrics should be adapted by popular maturity models in the industry such as CMMI in order to allow machine learning functionalities to aid the maturity level identification process. It is evident from this paper, that organisations that are following SOA principles must take Machine learning functionalities and Big Data analytics into consideration in order to improve their business processes. Thus, the unique nature of these maturity metrics compared to other metrics in CMMI is that it consists of machine learning and big data processing functionalities. Hence, the ultimate aim of these maturity metrics is to allow businesses to improve their processes and adapt to changing environments using machine learning and big data analytics.

3.2: ML for SOA Maturity Metrics

There are multiple methodologies that can be used to improve business processes. As it was discussed in part 1 of this paper, machine learning can be regarded and one of the newer approaches that can be used for these criteria. Therefore, machine learning is effectively used in this maturity metrics. This will allow the organisation to achieve key process indicators through the usage of machine learning methodologies. This could be accomplished by using machine learning tools that were introduced in the first part of the research paper. Following subsections proposes which machine learning metrics should be adopted by organisations to improve business processes to move up the maturity levels.

3.2.1: Level 1 to Level 2

Initial level of a maturity model specifically targets small organisations that use simple SOA methodologies. Therefore, in order to move from level 1 to level 2, the organisation must improve upon software performance, improve development/deployable environment, increase the quality of big data analysis etc. One of the ways that machine learning could be applied for this process is through big data analysis. Machine learning can be effectively used for this procedure to visualise data to help the organisation in its day to day activities. This could be achieved by dividing the Big Data collection process into two scenarios: acquisition and preparation [33]. Data acquisition involves the process of integrating appropriate databases. This will allow the organisation to visualise data to make intelligent business decisions. This process could be achieved through machine learning by using a platform such as Apache Hadoop Cluster, which allows storing unstructured data [33]. Moreover, Hadoop could be used for data preparation, which is combining data from multiple sources to identify the effectiveness of business processes, thus allowing the organisation to make further improvements.

Key Process Areas	**Machine Learning Tool**	**Business Process Improvement**
Integration with databases, Big data collection, Performance Management	Apache Hadoop Cluster	Data visualisation for business decisions, storage of unstructured data, data preparation

Table 5: ML metrics to achieve level 2

3.2.2: Level 2 to Level 3

In order to move from level 2 to level 3, the organisation must mainly adopt descriptive models, predictive models and improve upon data analytics processes. Descriptive modelling involves possessing data for knowledge management [34]. This implies that there should be software in place that consists of machine learning algorithms to process attribute sets and provide relevant answers [34]. Similarly, predictive modelling involves predicting answers to future problems by studying current input attributes [34]. Both of the above machine learning models should consist of machine learning algorithms developed using decision trees for the software to make intelligent decisions. Hence, the organisation will be able to define current and upcoming issues in the organisation. WEKA 3 allows users to develop machine learning tools that consist of decision trees [15]. Hence, it could be used by the organisation to implement above features and move up the maturity ladder.

Key Process Areas	**Machine Learning Tool**	**Business Process Improvement**
Application of descriptive model, Application of predictive models	WEKA 3	Enhanced knowledge management, predictions for future issues

Table 6: ML metrics to achieve level 3

3.2.3: Level 3 to Level 4

In order to move from level 3 to 4, the organisation should contain Machine Learning/Data analytics models. Therefore, the organisation should have appropriate analytics tools to analyse all the data that it contains in order to make intelligent decisions about business processes. Moreover, it should have an archive of historical data to be combined with current data to analyse and predict business processes [33]. This will allow the organisation to identify areas of business that requires more attention for improvement. Moreover, it will provide details about processes that should be undertaken and improved for future business success. Furthermore, the analytics model should be validated constantly to check whether machine learning algorithms functions accordingly for business process improvement. [33]. An example for this is Amazon Machine Learning platform. It contains machine learning algorithms to identify past search results of the user and to predict products that they might be willing to buy in the future. Thus, this logic can be similarly applied to businesses. This feature could be achieved via platforms such as Amazon ML, Azure ML and HP Haven which provides this functionality as one of their platform features.

Key Process Areas	**Machine Learning Tool**	**Business Process Improvement**
Data analytics methods, Data archive	Azure ML, Amazon ML, HP Maven	Learn from historical data and make predictions about future business processes, issues and success rates.

