IJCAT - International Journal of Computing and Technology, Volume 5, Issue 3, March 2018
ISSN (Online) : 2348-6090
www.IJCAT.org
Impact Factor: 0.835

A New Approach Based on Requirements Engineering in Software Projects Development Using Random Forest Algorithm

[1] **Vahab Yousefi Davood-Khani,** [2] **Soheil Afraz ***

[1] Department of Computer Engineering, Ardabil Branch, Islamic Azad University, Ardabil, IRAN

[2] Department of Computer Engineering, Ardabil Branch, Islamic Azad University, Ardabil, IRAN

Abstract - Due to the development and growth of information technology, software systems and server organizations have encountered a huge amount of information and new requirements. The requirements that change over time and elicitation of these requirements from an aggregation of needs are an important challenge. To overcome these problems, researchers for development of your projects presented different techniques and approaches. Software projects developers are among the pioneers in this field and are focusing on the requirements engineering approach using elicitation techniques. This paper presents a new approach based on requirements engineering has been developed in software projects by Random Forest algorithm. The approach consists of three phases. In the first phase, the integrated modeling language graphs were used to formal description of the architecture and system used to identify the project risks and bottlenecks. In the second phase, the features and diagnosis indicators as learning are given to Random Forest algorithm. In the last phase, a dynamic model is presented to evaluate the reliability of detection with colored petri Nets, which shows the risks and critical conditions in the project. In Comparison of proposed approach with another works, the results show that identification of software projects risks and bottlenecks is a critical and necessity to make progress in the software development based on Requirement Engineering activities.

Keywords: *Requirements Engineering, Random Forest Algorithm, Software Development, Machine Learning, Colored Petri-Nets*

1. Introduction

Changing and transforming in today's world in accordance with the needs and requirements of the clients is one of the main concerns of service providers. This has led to the concentration of organizations, institutions, and large companies in all areas of competition to meet these flexible needs. Many researches and studies have been conducted in this field to address the needs of the clients, but recognizing these needs is an important challenge.

One of the factors in sustainability in current business markets knowing the exact needs and examining them and providing a solution to these needs. Prioritizing needs and recognizing them is one of the key factors in this way. Emphasize of requirements is one of the important issues in software and information systems architecture. In the architecture of the software systems, after identifying the requirements that must be met by the system, taking into account the operational restriction of the project, must be emphasize the requirements according to their importance, so that the time and resources available to implement the project, they will be allocated according to the emphasize of the needs. This affords a more favorable use of the time and resources of the project and cased to in the time taken to implement the project to achieve a product that meets the higher priority requirements. On the other hand, the requirements of the entire software system are divided into two general categories of functional and non-functional requirements, the former being the ability to run, and the second to the quality of the execution of the system's tasks. The requirements for prioritizing applied the non- functional requirement.

Techniques for extracting requirements are the various tools that a mechanic has in his toolbox, a tool may be used more than the rest, but not suitable for anything, a mechanic with difficulty comes to determine which tools to use, so we, as a requirement engineer, should be able to apply the appropriate technique to suit the subject matter. The skill of a requirement engineer is to identify these real needs by using one or a combination of several requirements elicitation techniques. For example, use a method like interviews to hear what stakeholders are interested in and determine their correctness using a different method, such as examining the pertaining documentation.

In the last few years, various methods have been developed for the design of requirements in the development of software projects and have been dealt

with in a variety of studies. Several of the important researches have been investigated in this paper.

In a method for classifying quality requirements, they used machine learning techniques. In this research, many efforts have been devoted to engineering requirements to ensure to meet the requirements of the project. Among the tasks performed, automatic processing of requirements is needed to assess their quality. Of course, the challenge that can be considered for research is that the qualitative assessment interpreted according to the requirements that the experts and the demands of a specific project. The innovation of this research is that experts use learning techniques by utilizing the qualitative requirements defined by the project. In other words, the requirements are given to the learner as input and learner, and the learning system begins to predict according to the defined needs [11].

In [4] a Genetic Algorithm is proposed to solve the scheduling of related tasks, in which two important parameters are considered from the quality of service that they are time and cost. In this algorithm, random variables are used instead of producing the initial population. Combining the rates of genetic algorithm with turbulent variables has caused the solutions generated by this algorithm to be distributed throughout the search space and prevent early convergence of the algorithm. Better designs and products are obtained in shorter time and increased the convergence rate of the algorithm.

In another work [13] for the requirements engineering, anatomy science is used for the agile software process. In this way, one of the most important determinants of the success of software development is agility that speed and time are very important factors. In fact, with the help of this process, the relationship has been established between the developer and the stakeholders.
For quantitative security evaluation based on discrete-time Markov model, a method is proposed that calculates architectural security based on the component vulnerability and the frequency of meeting components during the implementation of the program. Also, to assess security using the Petri Nets, there is a model that is used to assess quantitative features of accessibility, performance, and security in the architecture. In other words, using the discrete-time Markov model, the reliability of the architecture is calculated based on the reliability of each component and the probability of transition [7].

