**INTELLIGENT SOFTWARE PROJECT MANAGEMENT SYSTEM**

Rathnayaka Gamage Hasith Kaushan Wijewickrama

IT13033538

Bachelor of Science Special (Honors) Degree in IT

Department of Information Technology

Sri Lanka Institute of Information Technology

Sri Lanka

September 2016

**INTELLIGENT SOFTWARE PROJECT MANAGEMENT SYSTEM**

Rathnayaka Gamage Hasith Kaushan Wijewickrama

IT13033538

Dissertation submitted in partial fulfillment of the requirements for the B.Sc. Special Honors Degree in IT

Department of Information Technology

Sri Lanka Institute of Information Technology

September 2016

# **Declaration**

“I declare that this is my own work and this dissertation1 does not incorporate without acknowledgement any material previously submitted for a Degree or Diploma in any other University or institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Also, I hereby grant to Sri Lanka Institute of Information Technology the nonexclusive right to reproduce and distribute my dissertation, in whole or in part in print, electronic or other medium. I retain the right to use this content in whole or part in future works (such as articles or books).

Signature: Date:

The supervisor/s should certify the dissertation with the following declaration.

The above candidate has carried out research for the B.Sc. Special (Hons) degree in IT/SE/CSN Dissertation under my supervision.

Signature of the supervisor: Date:

# **Abstract**

Software project management has become a crucial and vital task today since it plays a major role in software companies. It consists of a number of planning, organizing, staffing, directing, and controlling activities.

Human resources feature prominently in all of these activities and as a consequence they can affect and determine project management decisions. Therefore in order to guarantee the sustainability of the software project, managers must take into consideration this type of resource when performing the aforementioned activities. This document specifically investigates how the automated system is handling human resource allocation for a certain project and in particular focuses on the responsibilities of allocated developers. These responsibilities are often challenging to undertake because they are accompanied by time, budget and quality constraints, which software project managers find difficult to balance correctly. Here it is precisely mentioned how the automated system is going to generate predictions. What are the appropriate data mining algorithms and techniques that can be used to increase the accuracy level is briefly mentioned and also how to process a clean data set. A large proper data set is needed to maintain the accuracy level of predicted development team. So then the data set must be and has to be a valid and well organized format. Suitable team of developers for a certain project is always brings a successful project into reality ultimately. Project manager doesn’t need to worry on this task since system is going to handle it by itself and if project manager involves in this then he needs to go through many documents to analyze the skill levels of personnel to find the right team for a project. It is more costly and time taken traditional approach to master this and not perfect to proper project planning.

The purpose of this document is to find the right personnel to software projects and provide a supportive tool to better project planning. This perspective, in particular, sheds light on current and future trends, which lean towards incorporating human-centric aspects of software development in planning activities.

# **Acknowledgement**

The special thank goes to our helpful supervisor Mr. Tharindu de silva, who was given his fullest support and supervision throughout the entire project. Invaluable assistance was provided by Mr. Thilina Halloluwa to make our project a success. Mr. Jayantha Amararachchi for providing final years’ students with the knowledge that’s necessary for writing and presenting effective project progress reports. Our grateful thank goes to all the personnel of the Sri Lanka Institute of Information Technology who contributed essential information. Our heart full love goes to our dear parents for their encouragement throughout the project. Where this project succeeds, we share the credit with all. Furthermore, a special gratitude to our group members who supported me and invested their time and effort in the team towards achieving the goal.

Finally, a special thank goes to my family for their constant encouragement without which this project would not have been possible.

Thank you.

**Table of Contents**

[Declaration i](#_Toc465419216)

[Abstract ii](#_Toc465419217)

[Acknowledgement iii](#_Toc465419218)

[List of Tables v](#_Toc465419219)

[List of Figures vi](#_Toc465419220)

[List of Abbreviations vi](#_Toc465419221)

[1 INTRODUCTION 1](#_Toc465419222)

[1.1 Background context 2](#_Toc465419223)

[1.2 Research gap 9](#_Toc465419224)

[1.3 Research problem 11](#_Toc465419225)

[1.4 Research objectives 12](#_Toc465419226)

[2 METHODOLOGY 15](#_Toc465419227)

[2.1 Methodology 15](#_Toc465419228)

[2.1.1 Feasibility Study 17](#_Toc465419229)

[2.1.2 Generate project team 20](#_Toc465419230)

[2.2 Testing and implementation 25](#_Toc465419231)

[2.2.1 Unit Testing 27](#_Toc465419232)

[2.2.2 Integration Testing 27](#_Toc465419233)

[2.2.3 System Testing 27](#_Toc465419234)

[2.2.4 Building the Classifier or Model 28](#_Toc465419235)

[2.2.5 Using Classifier for Classification 29](#_Toc465419236)

[2.2.6 Classification and Prediction Issues 30](#_Toc465419237)

[2.3 Research findings 32](#_Toc465419238)

[2.3.1 Identify the appropriate classification algorithms 32](#_Toc465419239)

[2.3.2 Identify suitable language for data analysis 32](#_Toc465419240)

[3 RESULTS AND DISCUSSIONS 32](#_Toc465419241)

[3.1 Results 32](#_Toc465419242)

[3.1.1 Test Approach 41](#_Toc465419243)

[3.2 Discussion 44](#_Toc465419244)

[4 CONCLUSION 45](#_Toc465419245)

[References 46](#_Toc465419246)

[Glossary 49](#_Toc465419247)

[Appendices 50](#_Toc465419248)

# **List of Tables**

[Table 1 : Comparing existing and suggested system 10](#_Toc465420190)

[Table 2:Insert the project details 33](#_Toc465420191)

[Table 3:Select the job roles of the employees 34](#_Toc465420192)

[Table 4:Provide technologies that are needed to the software project 34](#_Toc465420193)

[Table 5:Upload documents which includes the details of the software project 35](#_Toc465420194)

[Table 6:Upload employee details including each employee’s skills and basic details 35](#_Toc465420195)

[Table 7:Retrieve the absolute file paths from the uploaded documents 36](#_Toc465420196)

[Table 8:Extract data from the uploaded documents and store in the database 36](#_Toc465420197)

[Table 9:Select the project that need to predict the team 37](#_Toc465420198)

[Table 10:Display a success message 37](#_Toc465420199)

[Table 11:Display an error message 38](#_Toc465420200)

[Table 12:Team selection attributes 39](#_Toc465420201)

[Table 13:Algorithm selection 39](#_Toc465420202)

[Table 14:Train the data set for Support vector machine algorithm 40](#_Toc465420203)

[Table 15:Train the data set for K-Nearest neighbors algorithm 40](#_Toc465420204)

[Table 16:Train the data set for Naïve Bayes algorithm 41](#_Toc465420205)

[Table 17:Predicted team which is generated by the algorithm stored in the database 41](#_Toc465420206)

[Table 18:Load the selected developers for the team in a graphical view 42](#_Toc465420207)

# **List of Figures**

[Figure 1 : Triple constraints of a project 4](file:///D:\IT\4thYear\Docs\final%20docs\my%20doc%20-%20IT13033538.docx#_Toc465419309)

[Figure 2: Software development life cycle 5](file:///D:\IT\4thYear\Docs\final%20docs\my%20doc%20-%20IT13033538.docx#_Toc465419310)

[Figure 3 : Work Break Down Structure 16](file:///D:\IT\4thYear\Docs\final%20docs\my%20doc%20-%20IT13033538.docx#_Toc465419311)

[Figure 4 : Extracting knowledge from data 22](file:///D:\IT\4thYear\Docs\final%20docs\my%20doc%20-%20IT13033538.docx#_Toc465419312)

[Figure 5 : Sample Data model 24](#_Toc465419313)

[Figure 6 : Team build function implementation structure 28](file:///D:\IT\4thYear\Docs\final%20docs\my%20doc%20-%20IT13033538.docx#_Toc465419314)

[Figure 7 : Building the Classifier or Model 29](#_Toc465419315)

[Figure 8 : Using Classifier for Classification 30](#_Toc465419316)

[Figure 9 : Support vector machine 42](#_Toc465419317)

[Figure 10 : Use case scenario 51](#_Toc465419318)

[Figure 11 : Insert project details 52](#_Toc465419319)

[Figure 12:Job roles selection 52](#_Toc465419320)

[Figure 13:Provide technology 53](#_Toc465419321)

[Figure 14:Upload employee data 53](#_Toc465419322)

[Figure 15:Project selection 54](#_Toc465419323)

[Figure 16:Team selection criteria 54](#_Toc465419324)

[Figure 17:Algorithm selection 55](#_Toc465419325)

[Figure 18:Predicted team for the software project 56](file:///D:\IT\4thYear\Docs\final%20docs\my%20doc%20-%20IT13033538.docx#_Toc465419326)

# **List of Abbreviations**

|  |  |
| --- | --- |
| Acronym | Definition |
| ISPM | Intelligent Software Project Management |
| XML | Extensible Markup Language |
| SRS | Software Requirement Specification |
| OWL | Ontology Web Language |
| UI | User Interface |
| JDK | Java Development Kit |
| SWRL | Semantic Web Rule Language |
| RAM | Random Access Memory |
| SPARQL | Simple Protocol and RDF Query Language |
| SPM | Software project management |

# **INTRODUCTION**

The main objective of any software company is to provide quality software to its customers. The best of software system is bound to fail without the right people working on it. Even though a talented, skillful team has selected if they are not guided properly then also the whole project will become a failure. Hence there should be an intelligent, good decision making, enthusiastic project manager to do the job. But when it comes to larger projects and bulk of developers to select for the team and analyze set of documents project managers cannot perform their duties effectively and efficiently.

Our system which we called it as Intelligent Software Project Management comes to play its role in this point. It is actually a tool to handle the software projects in a very precise manner. No need of worrying analyzing larger set of documents, information extracting, finalize software requirements, predicting a software development team, managing the project to the end, track the tasks of the developers; system itself is going to handle these crucial and serious tasks. Not only that but also it is performing these tasks automatically. Ultimately what will happen is our system is the software project manager for the company.

This dissertation is submitted in partial fulfillment of the requirements for the degree of Science as a draft so that the current status of the research is well documented for evaluation purposes as well as documentation purposes. It covers the basic requirements of a thesis including the research background, the identification of the research gap and problems, significance of the research, the procedure carried out in testing and implementation and etc. It also reveals the findings and results of the research so that it maps back to the research problems with the provided solutions. Even though it is a single research, there are several research areas being surfaced during the journey of it.

The main research components associated with the system are Information Extraction, Ontology based decision making, Predictive analysis. At present, the actual researching solutions have been identified and we are in the process of implementing the product while introducing new features.

## **Background context**

Software is a direct product of the cognitive processes of individuals engaged in innovative teamwork. Many of the procedures and techniques used in software project management are designed to facilitate communication and coordination among team members engaged in an intellectually intensive work. Software development is often characterized as a learning process in which knowledge is gained and information generated during the project. Dealing with people and conflicts, team building, knowledge sharing, and communication will be the determinants of good software project management.

