



People Clues

BI tool to dynamically select the group for a given project

Project ID: 16-073

Project Proposal



B.Sc. Special (Honors) Degree in Information Technology

Submitted on 2016-03-09

People Clues

BI tool to dynamically select the group for a given project.

Project ID: 16-073

Authors:

Student ID	Name	Signature
IT13082598	Perera B.G.A	
IT13009700	Perera L.N.P	
IT13068974	Samarasingha H.K.L	
IT13024314	Senaratna P.M.	

Supervisor

.....

Dr. Rohan Samarasinghe

DECLARATION

We declare that this is our own work and this project proposal 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 our 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.

.....

Perera B.G.A

.....

Perera L.N.P

.....

Samarasingha H.K.L

.....

Senaratna P.M

ABSTRACT

Today in Sri Lanka, in both apparel industry and IT industry since there are huge number of employees, workforce management is a critical factor. Managing individuals and groups in an effective way has a big impact on decision making process and productivity of the company. Unfortunately looking for optimal or near optimal teams and feasible or near feasible teams is costly task for humans due to the exponential number of outcomes. So far it has been a manual process where these companies spent a huge cost on consultation for expertise knowledge or great time and resource wastage to allocate expertise within the company. The proposal aims at describing processes of building cost effective, flexible, ease of use and productive enterprise focused Business Intelligence tool – **People Clues**. **People Clues** is a business intelligence tool which will select the optimal or feasible team for a given project depending on the attributes selected by the user. In this paper we present a desktop and a web application that facilitates the task of automating dynamic team generation depending on the optimality or feasibility based on different knowledge areas such as team dynamics, predictive modeling, business intelligence, data mining and team characteristics. An ETL (Extract, Transform and Load) tool will analyze the data and four prediction models will be designed in order to generate the optimal and feasible solutions for both apparel and IT industry. Depending on the relationship of the attributes (Internal and External) People Clues will dynamically visualize the most optimal or feasible team on the Dashboard.

TABLE OF CONTENT

1	INTRODUCTION.....	7
1.1	Background	7
1.2	Literature Review.....	8
1.2.1	Products Available in the Market.....	14
1.3	Research Gap & Research Problem	17
1.3.1	Research Gap	17
1.3.2	Research Problem	19
2	OBJECTIVES	20
2.1	Main Objectives	21
2.2	Specific Objectives	21
3	RESEARCH METHODOLOGY	23
3.1	System Architecture.....	24
3.2	Tools and Techniques.....	29
3.3	Testing.....	29
3.4	Gantt chart.....	30
3.5	Work Breakdown Structure	31
4	DESCRIPTION OF PERSONAL AND FACILITIES	32
5	BUDGET (If Any)	34
6	REFERENCES	35
7	APPENDICES (If Any).....	38

Table of Figures

Figure 1 -Level Selection	Figure 2 - Genetic composition	9
Figure 3 - POP TM		14
Figure 4 - CATME Team Maker		15
Figure 5 - Oracle Business Intelligence		15
Figure 6 - Birst		16
Figure 7 - Microsoft Share Point		16
Figure 8 - IBM Cognos		17
Figure 9 - People Clues system Architecture		24
Figure 10 - Work Breakdown Structure		31

1 INTRODUCTION

1.1 Background

People Clues - Business intelligence tool syndicates operational data with analytical tools to present complex and competitive information to decision makers, in order to dynamically select the best team for a given project.

Data in businesses can be very useful only if data is analyzed properly which will aid in strategic business decision making. Therefore, we must find solutions to automatically analyze data, classify and summarize it along with discovering and characterizing trends and flagging anomalies in order to ease the decision making process effectively in a company. **People Clues** targets at developing a set of tools, technologies and programmed products that are used to collect, integrate, aggregate and make data available for better, faster decision making.

Today in Sri Lanka, in both apparel and IT industries, workforce management is a critical factor since there is a huge workforce working under these industries. Managing individuals and groups effectively has a big impact on decision making processes and productivity of a company. This research project involves cutting-edge technical skills and domain knowledge in order to achieve what it is actually meant to be achieved as goals and it will be based on Team dynamics. “Team dynamics is the term used to define how people work and interact together in teams. Team dynamics are the hidden strengths and weaknesses that operate in a team between different people or groups and they affect how a team reacts, behaves or performs [1].” The effects of team dynamics are often very complex. There are various forces that could influence team dynamics, these include nature of the task, the organizational context, and team composition. Not only limited to ETL (Extract, Transform, Load) tool, this project involves the predictive model building, analysis and implementation in regards to the data set available.