In the resource for identifying the status of the requirements engineering modeling process has been talked about LMS that is a learning-based system. In this source, the method that is presented based on the machine learning and the work platform on the web and system management for learning with a very simple matrix method and the characteristics and requirements of the stakeholders are completed in accordance with the definitions and wishes in the form of a status matrix dynamically from the customer's review and the system is taught as an indicator or learner. So, in the post-learning stages, according to the characteristics, the system automatically displays the effect [12].

In a Structural method, object-oriented languages have been used for requirements engineering security in banking applications. The requirements engineering is expressed in an official language based on structured object-oriented (SRESOFL). This framework is looking for a method for promoting while paying particular attention to financial programs, and maintaining security. In this research, the mobile banking program is also considered as a platform [10].

In model-based requirements engineering [14] has been presented architecture for the needs of mentally-minded stakeholders. In the development of complex systems, this is a critical process. Currently, this architecture is a kind of new phenomenon, and is gradually transferring repetition in the operational sense of concept. MBSE provides a framework for effective and consistent systems from engineering and architecture perspectives Such as the integrated OPM approach relying on object-oriented. In this paper, a requirement-based engineering model (MBRE) is provided to facilitate the transition from the mental needs of the SHR beneficiaries to the needs of the SysReqs system and from SysReq to the architecture features of the proposed system. A case study was conducted based on a framework for the architecture of a robotic loading system at an international airport. In fact, has been proposed an ISO-based MBRE approach.

In [6] the engineering of integrated systems and software requirements engineering have been addressed systems for technical and customized systems. In fact, the development of complex software systems involves several engineering disciplines, such as mechanics, electricity, instrumentation control, and software engineering in the development of MBSE-based systems. In other words, a model based on system engineering provides the process of discipline. In this research, a model is included based on the understanding of requirements engineering and considered a case study by a car company.

In this paper we will investigate the existing system using the document analysis technique. Using this technique, we can obtain great information by studying the existing system and its related documentation, including the circulation of main and subservient work of system; diagrams, system guidance and reports, and we can get tighter information.

In other words, as our system studies the development of software projects, we consider a system such as education and describe the behavior and structure of this system with the integrated modeling language (UML). In fact, we describe the current process of a system first. So, to illustrate and extract the system requirements by document analysis technique, we will work on an official model using UML diagrams such as Component diagram and Sequential diagrams, and we will use the sequence diagram to show the behavior and the structure of the diagram. In the second phase, by obtaining the initial information from the initial process, considering the analysis of the documentation and the current history of the system, and we try to settle the restriction. In this way, we identify the characteristics that affect the results of the first phase and determine the extent of its functionality and its lack of functionality. Therefore, we need to find a way to illustrate the effective and risky needs among these extracted needs. Here we use the Random Forest algorithm. Among these extracted needs. Here we use the random algorithm. These features are characterized by the degree of impact. Once the project risks have been identified, we will work to fix it. In fact, by exploiting the extracted indicators, we are going to predict the risks and critical situations with advancing the goals of developing software projects. Finally, by providing a applicable model with the aid of colored petri nets, we will evaluate the reliability of the diagnosis in the proposed Approach.

The structure of the article is as follows: In the second part, we described in the subject the basic concepts of requirements engineering, the integrated modeling language, Random Forest algorithm, software development and machine learning. In the third section, a useful selection of previous studies is described in brief similar to the subject of the paper, and in the fourth section, the proposed Approach is described. To demonstrate the correctness of the approach presented in section fifth a case study was conducted to evaluate and modeling the proposed Approach and finally comparison table of previous works with our approach has been shown.

2. Fundamental Concepts

Nowadays, different definitions of concepts are given according to their application and their efficiency. In this article, concepts are defined according to the application and type of use.

2.1 Requirements Engineering (RE)

In fact, a systematic approach with defined rules for elicitation, organizing, documenting, analyzing, validating, and managing changes in system requirements is called requirements engineering. It is an interdisciplinary function that interfaces between the client and supplier domains to create and maintain the requirements that must be met by the system, software, or service. Requirements engineering considers discovery, elicitation, development, analysis and determination of authentication, validation, information, documentation and requirements management. Requirements engineering should be continuously performing its functions as an independent entity along with other business units. In Figure 1, it has been shown the relationship between the requirements engineering in the structure of the software project organization.