The principal nature of the challenges in software project management has not changed dramatically in the last 25 years. However, software-intensive systems of the twenty-first century increasingly vary in their content, size, complexity, and their degree of interaction with other systems. The technological and communication infrastructure to develop these systems is hard to compare with that available in the past. As a consequence, the concrete content of the project management challenges looks different from that of 25 years ago [1].

A Software Project is the complete procedure of software development from requirement gathering to testing and maintenance, carried out according to the execution of methodologies, in a defined period of time to achieve the specified software product. Software is defined as an intangible product. Software development is a method of new trend in world business and there’s a very little experience in developing in software products. The most important is that the defined technology changes and advances so quickly that experience of one product will not be applied to the other one. All such business and the developing environment bring risk in software development since it is essential to manage software projects in an efficient manner.

Software project management deals with software projects and the challenges of human-based development (as opposed to the more deterministic processes in traditional projects). The higher flexibility in software development approaches puts new demands on the capabilities of software project management. Weaknesses in planning, organizing, staffing, directing, and controlling are hard to be counterbalanced by more efficiency in technical development work. As Fred Brooks stated in 1987, “... today’s major problems with software development are not technical problems, but management problems” (Brooks 1987) [2].

Software development is both human-intensive and knowledge-intensive, which makes people the most important asset in any software development endeavor.

Software projects are different from other projects in a number of ways. Consequently, management of software projects cannot be done in the same way as in traditional project management and needs to be adjusted correspondingly [2]. Following

(PMI 2013b), some of the main differentiating factors are as follows:

* Software is an intangible product.
* Software is a cognitive and human-based development process that requires sharing of documents.
* There is a higher degree of uncertainty in the project and product scope.
* Communication and coordination within software teams and with project stakeholders often lacks clarity.
* The intellectual capital of software personnel is the primary asset of software projects and organizations.
* There is a degree of change of requirements in the course of the software project.
* The creation of software requires innovative problem solving to create unique solutions.
* Initial planning and estimation of software projects is challenging because these activities depend on requirements that are often imprecise or based on lacking information.
* The development and evolution of software-intensive systems is challenging because of the high complexity of software based on the enormous number of logical paths in program modules and all the combinations of interface details.
* Objective measurement and quantification of software quality is difficult.
* Learning and knowledge creation in software development is more difficult because processes, methods, and tools are constantly evolving.

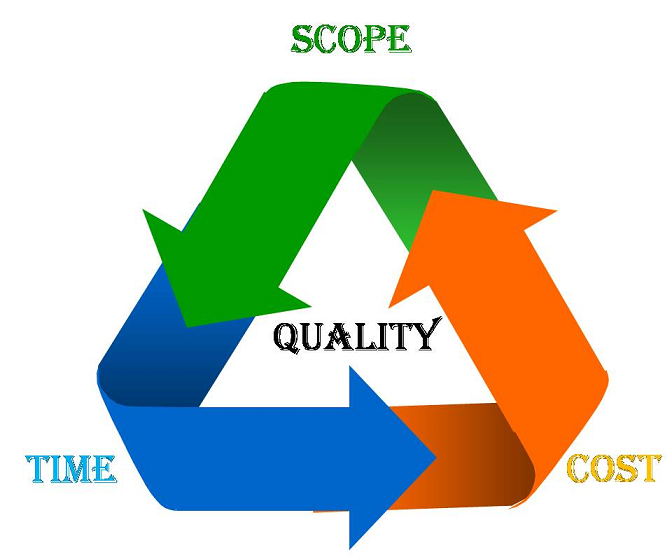


Figure 1 : Triple constraints of a project

The image above shows triple constraints for software projects. It is an essential part of software organization to deliver quality product, keeping the cost within client’s budget constrain and deliver the project as per scheduled. There are several factors, both internal and external, which may impact this triple constrain triangle. Any of three factor can severely impact the other two [2].

Factors such as poor requirements, lack of management support, and customers and users who do not make themselves readily available significantly affect software development success. The earlier in the development process problems arise the more serious is the outcome [3]. When there are limited resources, a project manager needs to know which factors are likely to have the most severe impact on the project. In order to investigate some of these early factors and their effect on the success or failure of software projects we develop an organizational case study.

There are several steps involved in the software development lifecycle. Software Development Life Cycle, SDLC for short, is a well-defined, structured sequence of stages in software engineering to develop the intended software product. They are:

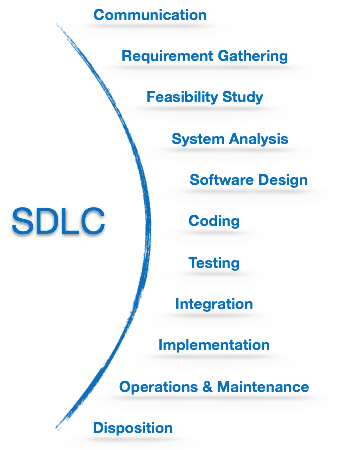


Figure 2: Software development life cycle

These steps should be managed and tracked accurately in order to come up with a better software product. That is where software project manager going to play his role. He should be the person to guide the development team to the correct path to achieve project goals.

A well-qualified project manager is able to address what may seem to be an overwhelmingly complex process by developing an organized approach where the process is broken down into manageable individual tasks and understanding how to keep those involved in the project dedicated to the ultimate goal of meeting and even exceeding the expected end result.

For any software development or other project initiative your company may be considering, it is critical to have in place and practice a set of effective and proven guidelines to ensure project success and delivery of the expected results: taking into consideration the role and responsibilities of a well-qualified project manager, knowledge of important business and financial aspects, and a step-by-step process that all contribute to the solid foundation and implementation of an effective project plan [4].

If the project manager dedicates the necessary time, effort, and resources to the preparation of an efficient, comprehensive, and practical approach, then the project team may progress with ease and confidence as they deliver on their individual tasks, having a solid foundation and strategic framework at the outset. Far too often, however, failures with such projects are the result of not only a poorly executed plan, but one that ultimately lacked the fundamental elements of a well-though-out approach rooted in adequate preparation and commitment from the project manager and project team.

Let us see few responsibilities that a project manager shoulders:

Managing People

* Act as project leader
* Liaison with stakeholders
* Managing human resources
* Setting up reporting hierarchy etc.

Managing Project

* Defining and setting up project scope
* Managing project management activities
* Monitoring progress and performance
* Risk analysis at every phase
* Take necessary step to avoid or come out of problems
* Act as project spokesperson

On a typical IT project, the overall project manager has to manage many different disciplines and functional areas that eventually have to work together for the success of the project. Compound this multi-disciplined deliverable set is always on the critical path, and it is easy to see why a project manager is extremely important to the success of the project [5]. Hence the tasks which is done by a project manager should be properly managed, well organized and most importantly should meet the deadlines and finally completed on time.

Automation is key to cost-savings, however; because of the automation capabilities, less employee time is spent on project management, requiring either less staff or allowing staff to focus on other, more profitable tasks [6]. Automation also means that information is available all the time. Instead of having to ask someone a question about a task, the user can go into the program and see the status. That way, if someone is on vacation, ill or tied up in meetings, the project still can continue to run.

So many projects, so much mismanagement. That's the refrain of many IT executives. Indeed, even with project management software, IT projects often wind up taking longer (much longer) than planned and costing more than budgeted.

The most famous research on prediction was done by Philip Tetlock of the University of Pennsylvania, and his seminal 2006 book Expert Political Judgment provides crucial background [7]. However several people has done researches on prediction and they came with absolute results. According to our research we are going to make predictions through the system itself. Normal scenario of making predictions in software project management is project manager will go through past completed projects and make decisions depend on them. It is more time consuming and not a realistic method. Even though there are some software project management tools there is no such a proper tool where we can get all kind of valuable predictions that we must know.

PMI’s Pulse of the Profession™ research, which is consistent with other studies, shows that "less than two-thirds of projects meet their goals and business intent (success rates have been falling since 2008), and about 17 percent fail outright. Failed projects waste an organization’s money: for every US$1 billion spent on a failed project, US$135 million is lost forever…unrecoverable." [8].We can clearly see that how making successful predictions are going to help to complete the project successfully and also keep the reputation of the company.

We now know that software project manager plays a key role when managing the project in the entire time period and he should bear the weight of the project ultimately. So what we are trying to achieve here is come up with a software project management tool as a replacement to the software project manager. Since it is an automated tool, in most crucial tasks like predicting the development team for a particular project through the system itself companies can generate the team. So our ultimate goal is to automate the major crucial tasks of a software project manager when managing the software projects and help to come up with a successful user friendly quality software product and deliver on time to the users.

## **Research gap**

Typically project managers in IT field has to face different kind of critical problems when they are dealing with software projects. IT projects are notoriously difficult to manage, and failure is an all too common outcome. The right IT project management software can help to manage and process the software projects in an effective and efficient manner even though they are large in size. A 2012 study by McKinsey & Co. and the University of Oxford tells a cautionary tale of large IT projects that are severely troubled: “On average, large IT projects run 45% over budget and 7% over time, while delivering 56% less value than predicted,” McKinsey reported gloomily, focusing on projects of $15 million or more [9].

As we all know generally a project manager has to involve in the project from the very beginning and manage it through the entire time period until it is completed. So, then his time is wasted unnecessarily and that is also a cost to the company too. Likewise, still project managers have to undergo several obstacles whether they like them or not. Currently there are several researches have been done related to project management and but most of them have limitations in some aspect. None of them are not implemented to do the project management process fully automatically. So here what we are trying to achieve is bridge the gap to automate the whole process of managing the project.

Current software project management tasks such as finalizing the development team for a particular project can be a crucial task when there are so many developers in the company and we can’t assign them directly without considering their past project success rates. Hence project managers have to undergo their past historical data and analyze them before assigning them for a certain task. That is very time consuming and traditional method of achieving that task. So, there is a need of generating the development team for a project more quick and accurate manner according to their performances.

Different types of software companies use various types of SPM systems to increase the efficiency and accuracy of their software product. Most of the SPM systems are working in manual way. As the example ‘Wrike’, ‘Huddle’, ‘Podio’, ‘goplan’ are the main SPM systems. Most of these are provide manual operation of SPM system [10]. In a particular company run a set of software projects at a time and needs inputs from teams or group of individuals for a multilevel development plan. Hence a good automated project management system is needed.

This automated SPM system will enter a major role in a large number of companies. Therefore, automated SPM system gives high quality application for customers and it will helps ensure the durability too. Developing a web based automated project management system helps users to handle projects in a convenient way without the help of any project manager. That is the main goal of this research.

Great people, people with sufficient functional skills and domain expertise can trump process, good or bad. Good process, process appropriate for the context, will help those people. But great people can overcome bad process to deliver a good product. Hence properly assigned development team for correct tasks is always leads a project to success. As mentioned before there’s a huge gap between existing software project management tools and our proposed system. There are no any accurate team prediction mechanisms existing currently. In our proposed system, most efficient classification algorithms are used to predict the development team and in a well-organized, clarified interface it’s going to view the predicted team. Our team believes that by implementing this team build function software companies can achieve numerous advantages and benefits since dynamically system itself assign the development team for a software project.