Table 7: ML metrics to achieve level 4

3.2.4: Level 4 to Level 5

In order to move from level 4 to level 5 of the maturity model, it is required to automate the big data process, analytics and predictive criteria without any human interactions. This could be achieved via stream analytics [33]. This means that the organisation should have machine learning software in place to develop, understand and release algorithms to identify and predict business activities [33]. This feature could be achieved by open source machine learning frameworks such as Microsoft Azure Stream Analytics, Apache Spark Streaming, IBM Infosphere Streams etc. The application of this methodology will allow the understanding of business operations and patterns in real time without the need for a model rebuild [33]. Furthermore, this process will help to reduce the human interactions required to identify business process improvements. Moreover, machine learning can be applied to identify changing business processes involving employees such as their work efficiency and this information could be obtained in real time to identify overall business process efficiency. Furthermore, real-time streaming data can be used to identify the required evolution pattern of the business. Thus, it will allow the businesses to grow and be more profitable.

Key Process Area	**Machine Learning Tool**	**Business Process Improvement**
Real Time event processing, analysis and predictions	Azure ML, Apache Spark Streaming, IBM Infosphere	Identification of live business process changes, identification of business evolution patterns.

Table 8: ML metrics to achieve level 5

IV. CONCLUSION AND FUTURE RESEARCH

This paper clearly depicts on the role of machine learning on Service Oriented Architecture and the usage of machine learning in business process improvement. The efficient usage of machine learning techniques in organisations allows them to keep the competitive edge when they are going against their rivals in the industry. Usage of a wide variety of machine learning tools such as Azure ML, Amazon ML, WEKA 3 etc. provides methodologies that can be implemented easily through organisations for business improvement without being proficient in machine learning and algorithms. Moreover, it is evident that due to the growth in digital data processing which was developed along with the growth of the internet in the past few decades have made it vital for businesses have methodologies to process big data. Ultimately, the efficient use of these methodologies helps businesses to improve its day to day processes. Thus, this paper provides a solution to combine business process improvement with machine learning technologies via maturity model. The usage of maturity metrics based on machine learning technologies implicates the role that machine learning can play for business process improvement. It clearly displays machine learning tools which could be used as platforms to achieve key process indicators and move up the maturity ladder. However, it is recommended to develop maturity metrics much further by adding more key process areas to identify further business aspects. Moreover, application of the above methodology through software will allow organisations to automatically identify and improve business processes through machine learning to improve the business organisation.

REFERENCES

[1]. Ramachandran, M., (2016), Service Oriented Architecture (SOA) Introduction [PowerPoint Presentation], Leeds Beckett University. Available from: < https://app.box.com/s/xzr4xgn1p71afp5icoc1> [Accessed 03 March 2016].

[2]. Costello, T., Prohaska, B., (2013), 2013 Trends and Strategies, IT Professional, vol. 15, no. 1, pp. 64–64.

[3]. Chiang, R. H. L., Goes, P., Stohr, E. A., (2012). Business Intelligence and Analytics Education, and Program Development: A Unique Opportunity for the Information Systems Discipline, ACM Transactions on Management Information Systems, vol. 3, no. 3, pp. 12:1–12:13.

[4]. Barnes, J., (2015), Azure Machine Learning, Microsoft Azure Essentials. 1st ed, Microsoft.

[5]. Alpaydin, E., (2010), Introduction to Machine Learning. MIT Press, pp 1-579.

[6]. Murphy. K., (2012), Machine Learning — A Probabilistic Perspective. The MIT Press, pp 1-25.

[7]. Fryer, A., (2015), Microsoft Azure Machine Learning [PowerPoint Presentation], Microsoft. Available from: <http://www.ecmwf.int/sites/default/files/elibrary/2015/13314-microsoft-azure-machine-learning.pdf> [Accessed 04 March 2016].

[8]. Szeliga, M., (2015), Pros and Cons of Azure Machine Learning [Internet] Available from:<http://blog.sqlexpert.pl/2015/01/25/pros-and-cons-of-azure-machine-learning/> [Accessed 04 March 2016]

[9]. Icorps, (2016), What Are the Pros and Cons of Microsoft Azure? [Internet] Available from: <http://blog.icorps.com/pros-and-cons-microsoft-azure> [Accessed 04 March 2016]

[10]. Google Cloud Platform, (2017), Google Cloud Platform [Internet] Available from: <https://cloud.google.com/> [Accessed 04 March 2016]

[11]. Dodda, M., (2016), The Top 3 Cloud Platform Providers: Pros and Cons [Internet] Available from:<http://bitmin.net/blog/the-top-3-cloud-platform-providers-pros-and-cons/> [Accessed 01 March 2016]

[12]. Pogorzelska, B., (2016), Machine Learning at Amazon [PowerPoint Presentation], Amazon Web Service. Available from: <https://aws-de-media.s3.amazonaws.com/images/slides/enterprise-summit-2016/track-cloud-native/1545_Amazon_MaschinellesLernen-FRAv4.pdf> [Accessed 03 March 2016].