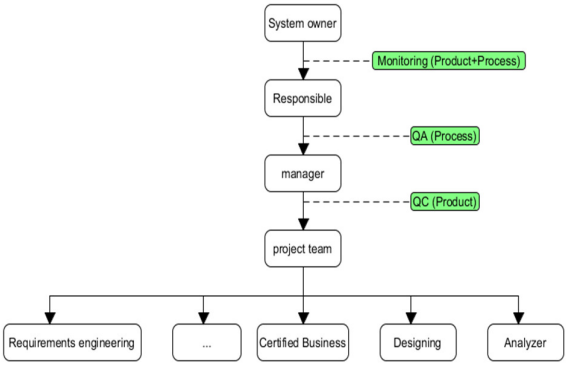

Figure 1. Relation of requirements engineering in the software project structure

2.2 Random Forest Algorithm

Random forest is an ensemble classifier that uses multiple decision trees to recognize a popular class [2]. A single decision tree suffers from variance or high deviation. In contrast, the random forest provides an unbiased estimation of the classification fault that is added to the forest in the form of trees. Also, the law of large numbers ensures that the random forest is resistant to over-fitting (Figure 2). One of the main methods of random forest in assessing the immediate safety of traffic is the estimation of variable importance [8, 15].

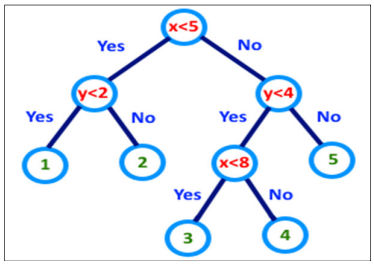

Figure 2. A simple example of a Random forest

3.2 Software Development

Software development is a set of software engineering activities designed to manage the life cycle of a software

product. In fact, software development started from the design phase of a conceptual solution to the problem (feasibility), after receiving the demands and analyzing the design system, and eventually this design becomes a real system with the help of implementation tools. On the one hand, the purpose of this process is to meet the needs of users and, on the other hand, ensure the proper quality of the system's operation, and therefore it must contain mechanisms for validation, that is, output in accordance with requirements and reliability and it is the correctness of the output function. The development process while giving freedom to the analyst must ensure to meet the scheduling of the implementation of the project.

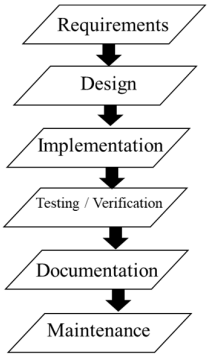

Figure 3. Software development steps

2.4 Machine Learning

Machine Learning is a science-based discipline that allows computers to learn without having to be specifically designed for that task. In fact, the process of data use it automatically generates a model, which serves as an input from a set of known features and provides something as a prediction as output. Figure (4) illustrates a simple example of machine learning.

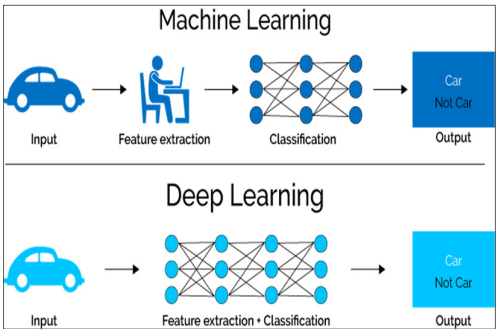

Figure 4. Machine learning Schema

2.5 Colored Petri Net

The colored Petri Nets provide a graphical representation of the system with a mathematical approach that can represent communication patterns, control patterns and information flows, providing networks for analyzing, validating and evaluating performance. The basis of the petri nets is based on the graph and, informally, we can say that it is a two-part directed graph consisting of two elements of the place and the transition. These networks are based on a situation, not event, which makes this model explicit the status of each item. Petri Networks offer models of structural and behavioral aspects of a discrete event system. It also provides a framework for analyzing, validating and evaluating performance and reliability [5].

3. The Proposed Approach

In this paper, a new approach based on requirements engineering in the development of software projects is presented using random forest algorithm. In this way, the formal description of the current process of the system is done using the integrated modeling language and then, extracted critical points according to the documentation and important indicators and given as input to the random forest algorithm. After applying the learning, the model predicts the states and the cases that lead to risk. In other words, the properties extracted by the RF algorithm apply to new data. Finally, using a colored petri nets is presented a dynamic and applicable model of the proposed Approach for evaluating reliability. In this way, is also evaluated the application of the proposed Approach. The following is described how the proposed Approach.

3.1 Describes the Behavior of the System

The purpose of this step is to describe the behavior of the study system. In fact, each process is characterized by a critical cycle, and follows a series of rules and frameworks to the destination. In each system, there is a normal and current flow, but identifying the indicators and characteristics that affect the system is an essential task. This identification should be converted into an understandable language and describe the exact description of the system's behavior. Therefore, integrated modeling language is used to identify these indices. To illustrate the behavior, is also used the sequence diagram.