According to our team’s perspective we have identified major gaps between existing software project management tools and proposed system, and they are briefly mentioned below.

Table 1 : Comparing existing and suggested system

|  |  |  |
| --- | --- | --- |
| Features | Current Applications | ISPM |
| 1. System runs without having a software project manager | X |  |
| 1. User can provide the requirements to the system easily | X |  |
| 1. Scanning and extract user requirements | X |  |
| 1. Maintain the knowledge based system for extracting appropriate data | X |  |
| 1. Identifying and avoiding the similar user requirements. Finalize the unambiguous user requirements | X |  |
| 1. Predict the development team for the project , success rate | X |  |
| 1. Defining the milestones for each and every tasks and visualize the progress of each and every task | X |  |
| 1. Sending emails, notifications to the users. Provide user friendly and attractive environment |  |  |
| 1. Users (sponsors, clients and developers) can access to the system and gain the progress of each and every task |  |  |

## **Research problem**

As mentioned earlier managing a software project and directing a development team to accomplish their tasks are not an easy task. A software project manager should be with the group to lead and supervise the project. He must be a skillful, talented person to handle any kind of project. So, what will happen if he is not capable of doing that? And let’s assume in the middle of a project, project manager has to leave the company due to unavoidable circumstances. So, what will happen next? If a new project manager has been recruited to the company, then it takes some time to him to adjust to the current progress of the project. So, it is time wasting and sometimes it can reduce the success rate of the project or even get fail. So, that companies need to rely on their project managers.

When selecting a development team for a software project how can we measure they are the right people for the project. If unsuitable people assigned to handle the project ultimately project will become a mess. Company’s reputation depends on the success of the project. Therefore, to complete a project successfully righteous people should be assigned to perform the tasks. And that is a major problem most of the IT companies are facing today.

When considering the team generation for a particular project it is obviously a critical task. To properly generate a balanced skillful development team for a project lot of facts need to be considered. In some cases bulk of documents have to be read to find the right people for a project. Most of the time it is so hard to analyze on what skills they are capable of. The above-mentioned things are done by manually by project manager. This improvement needs to be more reliable, efficient, secured and consistent. Better forecasting mechanisms should be taken into action to find the right personnel for a project. By using our product both clients and software developers don’t need to be afraid of the management of the project since from the very beginning onwards system itself going to handle the most critical tasks and mitigate the failures.

## **Research objectives**

**Main Objectives**

* Requirement Gathering

In here, requirements are gathered using the documents provided by the client and using text areas reserved for clients on the web application new requirements are gathered. What we are trying to emphasize is well structured template is given to the clients to capture software requirements in a well-organized manner.

* Resource Management

Resources management are done using allocating them to the correct places at correct time calculating in what amounts they have to be allocated. Mainly human resources are managed properly since in the team build function employees are assigned to a particular project based on their skills. By allocating right people fir the right tasks management of human resources is increased insidiously. And also notifying them within the assigned time period for a particular task is also a tremendous effort to manage human resources.

* Efficiency

Efficiency is handled using continuously monitoring the work done by employees and monitoring how budget is moving. When the tasks assigned for each developers are tracked properly efficiency of the project is getting higher. When the team for a particular project is generated through the system itself no need of analyzing documents manually and ultimately efficiency is maximized.

* Cost Reductions

Cost spent on project managers are reduced to a great extent since most of the critical tasks which are done by software project managers are minimizing because of the proposed system. No need to hire special skillful people to manage the projects and guide the project and even in assigning a team.

* Time Management

Time spent in requirement gathering and analyzing is reduced, time spent on user stories creating is reduced. Since there is an ontology engine is implemented in the system we can directly get what we need. Team build function helps to create a team in less than few seconds by analyzing all of the employee’s skills.

* Risk Management

Risk Management is done using taking better decisions. By assigning a perfect team for a project unknowingly success rate of the project is increased and risk is mitigated since obviously the project depends on the people who are going to handle. Large projects always come with higher risks and here it can be minimized up to a satisfactory level.

* Task tracking

Tasks which are assigned for the developers by the system are checked periodically by the system itself and send notifications automatically for each and every developers mentioning their remaining time period for the task.

* Introduce Team Management Tool

Introduce an intelligent system to decide the team composition based on the team’s historical outcomes and apply this system to compose project teams in IT industries. The system relies on historical data of the procedures performed in the past. Depending on the project characteristics given an optimal or feasible team will be generated. The optimal team composition is the one with the lowest probability of unfavorable outcomes an optimal solution is the theoretically proven solution. But it might not be the logically suitable solution and we might have to come up with the feasible team. Hence the tool has the option of providing the most feasible (possible and practical) solution as well.

**Specific Objectives**

* Learning about text mining and coming up with methods and algorithms to summarize data provided in the documents, find specific areas when stakeholders provide key words and matching the requirements provided with the output user stories from the ontology engine.
* Learning about ontology mechanism and coming with algorithms to make ease of the work happening inside the ontology engine, in here requirements gather from diagrams such as use case stories, activity diagrams etc. provided by the developers are analyzed and output user stories are sent to the text mining system to match with its requirements.
* Learning about predicting mechanisms, in here data from text mining system and ontology engine is taken as inputs to the predicting system. Employee details including their technology skills are filtered through prediction algorithms to find the best development team.
* Build a solution to select most optimal team for IT industry

ISPM will generate the most optimal team for IT industry considering attributes related to the IT sector. Depending on the project characteristics and the relationships between them most optimal team will be selected. The prediction is expected to be highly accurate and final decision is represented as a simulation.

* Build a solution to select most feasible team for IT industry

ISPM will generate the most feasible team for IT industry considering attributes related to the IT sector. Depending on the project characteristics and the relationships between them most feasible team will be selected. The prediction is expected to be highly accurate and final decision is represented as a simulation.

# **METHODOLOGY**

## **Methodology**

This section includes detailed descriptions about the techniques and mechanism employed to make ISPM a reality. The descriptions include how software implementation of our project is carried out, what are the materials and data needed, and how they will be collected. It also includes time frames and schedules that are required in achieving its objectives. In addition to them, the research areas that we have identified in order to carry out this project are explained rationally.

Automated project management system represents a fast-growing technology in IT field. With the involvement of users, who utilize the project management applications helps to build, web based project management system enter a major role in a large number of companies. Therefore, automated project management systems give high quality web applications for customers and it will helps ensure the durability too. Developing a web based automated project management system helps users to handle projects in a convenient way without the help of any project manager. The reliability and the robustness of the automated web based project management system offers lots of advantages to the organization or user who is willing to use our product.

When considering the team build function in the project there it is going to predict the most accurate team for a software project. First of all, details such as how many developers needed to the project, what are their job roles, what are the using technologies are given as an input to the system. And also, employee details including their individual skills and performances are also submitted. Finally, user can select the particular algorithm to predict the development team based on their accuracy levels which are support vector machine, k nearest neighbors, naïve bayes. After that most suitable team for the project is predicted and system itself manage the project until to the end.

**ISPM**

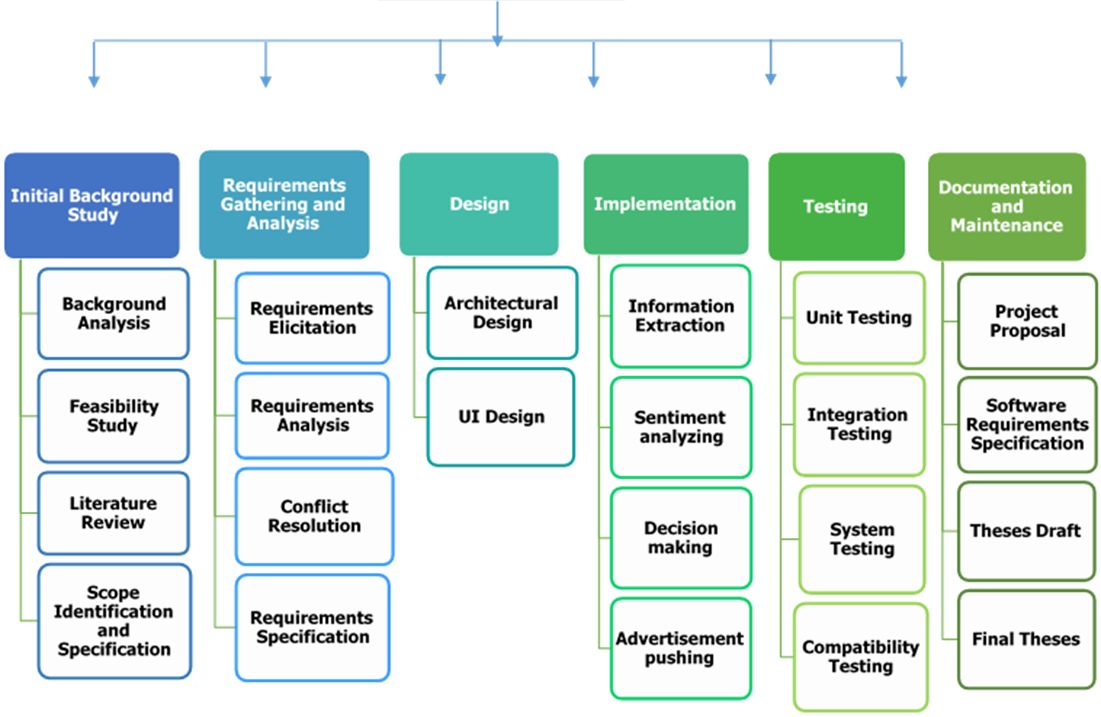


Figure 3 : Work Break Down Structure

### **Feasibility Study**

We went through an investigation into the goals and implications of the project. For very small scale projects this may not be necessary at all as the scope of the project is easily understood. In larger projects, the feasibility may be done but in an informal sense, either because there is not time for a formal study or because the project is a “must-have” and will have to be done one way or the other. Since this is a complex system we had to go via a feasibility study.

When a feasibility study is carried out, there are four main areas of consideration:

Technical – Is the project technically possible?

Financial – Can the team afford to carry out the project?

Organizational – Will the new system be compatible with existing practices?

Ethical – Is the impact of the new system socially acceptable?

To answer these questions, the feasibility study is effectively a condensed version of a fully blown systems analysis and design. The requirements and users are analyzed to some extent; some business options are drawn up and even some details of the technical implementation. The product of this stage is a formal feasibility study document. Document created using the feasibility study specifies the sections that the study should contain including any preliminary models that have been constructed and also details of rejected options and the reasons for their rejection.

Investigation of the current environment

In almost all cases there is some form of current system even if it is entirely composed of people and paper. Through a combination of having interviews with employees, circulating questionnaires, observations and existing documentation, the analyst comes to full understanding of the system as it is at the start of the project. This serves many purposes.