Finally, after processing all data set based on predefined predictive models, it generates comprehensive, self-descriptive and exploratory Dashboards which provide fact – driven visualization of the data to decision makers of a company. A heterogeneous or homogeneous team will be given as a solution to the provided task in a representative manner. “The terms “homogeneous team” and “heterogeneous team” used in the current literature cover many different aspects. It is important to note that while all agents in genetically homogeneous teams share the same genes, agents can nevertheless be behaviorally heterogeneous. Heterogeneous

teams are those in which agents are, on average, not genetically more similar to team members than to agents in the rest of the population [2].” Apart from that many software engineering principles and information technology knowledge areas are applied to this project. Especially desktop based application development skills, web development skills, data mining tools [3] and techniques are essential. In addition, some basic knowledge on business process and decision making process are also required throughout the development of the system.

1.2 Literature Review

Prior to the proposal of our research project we conducted a Literature survey on the existing platforms with similar functionalities and technologies. Some of the prominent researches are reviewed here.

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

The objective of this research is to develop a new model to analyze team diversity. It provides facts how to increase team performance with the help of team diversity. The performance of diverse teams increase as the size of the action space gets larger, and also that diverse teams converge faster than uniform teams. Therefore team diversity is important forming teams. This is basically based on the notion of spreading tail and non-spreading tail agents and they allow making many predictions about teams as the action Space or number of agent’s change [10].

2. Waibel, Markus, Keller, Laurent Floreano, Dario, 2009, {IEEE} {T}ransactions on {E}volutionary {C}omputation, **“Genetic Team Composition and Level of Selection in the Evolution of Cooperation”**

This paper provides an experimental demonstration of how the choice of genetic team composition and level of selection influences the performance of Multi Agent Systems in tasks with varying levels of cooperation that do not provide a benefit for specialization.

Genetic Team Composition	Heterogeneous		{11}, {13}, {14}, {15}, {16}, {17}, {18}, {19}, {10}, {11}, {12}, {13}, {14}, {15}, {16}, {17}, {18}, {19}, {20}, {21}, {22}, {23}, {24}, {25}, {26}, {27}, {28}, {29}, {30}, {31}, {32}, {33}, {34}, {35}, {36}, {37}, {38}
	Homogeneous	{17}, {39}	
Level of Selection	Individual	{3}, {17}, {29}, {39}, {40}, {41}, {42}, {43}, {44}, {45}, {46}, {47}, {48}, {49}, {50}, {51}, {52}, {53}, {54}	{17}, {24}, {40}
	Team		

Figure 1 -Level Selection

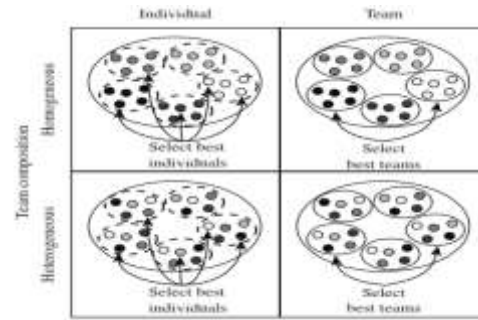


Figure 2 - Genetic composition

- Four possible combinations of genetic team composition and level of selection were formalized into four evolutionary algorithms
- Population size and team sizes are kept constant across generations

Identified and studied three different types of multi agent tasks depending on the amount of cooperation required between team members. Different combinations of genetic team composition as shown in the figure lead to different performances. No combination achieved optimal performance in all three types of task. For tasks that did not require cooperation, heterogeneous individual level selection achieved the highest performance. They performed poor in tasks that needed cooperation while homogeneous teams achieved highest in that. Homogeneous teams are a safe choice in tasks that do not benefit from specialization when the requirements for agent cooperation are difficult to estimate. Heterogeneous teams evolved with team selection were inefficient at selecting for good performance in all three types of tasks studied here and therefore cannot be recommended for cooperative tasks that do not require specialization [13].

3. Agustín-Blas, Luis E., Salcedo-Sanz, Sancho, Ortiz-García, Emilio G., Portilla-Figueras, Antonio, Pérez-Bellido, Ángel M., Jiménez-Fernández, Silvia, **2011**, Computers & Operations Research Journal, **“Team formation based on group technology: A hybrid grouping genetic algorithm approach”**

This paper presents a new model for team formation based on group technology (TFPGT). Specifically, the model is applied as a generalization of the well-known Machine-Part Cell Formation problem, which has become a classical problem in manufacturing in the last few years. In this case, the model presented is especially well-suited for problems of team formation arising in R&D-oriented or teaching institutions. A parallel hybrid grouping genetic algorithm (HGGA) is also proposed in the paper to solve the TFPGT. The performance of the algorithm is shown in several synthetic TFPGT instances, and in a real problem: the formation of teaching groups at the Department of Signal Theory and Communications of the Universidad de Alcalá in Spain [14].