Charting the message sequence is a language that models interaction between components with processes and interactions between samples and the environment [1]. In this scenario-based model, communication between the samples is described as sending and receiving messages and local events as well as the ordering of them. This language, while defining the partial behavior of the system, imposes restrictions on the data values transmitted and the occurrence time. On the other hand, the graph is arranged for the graphic display. The lifetime of each sample is shown in the

vertical dimension of this graph and the message is positioned as a horizontal arrow between the sender and receiver [1]. Figure 5 shows the flowchart of the proposed Approach and the process of execution.

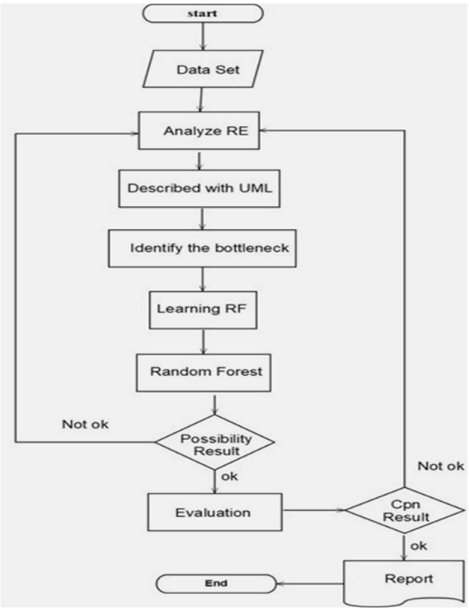

Figure 5. Proposed Approach Flowchart

3.2 Identification of Project Development Risks

In this section, according to the order of the diagram that describes the current behavior of the system, we analyze the documentation and identification of bottlenecks, in this way, the transactions of the components are specified that they described in the sequence diagram

And the components are considered as bottlenecks that make the most of the transaction and lead to queuing. In fact, to calculate the risk probability between the two components m and 1, if nms (1, m, j) is the number of interactions of these two components in the order of j, it is expressed by linteract (1, m, j). Therefore, using the probability of risk between two components (ψi), we calculate the reliability of ψmj) in accordance with equation (1) :

$$\psi_{1mj} = (1 - \psi i)^{linteract(1.m.j)} \quad (1)$$

So, we can say how much each component affects the risk. Of course, this diagnosis is only possible in the normal state of the system. So, for dynamic diagnosis, we need another technique or technique that follows.

3.3 Random Forest Algorithm

We obtained two types of data after identifying the parameters that influence the development of the studied projects. The data currently identified and recorded in the current process, and their results indicate that we give this data as input to the RF, so is performed the learning process of the algorithm. In the next step, the outputs are the data that identifies the system. In fact, the RF algorithm is used to categorize and predict possible risks based on the indicators defined here. Using algorithms, RF set of input parameters of the incoming initial (learning) are the expression X = (X1, X2, ... Xp) is a collection of P variable to predict the potential, Xj = (X1j, X2j, ... Xnj) T, and also assume that y is the under investigation attribute [15]. (In this research, there is a high-risk or non-occurrence risk) in which n is the number of machines in the sample. Also, B is the tree number and B is the total number of trees. We set the initial value b to one:

In step k, θk is an independent sample and is selected with the same distribution of the data set; this series is called learning or teaching. X is also a random sample of the set of predictive variables under study. The prediction function h (X, θk) is constructed using X and θk. Steps B Repeat to reach the tree number B. We assume in the proposed Approach to predict the risk or not, using the random forest algorithm:

$$P = \sum_{K=1}^{E} I\left(h(x.\theta k)\right) = 1\ ) \quad (2)$$

$$Q = \sum_{K=1}^{E} I\left(h(x.\theta k)\right) = 0\ ) \quad (3)$$

If P> Q, Random Fort predicts that x belongs to class 1 (in this risk study) and vice versa, if P <Q, RF predicts that x belongs to the class 0 (in this case Risk-free).

Algorithm 1 Random Forest

Precondition: A training set $S := (x_1; y_1); : : : ; (x_n; y_n)$, features F, and number of trees in forest B.

```
function RandomForest(S, F)
H ← ∅
for i 2 1; : : : ; B do
S(i) ← A bootstrap sample from S
hi ← RandomizedTreeLearn(S(i); F)
H ← H [ fhig
end for
return H
end function
function RandomizedTreeLearn(S, F)
At each node:
f ← very small subset of F
Split on best feature in f
return The learned tree
end function
```

3.4 Reliability Assessment

In fact, the purpose of evaluating this parameter is to mean that the system works correctly at the time interval [t0, t] Provided that the system is correct at the beginning of the interval (t0) and the system is expected to provide uninterrupted service, such as spatial applications. But in the system's availability at time (t)

(whenever needed), it is working properly and available and will do its job. The system of banks can be described as an example [9]. To evaluate reliability, is also used equation (4) [3].

$$f = [\, f \times (1-F)] \qquad (4)$$

In the above relationship, f is reliable and F is the fault rate or refractive index. After each time is calculated a reliability fault, from the new failure will be updated the average and reliability.