There are many similar products. Some of them are ‘Manage Engine Service Desk Plus’, ‘ZOHO Projects’, ‘Time Camp’, ‘Easy Red mine’, ‘exo Platform’ etc.

Manage Engine Service Desk Plus and ZOHO Projects

Manage Engine Service Desk plus is a completely web-based Help Desk and Asset Management Software. It offers an integrated package with Incident management (Trouble Ticketing), Asset Tracking, Purchasing, Contract Management, Self-Service Portal, and Knowledge Base at an affordable price point. Service Desk Plus provides all that you need to have a full-fledged IT Help Desk and a productive Help Desk staff. Manage Engine Service Desk plus is available in both normal edition and also ITIL edition [11].

Zoho Projects is the project management software from Zoho, a brand that enables 15 million users to work online. Businesses large and small, from every industry use the app to deliver great work on time. Plan your projects, assign tasks, communicate effectively, never miss an important update and view detailed reports on progress. You can add unlimited users on all plans at no extra cost [12].

Above 2 Project Management tools are the most famous project management tools in this year. As described above those are the basic functionalities of those project management tools. In our tool, there will be requirements mapping with tasks which is done using a larger data set and an ontology system and also information searching, analyzing and summarizing is available. Generation of the project team is also a major functionality in our system.

Business System Options

Having investigated the current system, team decides on the overall design of the new system. To do this, using the outputs of the previous stage, develops a set of business system options. These are different ways in which the new system could be produced varying from doing nothing to throwing out the old system entirely and building an entirely new one. The team may hold a brainstorming session so that as many and various ideas as possible are generated.

The ideas are then collected to options which are presented. The options consider the following:

• The degree of automation

• The boundary between the system and the users

• The distribution of the system, for example, is it centralized to one office or spread out across several?

• Cost/benefit

• Impact of the new system

Where necessary, the option will be documented with a logical data structure and a level 1 data-flow diagram.

The team chose a single business option. This may be one of the ones already defined or may be a synthesis of different aspects of the existing options. The output of this stage is the single selected business option together with all the outputs of the feasibility stage.

### **Generate project team**

When doing a feasibility study related to assigning a successful development team for a project we found that project managers are facing several issues. In general they need to analyze past data based on their performances and depend on that suitable people should be selected to do the work. In that traditional way sometimes bulk set of documents have to be analyzed to track their past work of experience. We identified that in our feasibility study there is no feasible way of doing it currently. Because of that proposed method of predicting team needed to be taken in to action.

Another important thing that we have discovered in this feasibility study is the benefits that users can reap by implementing this proposed method. Let’s consider a simple example like this. After examining employee’s details for a considerable time, project managers assign a team for a project. But they have no guarantee that they are the right people for the task. Because they just reviewed their past work and assigned to a fresh project. They didn’t have proof to show these are the right people because they choose them by themselves and no any accurate evidence of proving them. And let’s assume in the middle of the project, project manager realized that these are not the relevant people for complete the tasks. What has to be done next then? Anyhow project managers have to repeat the same process by examining the documents of employees to find another set of people to carry out the project. Then it can obviously see that lot of time, effort, money is wasting and also project success rate is going down. To avoid these failures and mitigate the risk this proposed method is feasible in most of the aspects. We discovered that by implementing this method in a practical scenario no of unnecessary tasks can be eliminated and project management process can be handled absolutely and conclusively.

In order to predict the team for a project there should be historical data based on the employee skills and performances. Otherwise it is so hard to achieve what we really expect. And also the gathered information should be organized precisely to given as an input to the algorithm. Capturing each individual’s technology skills had been a really crucial task since most accurate data should be taken to predict the most accurate team. Since we followed several techniques to gather the information based on developer’s skills.

* One-on-one Interviews

This is a most common and traditional technique to gather information. Here we interviewed each and every developer and ask them certain questions basically on the programming technologies.

* Group Interviews

Group interviews are similar to the one-on-one interviews, except that more than one person is being interviewed usually two to four. These interviews work well when everyone is at the same level or has the same role. We had more preparation and more formality to get the information we wanted from all the participants.

* Questionnaires

Questionnaires are much more informal, and they are good tools to gather accurate details from employees. Since we need a lot of requirements we prepared questionnaires and had dozens of people participated in that.

After generating the final requirements then the system is going to predict the development team for the software project. Then the finalized tasks will be distributed among the selected developers. The development team is being selected based on their performances. In order to achieve something crucial task like that a larger data set is needed. Hence a deep investigation is carried out upon similar non-critical projects from software industries to get the parameters for selection of project personnel. By going through successful surveys we were able to gather the data that we need for the system to do predictions.

Data mining, also popularly known as Knowledge Discovery in Database, refers to extracting or mining knowledge from large amounts of data. The sequences of steps identified in extracting knowledge from data are shown in Figure 10.

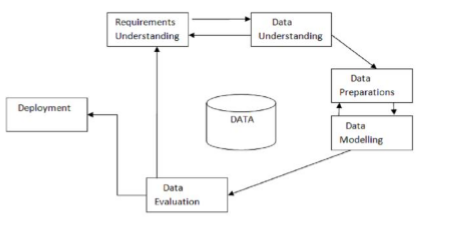


Figure 4 : Extracting knowledge from data

Various algorithms and techniques like Classification, Clustering, Regression, Artificial Intelligence, Neural Networks, Association Rules, Decision Trees, Genetic Algorithm, Nearest Neighbor method etc., are used for data mining process [14]. Our Techniques and methods in data mining need brief mention to have better understanding. Here what we are going to use classification algorithms to master the prediction process.

We have done several surveys in IT companies to gather the data that we need to develop the development team for a project. It is a known fact that collecting information such as people skills is always a challenging, vigorous task and if it is done properly then that is a tremendous achievement since we can originate more accurate predictions.

What is classification..?

Classification is a data mining function that assigns items in a collection to target categories or classes. The goal of classification is to accurately predict the target class for each case in the data. For example, a classification model could be used to identify loan applicants as low, medium, or high credit risks.

A classification task begins with a data set in which the class assignments are known. For example, a classification model that predicts credit risk could be developed based on observed data for many loan applicants over a period of time. In addition to the historical credit rating, the data might track employment history, home ownership or rental, years of residence, number and type of investments, and so on. Credit rating would be the target, the other attributes would be the predictors, and the data for each customer would constitute a case.

Classifications are discrete and do not imply order. Continuous, floating-point values would indicate a numerical, rather than a categorical, target. A predictive model with a numerical target uses a regression algorithm, not a classification algorithm.

The simplest type of classification problem is binary classification. In binary classification, the target attribute has only two possible values: for example, high credit rating or low credit rating. Multiclass targets have more than two values: for example, low, medium, high, or unknown credit rating.

In the model build (training) process, a classification algorithm finds relationships between the values of the predictors and the values of the target. Different classification algorithms use different techniques for finding relationships. These relationships are summarized in a model, which can then be applied to a different data set in which the class assignments are unknown.

Classification models are tested by comparing the predicted values to known target values in a set of test data. The historical data for a classification project is typically divided into two data sets: one for building the model; the other for testing the model.

Scoring a classification model results in class assignments and probabilities for each case. For example, a model that classifies customers as low, medium, or high value would also predict the probability of each classification for each customer.

Classification has many applications in customer segmentation, business modeling, marketing, credit analysis, and biomedical and drug response modeling.

A Sample Classification Problem

Suppose you want to predict which of your customers are likely to increase spending if given an affinity card. You could build a model using demographic data about customers who have used an affinity card in the past. Since we want to predict either a positive or a negative response (will or will not increase spending), we will build a binary classification model.

This example uses classification model, dt\_sh\_clas\_sample, which is created by one of the Oracle Data Mining sample programs. Figure 5–1 shows six columns and ten rows from the case table used to build the model. A target value of 1 has been assigned to customers who increased spending with an affinity card; a value of 0 has been assigned to customers who did not increase spending.

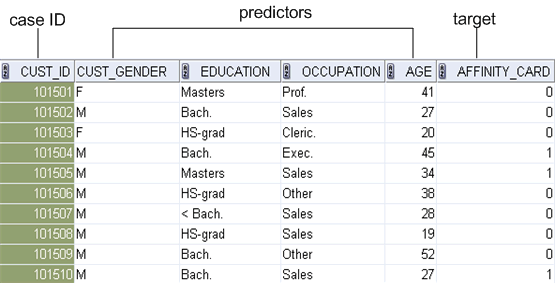


Figure 5 : Sample Data model

There are certain classification algorithms used for data mining purposes such as;

* Decision tree
* Naive Bayes
* Generalized linear models (GLM)
* Support vector machine (SVM)

Here what we are going to use support vector machine algorithm to predict the development team. It is more accurate and realistic approach to reach our target.

Support vector machine

Support Vector Machine (SVM) is a powerful, state-of-the-art algorithm based on linear and nonlinear regression. SVM has strong regularization properties. Regularization refers to the generalization of the model to new data. SVM models have similar functional form to neural networks and radial basis functions, both popular data mining techniques. However, neither of these algorithms has the well-founded theoretical approach to regularization that forms the basis of SVM. The quality of generalization and ease of training of SVM is far beyond the capacities of these more traditional methods.

SVM can model complex, real-world problems such as text and image classification, hand-writing recognition, and bioinformatics and bio sequence analysis.

SVM performs well on data sets that have many attributes, even if there are very few cases on which to train the model. There is no upper limit on the number of attributes; the only constraints are those imposed by hardware. Traditional neural nets do not perform well under these circumstances.

## **Testing and implementation**

After a successful feasibility study it was clear that proposed method should be implemented. Then UML diagrams were designed to the function for analysis, design, and implementation of the system as well as for modeling business and similar processes. Object oriented concepts have been used in the function implementation process. Use case diagrams and different use case scenarios were drawn in order to identify which features to be implemented and how to resolve the errors.

Software testing is an investigation conducted to provide stakeholders with information about the quality of the product or service under test. Software testing can also provide an objective, independent view of the software to allow the business to appreciate and understand the risks of software implementation. Test techniques include, but are not limited to the process of executing a program or application with the intent of finding software bugs (errors or other defects).

It involves the execution of a software component or system to evaluate one or more properties of interest.

### **Tools and Libraries**

* Eclipse

Eclipse is a Java-based open source platform that allows a software developer to create a customized development environment (IDE) from plug-in components built by Eclipse members. Eclipse is managed and directed by the Eclipse.org Consortium. Eclipse is famous for Java Integrated Development Environment (IDE), but C/C++ IDE and PHP IDE are pretty cool too. You can easily combine language support and other features into any of our default packages, and the Eclipse Marketplace allows for virtually unlimited customization and extension.

* JetBrains PyCharm Community Edition 2016.2.2

PyCharm is an Integrated Development Environment (IDE) used in computer programming, specifically for the Python language. It is developed by the Czech company JetBrains.[\*\*\*\*] It provides code analysis, a graphical debugger, an integrated unit tester, integration with version control systems (VCSes), and supports web development with Django. PyCharm is cross-platform, with Windows, Mac OS X and Linux versions. The Community Edition is released under the Apache License,[\*\*] and there is also Professional Edition released under a proprietary license - this has extra features.