4. E. Semsar-Kazerooni, K. Khorasani, **2009**, American Control Conference, 2009. ACC '09., **“A game theory approach to multi-agent team cooperation”**

The main goal of this work is to design a team of agents that can accomplish consensus over a common value for the agents' output in a cooperative manner. First, a semi-decentralized optimal control strategy introduced recently by the authors is utilized which is based on minimization of individual costs using local information. Cooperative game theory is then used to ensure team cooperation by considering a combination of individual costs as a team cost function. Minimization of this cost function results in a set of Pareto-efficient solutions. The choice of Nash-bargaining solution among the set of Pareto-efficient solutions guarantees the minimum individual cost. The Nash-bargaining solution is obtained by maximizing the product of the difference between the costs achieved through the optimal control strategy and the one obtained through the Pareto-efficient solution. The latter solution results in a lower cost for each agent at the expense of requiring full information set. To avoid this drawback additional constraints are added to the structure of the controller by using the linear matrix inequality (LMI) formulation of the minimization problem. Consequently, although the controller is designed to

minimize a unique team cost function, it only uses the available information set for each agent [15].

5. Chen, Shi Jie Gary, Lin, Li, **2004**, IEEE Transactions on Engineering Management, **“Modeling team member characteristics for the formation of a multifunctional team in concurrent engineering”**

In concurrent engineering, project tasks generally involve the establishment of multifunctional design teams in order to simultaneously consider various activities throughout the entire product life cycle. Team members from different functional departments of the company interact in every phase of development tasks to design the products and processes concurrently. To ensure a successful multifunctional team, it is important to understand the characteristics of team members. Three fundamental descriptors of team members are developed in this research. The first is to represent the multifunctional knowledge of team members due to the need of concurrent engineering. Second, to build a successful project team, teamwork capability of team members is needed by taking their experience, communication skills, and flexibility in job assignment into account. Multifunctional knowledge and teamwork capability ratings are captured from each member using analytic hierarchy process. Third, since the team members work closely, their collegiality directly affects team performance, regardless of their knowledge. Thus, the working relationship model is developed to provide such a metric. Personality profiling using Myers-Briggs type indicator serves as the basis of assessing each team member's abilities to work with others. Finally, we complete this paper by providing a step-by-step procedure with an example for selecting the best multifunctional team using the three ratings. This research helps establish an efficient multifunctional team because every team member will be capable of communicating to each other within the team due to their multifunctional knowledge, teamwork skills, as well as an established good working relationship. The research challenge to accomplish is to “assign the right team members to the right tasks at the right time” [16].

6. Xu, Y, Ma, J, Guo, X, **2012**, Proceedings - Pacific Asia Conference on Information Systems, PACIS 2012, **“Modeling researchers' characteristics for the formation of research team”**

Most of research is completed through a research team where team members' expertise and knowledge complement with each other. In academic contexts, research project generally involve the establishment of multi-disciplinary and competent research teams to simultaneously consider various activities throughout the entire research project. To ensure a successful research team, it is important to understand and model researchers' characteristics in an appropriate and systematic way. Three fundamental descriptors of researchers are developed in this research. The first is to represent researchers' expertise level in areas related to project. The second is to consider the team members' diversity in their expertise. The third is to develop a measurement of team members' working relationships based on their previous co-authorships and personality characteristics. This research helps construct a team for a research project based on these characteristics [17].

7. Becchetti, Luca, Castillo, Carlos, 2012, Journal Article, **“Online Team Formation in Social Networks”**

This research studies the problem of online team formation. They consider a setting in which people possess different skills and compatibility among potential team members is modeled by a social network. A sequence of tasks arrives in an online fashion, and each task requires a specific set of skills. The goal is to form a new team upon arrival of each task, so that (i) each team possesses all skills required by the task, (ii) each team has small communication overhead, and (iii) the workload of performing the tasks is balanced among people in the fairest possible way. They propose efficient algorithms that address all these requirements: our algorithms form teams that always satisfy the required skills, provide approximation guarantees with respect to team communication overhead, and they are online-competitive with respect to load balancing. Experiments performed on collaboration networks among film actors and scientists; confirm that their algorithms are successful at balancing these conflicting requirements.