3.5 Create an Applicable Model

In this paper, the sequence diagram and Matlab and the CpnTools tools are used to create the applicable model. In fact, an applicable model of the current system is a formal description of the system through which one can evaluate the final behavior before implementing the target system and it was informed about the problems and inefficiencies and is more confident about the implementation of the new system and it avoids additional costs, even the failure.

4. Control Experiment

The used Case Study is a comprehensive and as small as possible example has been chosen from the student registration system for education. As well as the lack of complexity, as little as possible, is a small sample of other real and large examples. In this paper, to demonstrate the feasibility and the accuracy of the proposed Approach, we performed the research using the proposed Approach and simulation of the applicable model in the MatLab and CPN Tools. Figure 5 shows a description of the transition of the test to the student records system in sequence diagram**.**

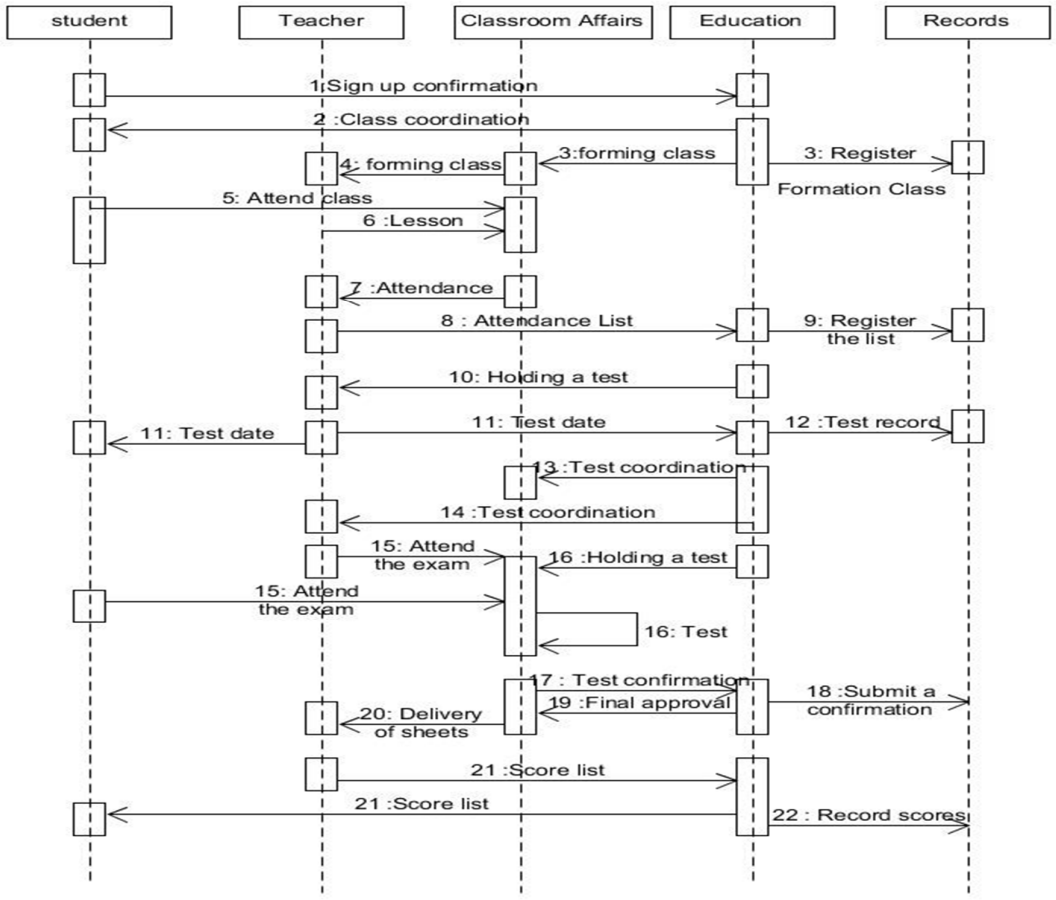

Figure 6 . Sequence Diagram: Description of the transition test from the student registration system

In accordance with the proposed Approach, the behavior of the system is described in Fig. 6. The project risks are Identified further. Table (1) shows the indicators that affect the emergence of risk. These indicators are extracted from the analysis of documentation and

behavioral descriptions. In fact, the success and failure of students in the exam is the main index of this section.