### **Unit Testing**

Unit testing, also known as component or module testing, refers to tests that verify the functionality of a specific section of code, usually at the function level. We have carried out testing individual modules precisely and in the latter part of the documentation it is briefly described.

### **Integration Testing**

Integration testing works to expose defects in the interfaces and interaction between integrated components (modules). Progressively larger groups of tested software components corresponding to elements of the architectural design are integrated and tested until the software works as a system.

### **System Testing**

System testing, or end-to-end testing, tests a completely integrated system to verify that it meets its requirements. We intend to perform a system testing to ensure that we have achieved all the objectives of our research up to the level of performance expected. First of all we created a sketch of the system. Then we designed interfaces of the system. After that one by one we tested each and every areas and finally we tested the whole system.

### **Implementation**

Here it is mentioned briefly step by step how the team build function implemented from the very beginning to the end. In the latter part of the documentation including test cases it is mentioned more verbosely.

Organize surveys/interviews to gather employee details

Organize the data set

Reorganize the data and model a data structure

Implement the algorithm

Support vector machine

K-Nearest Neighbors

Naïve Bayes

Train the data set

Predict the team

Figure 6 : Team build function implementation structure

### **Building the Classifier or Model**

This step is the learning step or the learning phase.

In this step the classification algorithms build the classifier.

The classifier is built from the training set made up of database tuples and their associated class labels.

Each tuple that constitutes the training set is referred to as a category or class. These tuples can also be referred to as sample, object or data points.

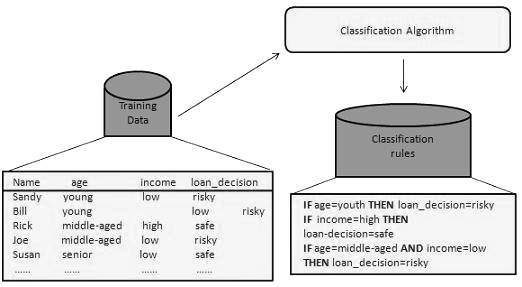


Figure 7 : Building the Classifier or Model

### **Using Classifier for Classification**

In this step, the classifier is used for classification. Here the test data is used to estimate the accuracy of classification rules. The classification rules can be applied to the new data tuples if the accuracy is considered acceptable.

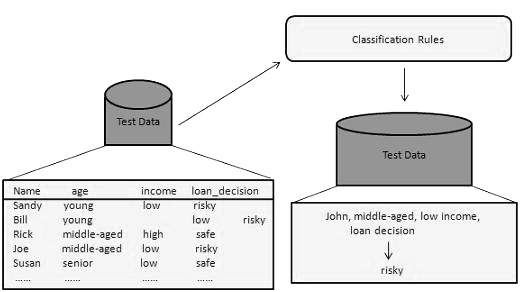


Figure 8 : Using Classifier for Classification

### **Classification and Prediction Issues**

The major issue is preparing the data for Classification and Prediction. Preparing the data involves the following activities −

* Data Cleaning − Data cleaning involves removing the noise and treatment of missing values. The noise is removed by applying smoothing techniques and the problem of missing values is solved by replacing a missing value with most commonly occurring value for that attribute.
* Relevance Analysis − Database may also have the irrelevant attributes. Correlation analysis is used to know whether any two given attributes are related.
* Data Transformation and reduction − The data can be transformed by any of the following methods.
  + - * + Normalization − The data is transformed using normalization. Normalization involves scaling all values for given attribute in order to make them fall within a small specified range. Normalization is used when in the learning step, the neural networks or the methods involving measurements are used.
        + Generalization − The data can also be transformed by generalizing it to the higher concept. For this purpose we can use the concept hierarchies.

Here I am using python to data analysis process. Python has gathered a lot of interest recently as a choice of language for data analysis. Here are some reasons which go in favor of learning Python:

* Open Source – free to install
* Awesome online community
* Very easy to learn
* Can become a common language for data science and production of web based analytics products

After analyzing the past data, team is going to be predicted using support vector machine algorithm and it will display in a well-organized interface. It is further described in the latter part of the document.

Few algorithms are going to use here in order to calculate the developer efficiency. For an example to calculate the developer efficiency;

No of days given for task (n)

Developer efficiency = ----------------------------------- x 100 %

No of days to complete task (c)

If c > n

No of days given for task (n)

Developer efficiency = -------------------------------------- x 100 %

No of days to complete task (c)

## **Research findings**

While doing implementation of the project there are lots of new things development team had to learn. Since this project covers huge area lots of technologies had to mix up and integrated in a compatible way. We have tried different kind of approaches to generate accurate predictions by experimenting different kind of tools and techniques. So it took time to identify the suitable mechanism to increase the accuracy level of prediction.

### **Identify the appropriate classification algorithms**

According to our scenario the best algorithm to master the predictions is the support vector machine. Accuracy level of SVM is higher than the other classification algorithms.

### **Identify suitable language for data analysis**

Python has been selected to use for this purpose since it runs as scripts it can be called inside any programming platform and it is a common language for web based analytics products.

# **RESULTS AND DISCUSSIONS**

## **Results**

This section conveys about the results of the research. The purpose of this topic is to give the reader to compare and contrast the research topic and the implemented system. After reading this section reader can understand the research completely and the implemented system correctly.

Let’s see how this team build function works in a real scenario. To do that we have selected a software company called “eaglejapan”. Basically in this company they are managing software projects which are receiving from clients. So that we have selected this company to test the team build function to find the optimal team for a software project. After going through several steps we were managed to generate the accurate team for the software project. Organization management also satisfied with the outcome result because they also approved that it was the feasible team.

Here I have listed test cases briefly according to the execution order.

**Test cases**

Table 2:Insert the project details

|  |  |
| --- | --- |
| Test case ID | 01 |
| Test case description | Insert the project details |
| Inputs | Project name, Client name, Company name, Project duration, Project Category |
| Expected output | Display a successful message after submitting data |
| Actual output | Data stored in the database and displayed the success message |
| Test status | Pass |

Table 3:Select the job roles of the employees

|  |  |
| --- | --- |
| Test case ID | 02 |
| Test case description | Select the job roles of the employees |
| Inputs | Submit no of employees need from selected job roles to the project development |
| Expected output | Display a success message after submitting the data |
| Actual output | Data stored in the database and displayed the success message |
| Test status | Pass |

Table 4:Provide technologies that are needed to the software project

|  |  |
| --- | --- |
| Test case ID | 03 |
| Test case description | Provide technologies that are needed to the software project |
| Inputs | Estimate cost, programming languages, databases, |
| Expected output | Display a success message after submitting the data |
| Actual output | Data stored in the database and displayed the success message |
| Test status | Pass |

Table 5:Upload documents which includes the details of the software project

|  |  |
| --- | --- |
| Test case ID | 04 |
| Test case description | Upload documents which includes the details of the  project |
| Inputs | SRS file, Requirement file, URS file |
| Expected output | Get the file path of the uploaded document |
| Actual output | Message was displayed after uploading the document and was able to get the file path |
| Test status | Pass |

Table 6:Upload employee details including each employee’s skills and basic details

|  |  |
| --- | --- |
| Test case ID | 05 |
| Test case description | Upload employee details including each employee’s skills and basic details |
| Inputs | Employee data files |
| Expected output | Display a success message after uploading all the data files |
| Actual output | Data files were uploaded successfully |
| Test status | Pass |

Table 7:Retrieve the absolute file paths from the uploaded documents

|  |  |
| --- | --- |
| Test case ID | 06 |
| Test case description | Retrieve the absolute file paths from the uploaded documents |
| Inputs | Employee data files |
| Expected output | Get the file path of the documents |
| Actual output | Successfully retrieved the file paths of the uploaded documents |
| Test status | Pass |

Table 8:Extract data from the uploaded documents and store in the database

|  |  |
| --- | --- |
| Test case ID | 07 |
| Test case description | Extract data from the uploaded documents and store in the database |
| Inputs | Employee data files |
| Expected output | Display a success message after reading and storing all the data files in the database |
| Actual output | All the data has been extracted from the data files and stored in the database |
| Test status | Pass |

Table 9:Select the project that need to predict the team

|  |  |
| --- | --- |
| Test case ID | 08 |
| Test case description | Select the project that need to predict the team |
| Inputs | Submitted new projects |
| Expected output | Load the new projects in the combo box |
| Actual output | Newly added projects loaded to the combo box |
| Test status | Pass |

Table 10:Display a success message

|  |  |
| --- | --- |
| Test case ID | 09 |
| Test case description | Display a success message |
| Inputs | Submitted new projects |
| Expected output | When user select a project from the combo box and then continue success message should be displayed with the selected project |
| Actual output | Selected project was displayed as the successful message |
| Test status | Pass |

Table 11:Display an error message

|  |  |
| --- | --- |
| Test case ID | 10 |
| Test case description | Display an error message |
| Inputs | Submitted new projects |
| Expected output | When user didn’t select any project from the loaded projects error message should be displayed |
| Actual output | Error message was displayed when user didn’t select any project from the combo box |
| Test status | Pass |

Table 12:Team selection attributes

|  |  |
| --- | --- |
| Test case ID | 11 |
| Test case description | Team selection attributes |
| Inputs | Employee’s characteristics such as Technology distribution, Soft skills, Previous project success rates |
| Expected output | Pass the parameters to the classification algorithm |
| Actual output | Attributes have been successfully passed to the algorithm |
| Test status | Pass |

Table 13:Algorithm selection

|  |  |
| --- | --- |
| Test case ID | 12 |
| Test case description | Algorithm selection |
| Inputs | Classification algorithms which are Support vector machine, K-Nearest neighbors, Naïve bayes |
| Expected output | Selected algorithm should be executed to predict the development team |
| Actual output | Algorithm executed correctly which is selected by the user |
| Test status | Pass |

Table 14:Train the data set for Support vector machine algorithm

|  |  |
| --- | --- |
| Test case ID | 13 |
| Test case description | Train the data set for Support vector machine algorithm |
| Inputs | Test data set, Train data set |
| Expected output | Generate most suitable team for the project based on the probability |
| Actual output | Most accurate team predicted for the software project |
| Test status | Pass |

Table 15:Train the data set for K-Nearest neighbors algorithm

|  |  |
| --- | --- |
| Test case ID | 14 |
| Test case description | Train the data set for K-Nearest neighbors algorithm |
| Inputs | Test data set, Train data set |
| Expected output | Generate most suitable team for the project based on the probability |
| Actual output | Most accurate team predicted for the software project |
| Test status | Pass |

Table 16:Train the data set for Naïve Bayes algorithm

|  |  |
| --- | --- |
| Test case ID | 15 |
| Test case description | Train the data set for Naïve Bayes algorithm |
| Inputs | Test data set, Train data set |
| Expected output | Generate most suitable team for the project based on the probability |
| Actual output | Most accurate team predicted for the software project |
| Test status | Pass |

Test case 13, 14, 15 are repeated several times to train the data set again and again in order to increase the accuracy of team prediction. When training a data set more and more data set is trained precisely and that leads to predict the optimal development team for a particular project.