8. **Three Fundamental Pillars of Multi-agent Team Formation**

This research paper addresses the three fundamental issues when forming Teams of voting agents

- I. Selecting which agents should form a team
- II. Aggregating the opinions of the agents
- III. Assessing the performance of a team

Agent Selection is picking a limited number of agents to form a team. Aggregation of Opinions is combining the opinions of each member of the team into a final team decision.

Team Assessment is verifying the performance of a given team

9. A Detailed Analysis of a Multi-agent Diverse Team

This work is related mainly to the study of team formation, diversity and voting.

Introduction

- show that a diverse team can overcome a uniform team and we give the necessary conditions for it to happen
- present optimal voting rules for a diverse team
- perform synthetic experiments that demonstrate that both diversity and strength contribute to the performance of a team

Results

- Diverse teams can outperform teams composed by copies of the best player.
- A team composed by strong but very similar agents is not necessarily optimal.

Limitations:

- However, it is still a challenge to find the best possible teams
- Moreover, in many complex scenarios cannot even easily enumerate all the possible states of the world.

10. Nahid Saberi, Mohsen Mahvash, and Marco Zenati, Member, IEEE, “An artificial system for selecting the optimal surgical team ”

The main objective of this research is to develop an intelligent system to optimize a team composition based on the team’s historical outcomes and apply this system to compose a surgical team. Forming a group of skilled professionals is practically difficult. Hence they have developed a system to identify the most effective surgical team to conduct surgeries for patients. At First they obtain the probability of unfavorable outcome for all team compositions that have operation records. A probability value is then assigned to each person of the database based on his/her record on all participated team compositions. A model is introduced to calculate the probability for any team composition based on the probability values for the members of the team. According to that an optimal team is selected by finding the team which has the minimum probability. But the drawback of this research is that they have only considered about the past recorded unfavorable outcomes of the surgical teams. The system can generate results on the performance of the staff of the surgical team based only on the recorded data [11].

1.2.1 Products Available in the Market

1. POP™ – Sales People Selection & Recruiting Tool



Figure 3 - POP™

- It only provides information about the individual’s fit to a sales role.
- Applicable only for an individual and not a good choice of selection when creating a group since it doesn’t address team characteristics.

2. A Team Formation Tool for Educational Environments

- Information for the predictions are provided by the students and not taken from the past activities or behavior.
- Each and every user has to log in to the system and interact for the analytics.
- No Predictive Analysis.
- Focused only for the optimal solution doesn't consider about the feasible solution [21].

3. CATME Team Maker



Figure 4 - CATME Team Maker

- Need to gather information from users
- Consider only attributes related to students
- It doesn't have Data Extraction, Transformation, Load Tool instead of that gathers data using surveys and questionnaires.
- Lack of user friendly interfaces
- It only provides the facility to divide students into teams and doesn't consider about their roles

4. Oracle Business Intelligence


	Vendor:	Oracle	Customers:	400,000
	Founded:	1977	Deployment Model:	On-premise, Cloud
	Headquarters:	Redwood Shores, CA	Intended Users:	Small, Medium, Large
	Ownership:	Public	Free Trial:	Yes

Figure 5 - Oracle Business Intelligence

- It does not have issue indicators
- Customizable Features are less according to the attributes
- Does not have a direct feature to generate optimal or feasible employee groups
- Visualization needs to be improved [20].

5. Birst


	Vendor:	Birst	Customers:	1,000+
	Founded:	2004	Deployment Model:	On-premise, Cloud
	Headquarters:	San Francisco, CA	Intended Users:	Small, Medium, Large
	Ownership:	Private	Free Trial:	Yes

Figure 6 - Birst

- Needs some time to get used to it as it is not so user friendly.
- Reports are limited to basic ad-hoc reports.
- Data processing is slow after uploading new data or making changes to existing data model. It takes lot of time to create data warehouse [19].

6. Microsoft Share Point

	Vendor:	Microsoft	Customers:	Unknown
	Founded:	1975	Deployment Model:	On-premise, Cloud
	Headquarters:	Redmond, WA	Intended Users:	All
	Ownership:	Public	Free Trial:	Yes

Figure 7 - Microsoft Share Point

- There is no way to access on premise data
- Lack of online analytical processing
- There is no dynamic or automatic data analyzing process [19].