Table 1: Indicators affecting risk

Description	Property
Specifies the student number	Student code
Registered attendance list for each lesson	Attendance meetings
According to the scientific level of in 4 categories, respectively, from associates degree to Ph.D	Teacher level
Depending on the geographical location, it is shown in 4 sections	School Code
In four levels, weak, average, good and excellent respectively	Average
For students, the average grades in the class are shown in four levels	Placement
Formation and absence of a compensatory class	Compensatory workshop
Specifies the teacher's code	Teacher code
At three levels of quality, the grade is poor, good and excellent	Class type

With regard to the obtained indicators, we show the learning stage in the system using the RF algorithm.

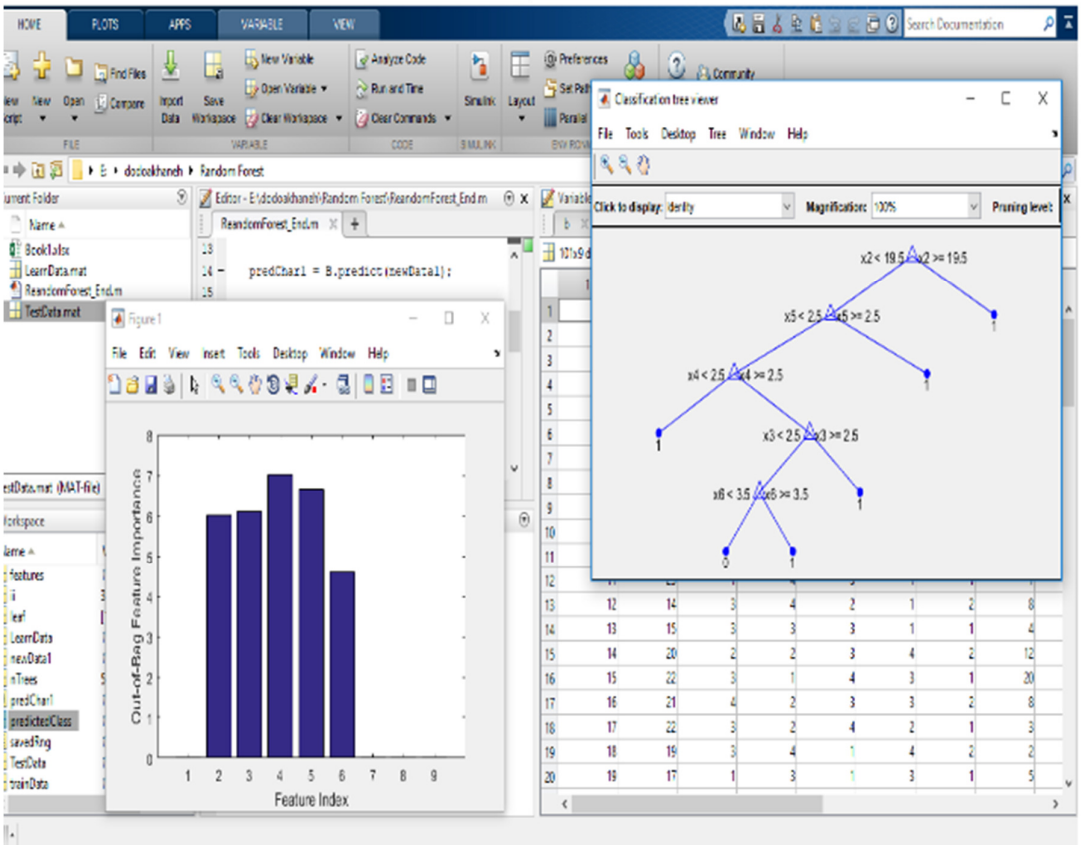

Figure 7. RF algorithm for learning and prediction

Due to the test carried out of 1500 students with the characteristics according to Table (1), the algorithm performed the learning action by examining the indicators and parameters and is also projected the test output per student. Figure 6 shows the output from the implementation of the algorithm. Figure 8 shows the effect of the extracted indices. According to the results of the proposed Approach, we evaluate the proposed Approach to examine its reliability and accuracy. In fact, this feature is also evaluated with the help of colored Petri Networks.

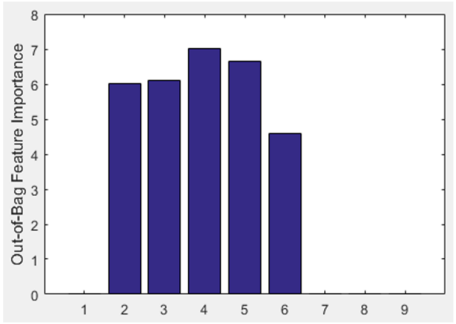

Figure 8. Effect of extracted indicators

In this paper, the results are evaluated with twenty, fifty, one hundred students to evaluate the reliability, accuracy and effectiveness of the proposed Approach. The results are shown below according to the proposed Approach and extracted features. Figure 9 shows the output of a model that is implemented by a colored Petri nets. The results of the three experiments are presented in Table (2).