Table 17:Predicted team which is generated by the algorithm stored in the database

|  |  |
| --- | --- |
| Test case ID | 16 |
| Test case description | Predicted team which is generated by the algorithm stored in the database |
| Inputs | Test data set, Train data set |
| Expected output | Store the predicted tea in the database |
| Actual output | Successfully predicted team recorded in the database |
| Test status | Pass |

Table 18:Load the selected developers for the team in a graphical view

|  |  |
| --- | --- |
| Test case ID | 17 |
| Test case description | Load the selected developers for the team in a graphical view |
| Inputs | Predicted employee details |
| Expected output | View the generated team in the interface |
| Actual output | Predicted team successfully loaded in the interface |
| Test status | Pass |

### **Test Approach**

While implementing the system we manually did the same things happening in the system by taking data from organizations, conducting interviews, conducting surveys etc. Then the manually finalized data are checked with the data finalized by our system.

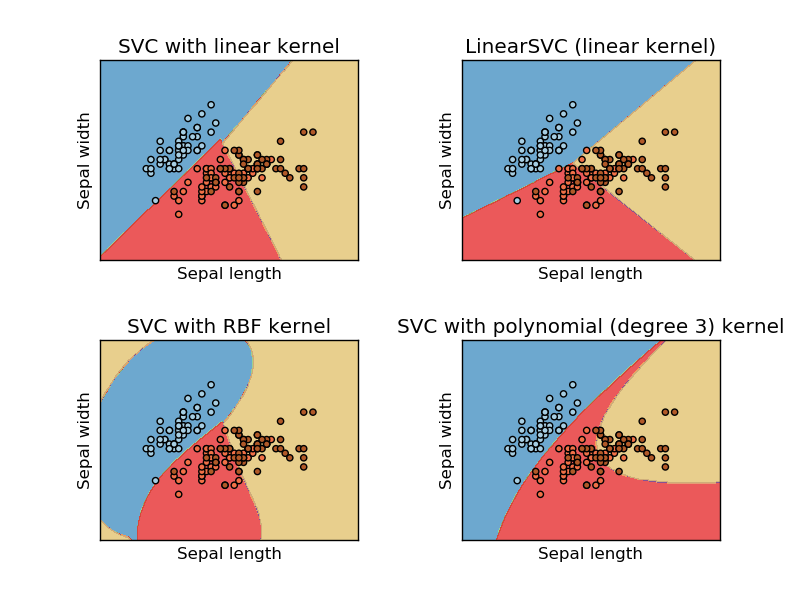


Figure 9 : Support vector machine

Stage 1: Descriptive Analysis / Data Exploration

In order to perform this task it takes more time. When the data set is getting higher it will need more time to complete this step. There are several steps of data exploration and preparation.

* Variable Identification
* Univariate Analysis
* Bi-variate Analysis
* Missing values treatment
* Outlier treatment
* Variable transformation
* Variable creation

Stage 2: Data treatment (missing values treatment)

* Create dummy flags for missing values
* Impute missing value with mean/ median/ any other easiest method
* Impute missing value of categorical variable

Stage 3: Data modeling

Here we are using SVM technique since it is more effective on prediction

Stage 4: Estimation of performance

There are several methods to validate model performance. So here we need to divide the data set into training set and testing set.

Training set and testing set

While experimenting with any learning algorithm, it is important not to test the prediction of an estimator on the data used to fit the estimator as this would not be evaluating the performance of the estimator on new data. This is why datasets are often split into train and test data [18].

## **Discussion**

To judge system performance it would be important to compare results with those obtained with other systems in similar conditions. In our case it is not possible to refer to a previous work on the automated software project management systems since there are no any similar tools implemented before.

According to the surveys we have done there is need of a having proper automated software project management tool. In there also generating predictions is much more important since basically whole project is depending on the predictions which is generated by the system and employees can get a clear idea about the project before start of the project. It is a needed fact to produce sustainable projects bringing to existence. As I mentioned earlier assigning a development team for a particular software project is always a crucial task which has to be done by a project manager. If the project is urgent and large in size, then development team should be assigned as soon as possible and it must be the best team.

When we tested this function above mentioned company they were really satisfied with our product and it seemed, there was a need of a function like this to their company. We realized that when we are discussing with them they can harness no of benefits using team build function. We also believe that implementing a function like this most of the software companies can save and manage their time, cost, risk, effort efficiently and ultimately generate the precise team for the project dynamically through the system itself. There are several improvements have to be made to increase the accuracy level of the data set. It was examined that when we were running the function for several times. Finally according to the tested company’s and the group member’s view this team build function was identified as a much needier characteristic in project management.

And also we had to work more hard in order to gather the employee skills including their working experiences and knowledge based on programming technologies etc. That had to be more realistic and efficient since depend on that development team is predicted.

# **CONCLUSION**

The objective of this study is to support the role of the Project Manager in a software company. In order to meet this objective, we had to move with several steps such as feasibility study, planning, designing, implementation and testing. We used limited number of resources in order to make our product more unique from other products. This report contents the way to create a tool related to project management step by step. This report can be used by any person who is willing to continue in this research area. This project can be expanded to cater more functionalities.

This project caters the functionalities of how to use the programming languages to handle future predictions, how to generate tasks from requirements, how to reduplicate requirements using the ontology engine, how to analyze and summarize information. All these functionalities can be further expanded by adding more research components. A very high business value is addressed on this system, since this research can be further expanded and establish at company systems. Then can have the results and can see how this system will reduce the total cost of the system.

In this paper a strategy is proposed to develop an IT enabled system for Software Project Management which has got a huge impact on the lifecycle of Project Management in Sri Lanka. This leads to the generation of a product which provides a number of functions. Due to these various functionalities of the product the visually impaired persons of Sri Lanka are going to be facilitated.

The system is developed after quite a lot of background study and reference work done by the group members. Each and every member of the group gained additional knowledge during the whole development period regarding practical work, presentations and documentations. Each team member took attentive lessons on the program languages they needed to develop the system. And the theoretical knowledge was put into practice.

# **References**

[1] M. Shepperd, "Cost prediction and software project management," in Software Project Management in a Changing World. Springer Science + Business Media, 2014, pp. 51–71.

[2] G. Ruhe and C. Wohlin, Eds., Software project management in a changing world. Springer Science + Business Media, 2014.

[3] Daya Gupta, Rajni Jindal, Vaibhav Verma, Dilpreet Singh Kohli, Shashi Kant Sharma, Predicting student‘s behaviour in education using J48 algorithm Analysis tools in WEKA environment‖, BIDW, 2010.

[4] E. 2008, "9 steps to a hassle free and effective software development project," Project Smart, 2008. [Online]. Available: https://www.projectsmart.co.uk/9-steps-to-a-hassle-free-and-effective-software-development-project.php. Accessed: Sep. 1, 2016.

[5] "Software development life cycle," www.tutorialspoint.com, 2016. [Online]. Available: http://www.tutorialspoint.com/software\_engineering/software\_development\_life\_cycle.htm. Accessed: Sep. 1, 2016.

[6] T. Clark, "4 common IT project management challenges and 4 solutions," in Project Management, LiquidPlanner, 2015. [Online]. Available: https://www.liquidplanner.com/blog/4-common-it-project-management-challenges-and-4-solutions/. Accessed: Sep. 1, 2016.

[7] Ondiappan Arivaz , "Making better estimates of project durations using Monte Carlo Analysis," 2015. [Online]. Available: http://blog.minitab.com/blog/statistics-in-the-field/making-better-estimates-of-project-duration-using-monte-carlo-analysis. Accessed: Sep. 1, 2016.

[8] B. Jackson, "5 ways to more accurately forecast with project management software," IT Business. [Online]. Available: http://www.itbusiness.ca/blog/5-ways-to-more-accurately-forecast-with-project-management-software/50306. Accessed: Sep. 1, 2016.

[9] . [Online]. Available: https://www.oracle.com/solutions/business-analytics/businessintelligence/index.html. Accessed: Sep. 1, 2016.

[10] Leandro Soriano Marcolino, Haifeng Xu, Albert Xin Jiang, Milind Tambe, Emma Bowring, “Give a Hard Problem to a Diverse Team: Exploring Large Action Spaces,”

[11] C. F. Chien and L. F. Chen, "Data mining to improve personnel selection and enhance human capital: A case study in high-technology industry," Expert Systems with Applications, vol. 34, pp. 280-290, 2008

[12] M. Hertzum and A. M. Pejtersen, "The information-seeking practices of engineers: searching for documents as well as for people,". [Online]. Available: https://scholar.google.com/scholar?q=searching+inside+documents&btnG=&hl=en&as\_sdt=0%2C5. Accessed: Sep. 1, 2016.

[13] M. Kuba, "OWL 2 and SWRL Tutorial," 2012. [Online]. Available: http://dior.ics.muni.cz/~makub/owl/. Accessed: Sep. 1, 2016.

[14] J. Han ,M. Kamber, M. Kaufmann, “Data Mining: Concepts and Techniques”, 2000.

[15] V. Suma, and T.R. Gopalakrishnan Nair, ―Effective defect prevention approach in software process for achieving better quality levels‖, Fifth International Conference on Software Engineering, vol. 32, pp.2070–3740, August 2008.

[16] "ETL,” [Online]. Available: http://datawarehouse4u.info/ETL-process.html. Accessed: Sep. 1, 2016.

[17] Pankaj Jalote , ‖Software Project Management in Practice‖,Addison Wesley 2002

[18] S.Gupta1, Suma, “Empirical Study on Selection Of Team Members For Software Projects – Data Mining Approach”

[\*\*\*\*] "Jetbrains Strikes Python Developers With Pycharm 1.0 IDE". Eweek.com. N.p., 2016. Web. 28 Oct. 2016.

[\*\*]"Pycharm 3.0 Community Edition Source Code Now Available | Pycharm Blog". Blog.jetbrains.com. N.p., 2016. Web. 28 Oct. 2016.

# **Glossary**

Python - Programming language

Java - Programming Language

Ontology - Advanced Database for user to extract knowledge

HTML - Standard Markup Language for creating web pages and web applications

# **Appendices**

Below use case diagram shows how the predictions are generated inside the system by going through several steps.