[13]. Amazon, (2017), Amazon Machine Learning Developer Guide. 3rd ed, Amazon.

[14]. Kalavagunta, G., (2015), Review of Cloud based Machine Learning platforms [Internet] Available from: <https://www.linkedin.com/pulse/review-cloud-based-machine-learning-platforms-ganapriya-kalavagunta> [Accessed 02 March 2016]

[15]. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I. H., (2009). The WEKA Data Mining Software: An Update, SIGKDD Explorations, Volume 11, Issue 1.

[16]. Shearer, C., (2000), The CRISP-DM model: The new blueprint for data mining. Journal of Data Warehousing, 5(4).

[17]. Ayyoob, M. P. (2015), Tools Pros and Cons of Clustering Algorithms using Weka Tools. International Journal of Computer Applications: National Conference on Advances in Computing Communication and Application, pp- 13-16.
[18]. Cukic, B., (2005), Guest Editor's Introduction: The Promise of Public Software Engineering Data Repositories, IEEE Software, vol. 22, no. 6, pp. 20-22.
[19]. Cheikhi, L., Abran, A. (2013). Promise and ISBSG Software Engineering Data Repositories: A Survey. In Software Measurement and the 2013 Eighth International Conference on Software Process and Product Measurement (IWSM-MENSURA), pp. 17-24).
[20]. Cleland-Huang, J., Settimi, R., Zou, X., Solc, P., (2006). The Detection and Classification of Non-Functional Requirements with Application to Early Aspects, in IEEE RE.
[21]. Leandro, Minku, L., Yao, X., (2011). A Principled Evaluation of Ensembles of Learning Machines for Software Effort Estimation, PROMISE ’11, Proceedings of the 7th International Conference on Predictive Models in Software Engineering, September 20-21, Canada.
[22]. Morisio, M., Ezran, M., Tully, C., (2002). More Success and failure factors in software reuses, IEEE Transactions on Software Engineering, vol. 28, 4, pp. 340-357.
[23]. Rodriguez, D., Herraiz, I., Harrison, R., (2012), On Software Engineering Repository and their Open Problems [PowerPoint Presentation], Oxford Brookes University. Available from: <http://promisedata.org/raise/2012/slides/RAISE12_Rguez.pdf> [Accessed 04 March 2016].
[24]. Chen, C. L. P., Zhang, C. Y., (2014), Data-intensive applications, challenges, techniques and technologies: A survey on Big Data, Information Sciences, vol. 275, pp. 314–347.
[25]. Kraska, T., (2013) Finding the Needle in the Big Data Systems Haystack, IEEE Internet Computing, vol. 17, no. 1, pp. 84–86.
[26]. LaValle, S.,Lesser, E., Shockley, R., Hopkins, M. S., Kruschwitz, N., (2011) Big data, analytics and the path from insights to value, MIT Sloan Management Review, vol. 51, no. 2, pp. 21–32.
[27]. Vera-Baquero, A., Colomo-Palacios, R., Molloy, O. &Elbattah, M. (2015). Business process improvement by means of Big Data based Decision Support Systems: a case study on Call Centers. International Journal of Information Systems and Project Management, 3 (1), pp- 1-22
[28]. Van der Aalst, W., Damiani, E., (2015) Processes meet big data: connecting data science with process science, IEEE Trans. Serv. Computing. pp 810–819.
[29]. Wijk, J., (2005) The Value of Visualization, in Visualization 2005. IEEE CS Press, pp. 79–86.
[30]. Welke, R., Hirschheim, R., Schwarz, A. (2011). Service-oriented architecture maturity. IEEE Computer, 44(2), pp-61-67.
[31]. Harris (2008), A SOA Maturity Model, Research Report, University of Twente, Netherland.
[32]. Inaganti, S., Aravamudan, S. (2007), SOA maturity model, available at: http://bpmg. orgwww.bptrends.com/publicationfiles/04-07-ART-The%20SOA%20MaturityModelInagantifinal.pdf/
[33]. Waehner, K., (2016), How to Apply Big Data Analytics and Machine Learning to Real Time Processing [PowerPoint Presentation], TIBCO. Available from: <http://public.jugru.org/jpoint/2016/msk/day_1/track_2/2016_04_JPoint_Moscow___MachineLearning_EventProcessing_Kai_Waehner.pdf> [Accessed 03 March 2016].
[34]. Somogyi, Z., (2016), ARTIFICIAL INTELLIGENCE IN BUSINESS PROCESS IMPROVEMENT [Internet] Available from:<https://zoltansomogyi.blogspot.co.uk/2016/08/artificial-intelligence-in-business.html/> [Accessed 05 March 2016]