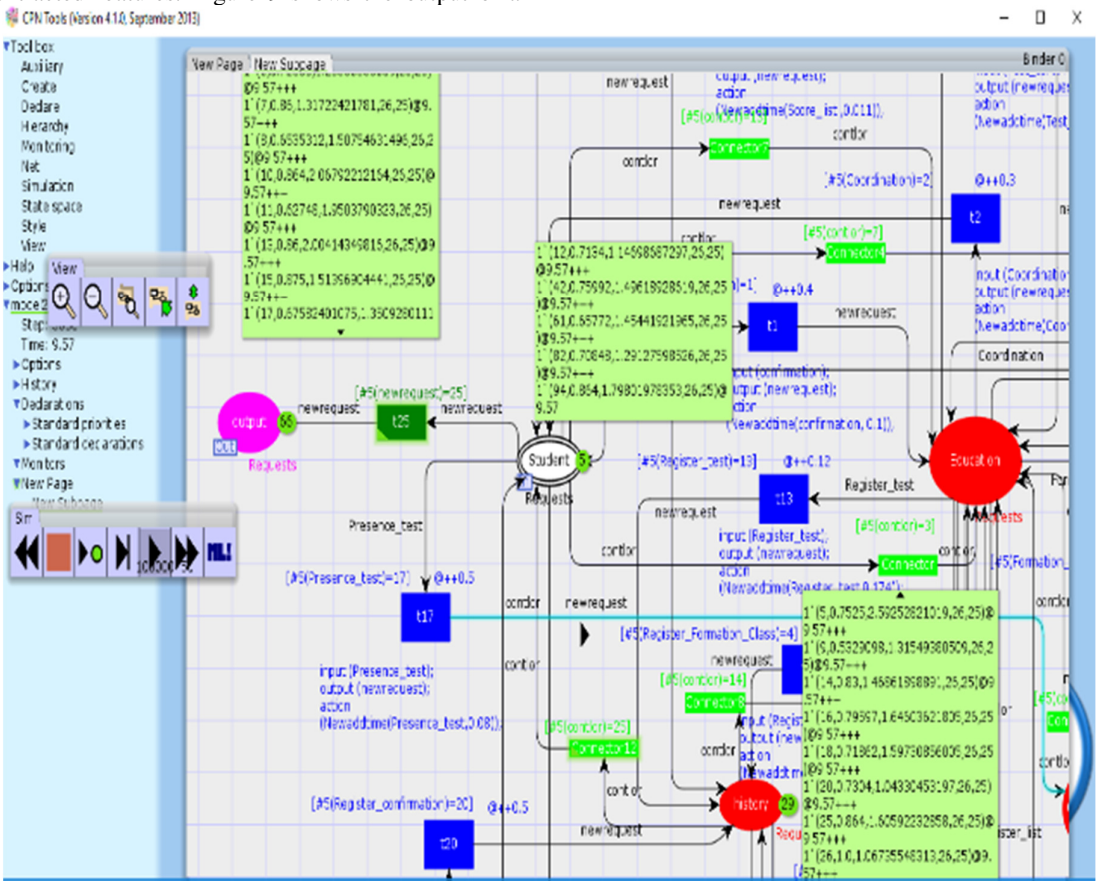

Figure 9 . Applicable model of the proposed Approach

Table 2: Average of all three different implementations

Number of student	*Reliability*	*Response time*
20	0.95557614	1.599106946
50	0.94964838	1.674033117
100	0.944333889	1.634504887

Table (2) shows the results of implementing the proposed Approach by the applicable model for the number of different users. Finally, the average reliability and response time are calculated from each performance and illustrated the results. The model that performed by the CpnTools is dynamic and has the ability to test with several different requests.

Figure 10 ,11 shows the average reliability and availability of this work in three different runs respectively. The proposed approach did not support Agile-Based Development methodology and not meet Learning based on Privilege and high Risk Condition.