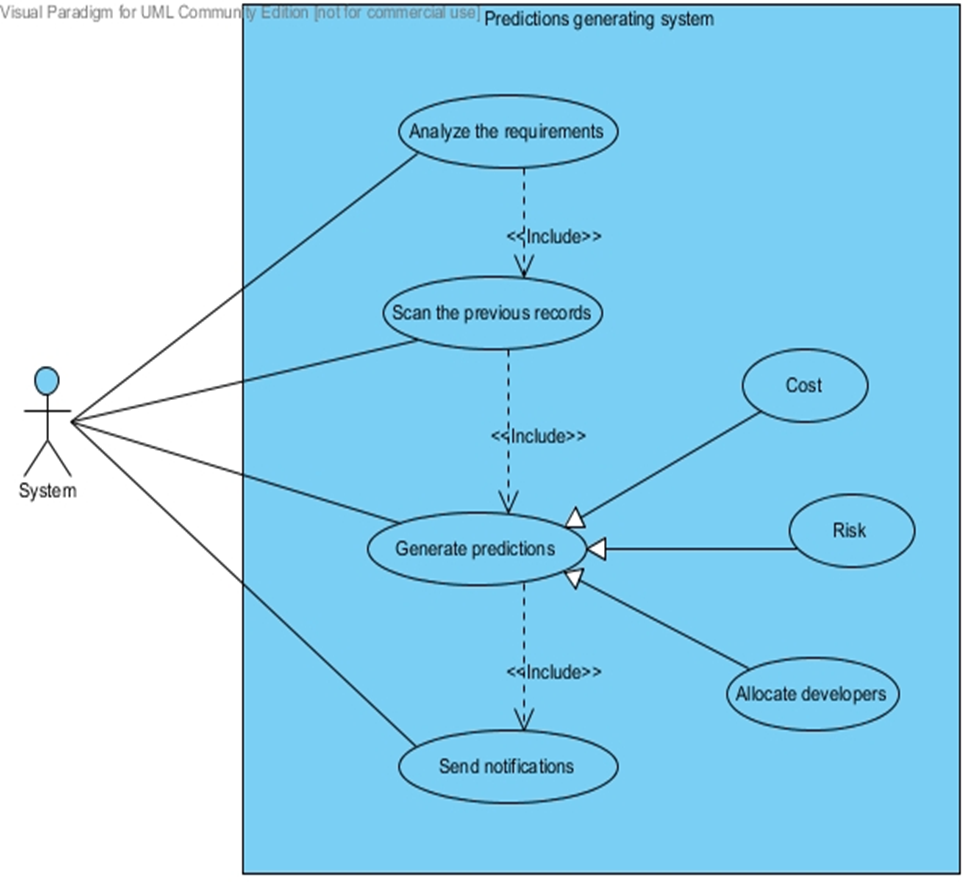


Figure 10 : Use case scenario

When finding right personnel for a software project following steps have to be completed by the user.

1.Insert project details.

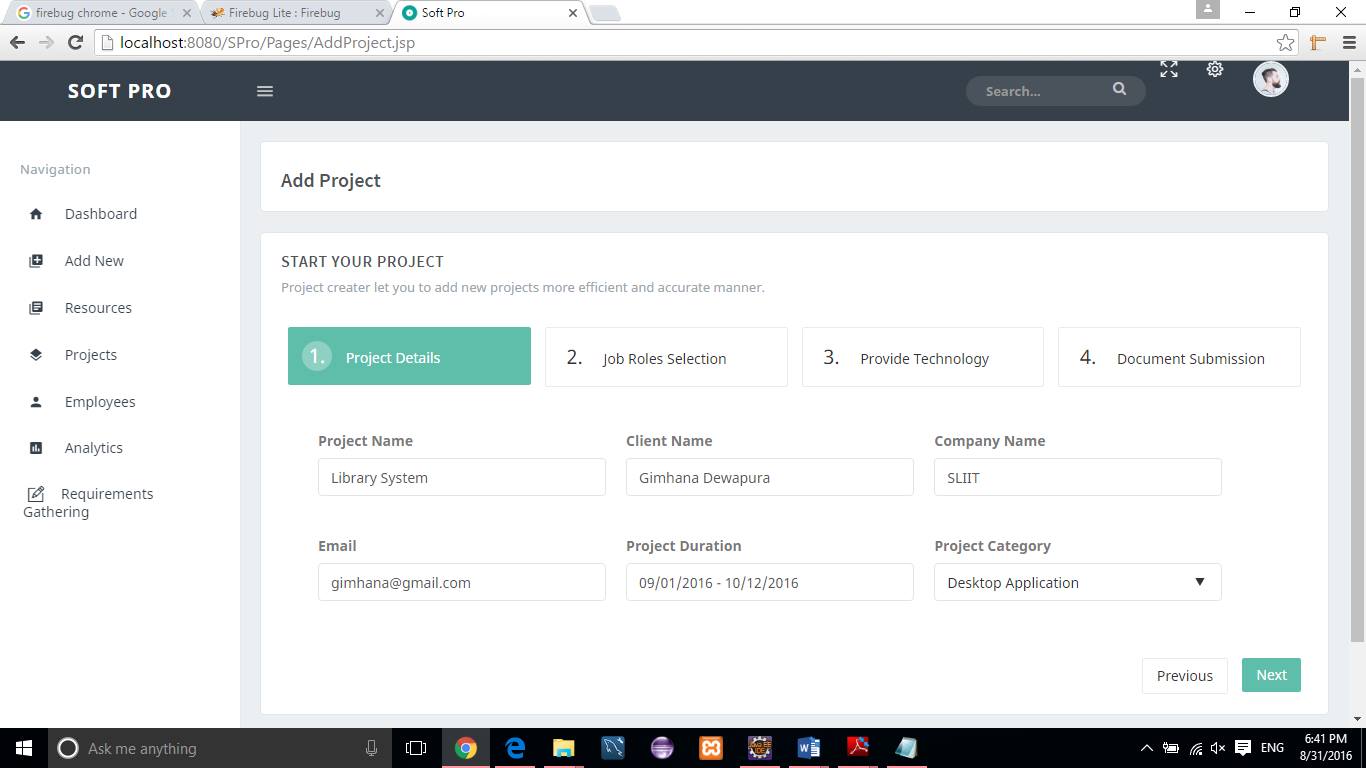


Figure 11 : Insert project details

2.Select needed job roles for the project.

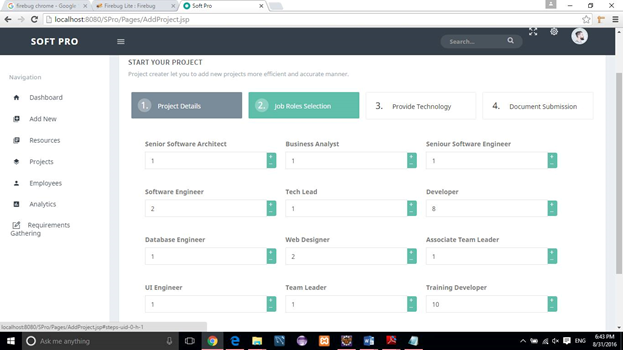


Figure 12:Job roles selection

3.Provide technologies including databases, programming languages etc.

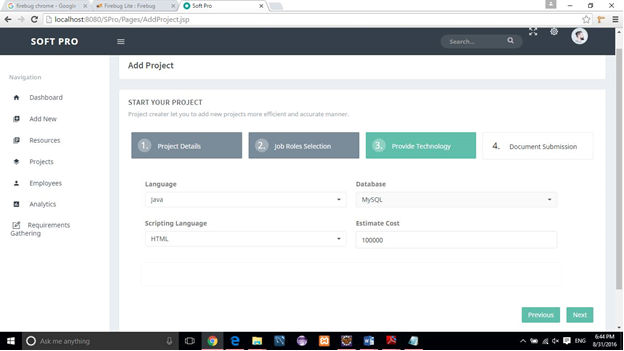


Figure 13:Provide technology

4.Upload employee data files.

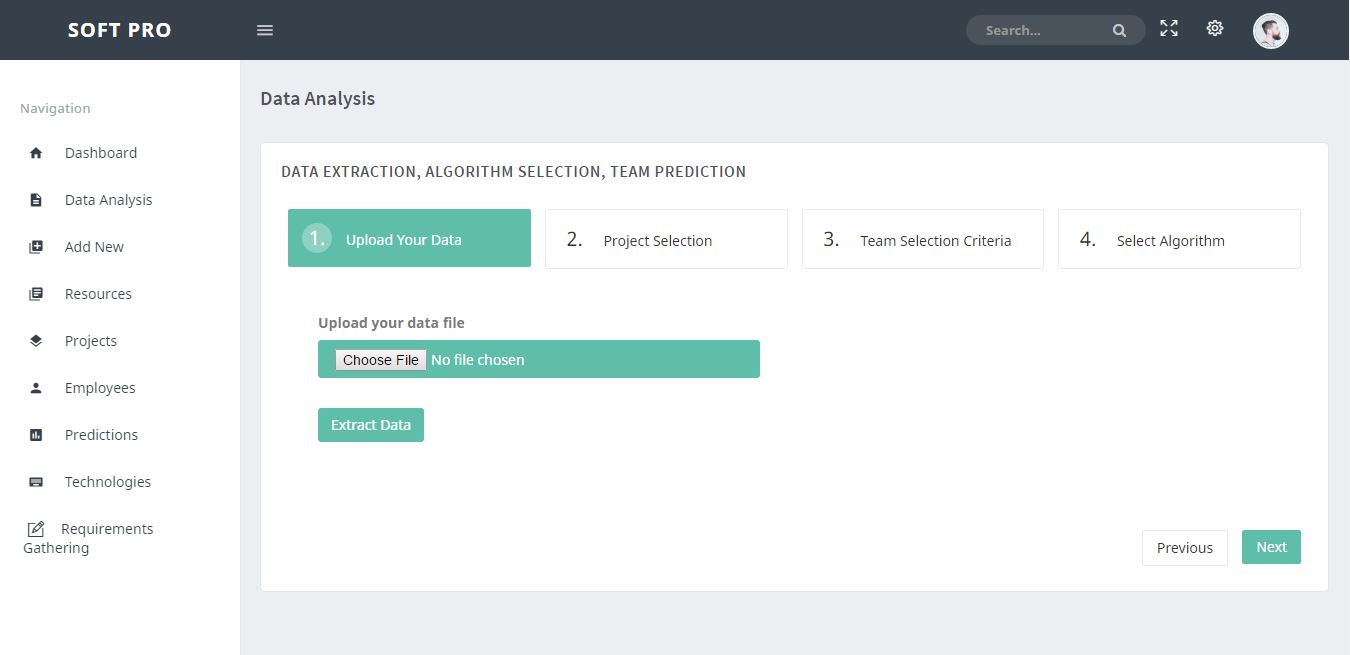


Figure 14:Upload employee data

5.Select the project that need to generate the team.