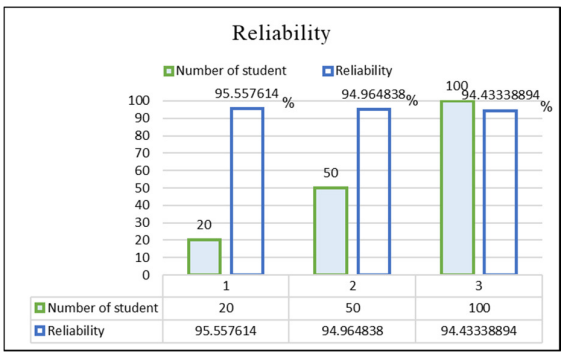

Figure 10. Average reliability with three runs

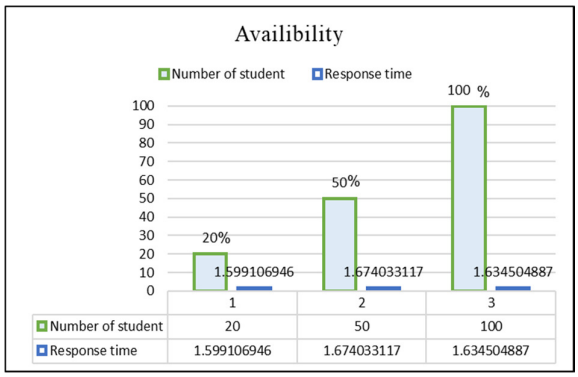

Figure 11. Average reliability with three runs

5. Conclusion and Future works

In this paper, an approach was presented based on requirements engineering in the development of software projects using random forest algorithm. The proposed approach using the documentation and analysis of the indicators affecting the risk of bottlenecks was identified with the help of random forest algorithm and reveals the effects of indicators in the software development process are based on the requirements engineering. A dynamic Petri-Nets based network model was proposed to assess and evaluate the reliability and availability of the proposed approach. The results of the three different runs indicate that the proposed approach has high reliability and availability for development in software projects. The disadvantages and shortcomings of this approach are the lack of support for the agile software development methodology and the high response time. Future works are focused on supporting different software development methodologies and also the study of privileges indicators with high power of learning.

References

[1] Allen, R., Douence, A. ,"Specifying Dynamism in Software Architectures", Journal of Systems Engineering ,Vol. 6, No. 4, pp. 52-94,2007.

[2] B. Leo. "Random forests." Machine learning 45.1, 5-32.2001.

[3] Farjaminejad.F and Harounabadi.A, "Modeling and Evaluation of Performance and Reliability of Component-based Software Systems using Formal" International Journal of Computer Applications Technology and Research Volume 3– Issue 1, 73 - 78, 2014.

[4] Gharooni fard, G., Moein darbari, F., Deldari, H., Morvaridi, A., "Scheduling of scientific workflows using a chaos- genetic algorithm", Procedia Computer Science, Elsevier, Vol. 1, No.1, pp. 1445- 1454, 2010.

[5] Jensen, K., "Colored Petri Nets: Basic Concepts, Analysis Methods and Practical Use", EATCS Monographs on Theoretical Computer Science, Vol. 29, No .2, pp70-120,2013.

[6] Jörg, H., Ruslan, B., Matthias, M., Schmelter, D., Tschirner, C.,"Integrated Systems Engineering and Software Requirements Engineering for Technical Systems", Tallinn, Estonia,,ACM, 978-1-4503-3346-7/15/,2015.

[7] Kryftis, Y., Mastorakis, G., Mavromoustakis, C., Mongay Batalla, J., Pallis, E. and Kormentzas, G.,"Efficient Entertainment Services Provision over a Novel Network Architecture". To be published in IEEE Wireless Communications Magazine.2016.

[8] M. Abdel-Aty, and K. Haleem. "Analyzing angle crashes at unsignalized intersections using machine learning techniques." Accident Analysis & Prevention 43.1,461-470,2011.

[9] Motameni, H., Movaghar, A. Siasifar, M.,"Analytical evaluation on Petri net by using Markov chain theory to achieve optimized Model". World Appl. Sci. J. 3 (3) 504–513,2008.

[10] Onesmus Emeka, B., Liu, Sh.,"Security Requirement Engineering using Structured Object-oriented Formal Language for M-banking Applications", IEEE International Conference on Software Quality, Reliability and Security, 978-1-5386-0592-9/17,2017.

[11] Parra, E., Dimou, C., Llorens, J., Moreno, V., Fraga A., "A methodology for the classification of quality of requirements using machine learning techniques", Information and Software Technology, elsevir,180-195,2015.

[12] Prihartini,N., Laksmiwati, R., Rendradjaya, B.,"Identifying Aspects of Web e-Leaming in LMS-based for Requirement Engineering Process Modeling", IEEE Future of Software Engineering,978-1-5090-5671,2016.

[13] Sitthithanasakul, S., Choosri, N.,"Using Ontology to Enhance Requirement Engineering in Agile Software Process", International Conference on Software, Knowledge, Information Management & Applications (SKIMA),IEEE, 978-1-5090-3298-3,2016.

[14] Yaniv, M., Dov, D.,"Model-Based Requirements Engineering: Architecting for System Requirements with Stakeholders in Mind", IEEE,978-1-5386-3403-5/17,2017.

[15] Yu. Rongjie, and M. Abdel-Aty. "Analyzing crash injury severity for a mountainous freeway incorporating real-time traffic and weather data." Safety science 63 ,50-56,2014