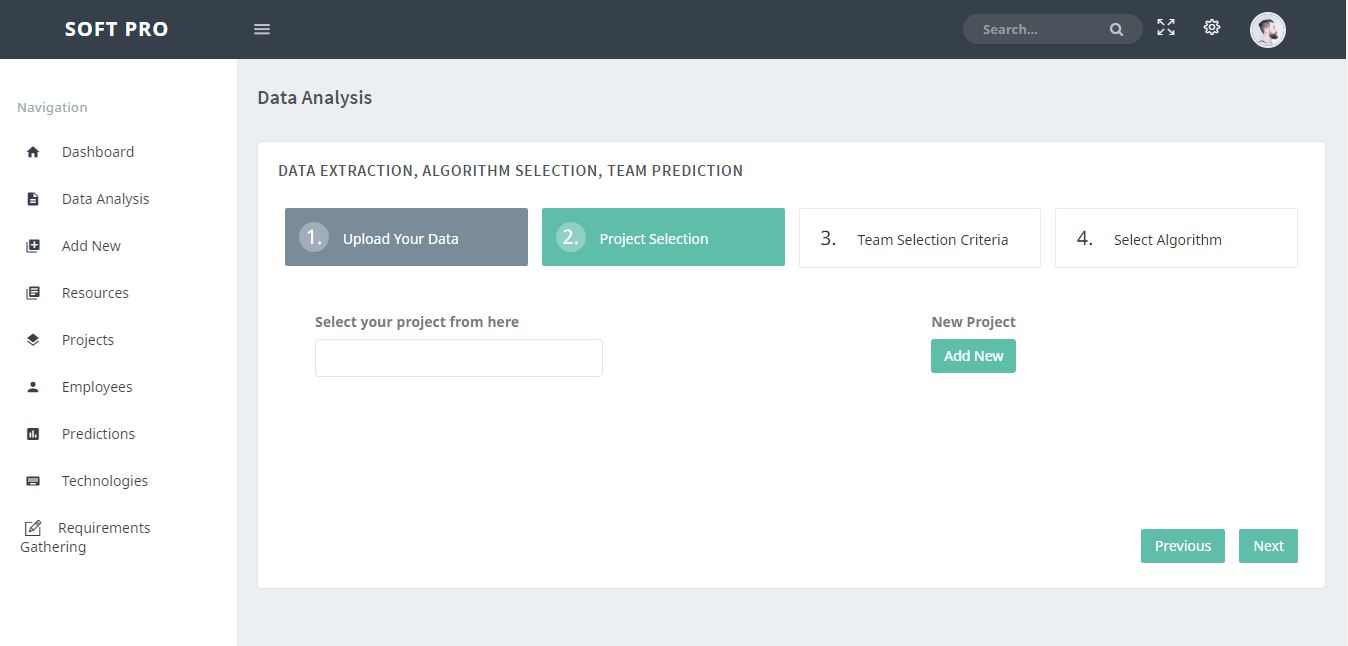


Figure 15:Project selection

6.Select attributes which need to pass to the algorithm as parameters.

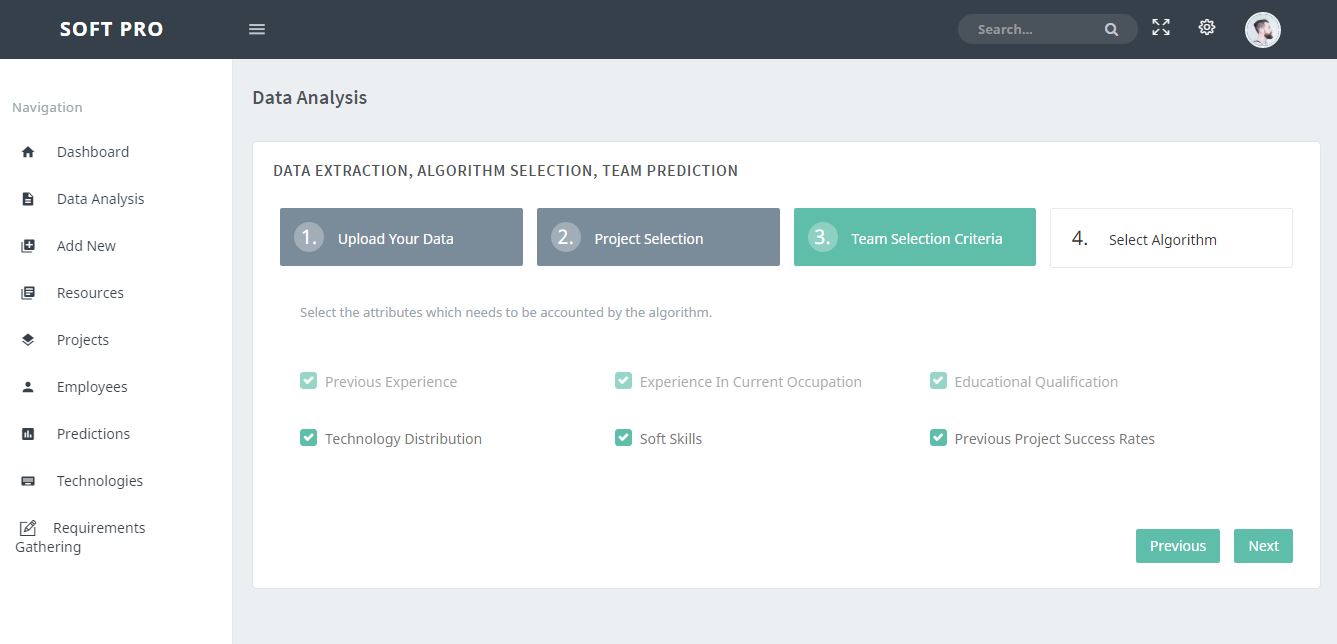


Figure 16:Team selection criteria

7.Choose the classification algorithm to execute.

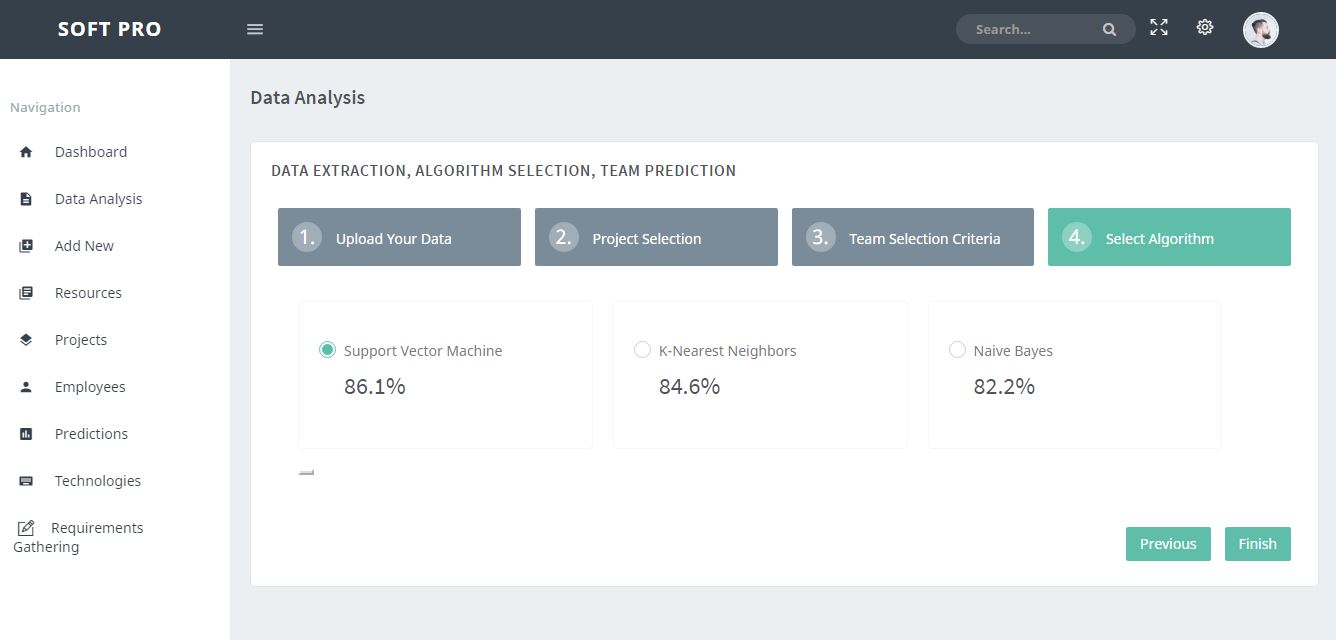


Figure 17:Algorithm selection

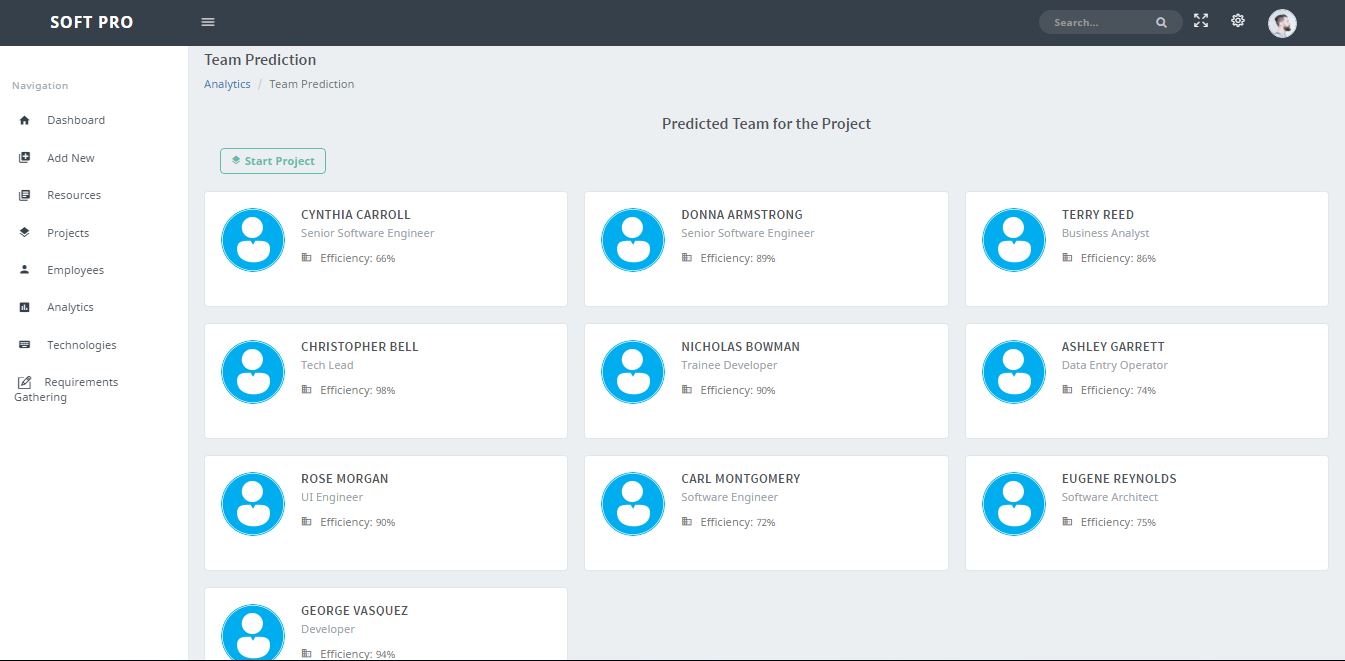
8.Predicted team for the software project.

Figure 18:Predicted team for the software project