7. IBM Cognos


	Vendor:	Cognos	Customers:	23,000
	Founded:	1911	Deployment Model:	Cloud
	Headquarters:	Armonk, NY	Intended Users:	All
	Ownership:	Public	Free Trial:	Yes

Figure 8 - IBM Cognos

- It does not support its own ETL Tool and data quality software, as they available as separate platforms.
- Reporting feature is not user friendly.
- There is a problem with visual analytics [19].

1.3 Research Gap & Research Problem

1.3.1 Research Gap

Even though there are existing proposed products in the market area, they do not address most of the problems that the proposed system is going to address. The following table shows a comparison of features between the existing products or applications and the proposed solution “People Clues”.

Features	POP™	A Team Formation Tool	CATME Team Maker	Oracle BI	Birst	MS Share Point	IBM Cognos	People Clues
ETL Tool (Automated)								✓
optimal solution	✓							✓
Feasible Solution								✓
Consider about role of each team member	✓							✓
applicable for many roles		✓	✓	✓	✓	✓	✓	✓
All the user does not have to interact with the system	✓			✓	✓	✓	✓	✓
Visualization of data in an abstract way					✓	✓		✓
Predictive Analytics	✓		✓	✓	✓	✓	✓	✓

1.3.2 Research Problem

The world is evolving and emerging in the hub of a wave of revolutions, spanning from massive changes in global business to the ever-exploding presence of information technology. As businesses get bigger day by day, the complexity is also getting higher. New trends are being incorporated with business ecosystems. Customer behavior is changing, new technologies are evolving rapidly in a dynamic nature and it has significant influence in the organizational processes. So each and every day, small to large scale companies need to update themselves in terms of resources, assets, manpower, capital and infrastructures in order to maintain a competent and healthy business ecosystem.

It is a known and proven fact that there is a need of information-driven culture in a company in order to cater the needs of customers in today's world. For achieving this, they need a tool or service to discover and prioritize business challenges across their organization with these new assessment methodologies and bring information closer to them so that they can make wiser decision.

Choosing the right individual for the right job can change the outcome of a lot of factors in a company aimed towards the effectiveness and good productivity. Each individual should be grouped or assigned to work depending on their performance to get the maximum benefit. If an individual is not best suited in their designated line of work, he/she should be transferred to a team which can take the most use out of those individuals' knowledge and expertise. When the minority of a particular team has the most knowledge of a particular subject, the knowledge will not be equally divided within the team.

When a knowledge person resigns, the company will have to spend time, effort and cost on grooming a new candidate. For the best interests of the company, this particular occurrence should be reduced as well. Nowadays, a lot of apparel and IT companies spend a huge cost on consultation in order to resolve the above mentioned factors. It takes a lot of time and effort for a human to judge an employee's skill and talent. This needs to be addressed as well. Therefore the purpose of this research is to address such mentioned issues.

Hence, knowing more about business and knowing it relevantly faster than others is the best way to gain an edge on business competition. For this there is need of a Business Intelligent tool which can cater the need of today's world. Companies have understood the importance of enforcing achievements of the goals defined by their business strategies through business intelligence concepts [7].

There are commercially available BI tools but below are the drawbacks.

- They are far too expensive which are being developed by big vendors and often target to big client.
- Small and medium scale companies neither can afford cost nor can hire any Business Analyst in order to analyses and predict the behavior, patterns and trends of their business

It has been found that \$150 000 is the average cost of business intelligence software for ETL software (Extract transform and load) which is far too expensive for a small and medium scale companies .And the rule of thumb for cost of effort and services is five times the software

Considering the above facts, we can come to a point that there is a real need of cost-effective Business Intelligent(BI) tools that can cater the need of today's business ecosystem addressing the need of all types of companies ranging from small scale to higher scale.

2 OBJECTIVES

For a project to sustain its way towards success it is a common understanding that the outcome of a team operation depends on the team members and their composition .a proper team should be selected and team combination should be perfect. The Team Dynamics Concepts can be used to achieving this goal. The Proposed **People Clues** is a research project focused on several objectives. With the completion of the project we are supposed to fulfill these research objectives. The main objectives of the research project are mentioned as follows.

2.1 Main Objectives

Objective 1: 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 Apparel and 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.

Objective 2: Adaptable and easy going

The **People Clues** enables users to generate a team for a given project by selecting relevant attributes for team formation depending on their company or industry. Most of the BI tools in the current market have a very complex interface where professional knowledge is required to perform the tasks. **People Clues** will not require technical knowledge or professional expertise to interact and it has been designed in a simple way with less number of controls in order to increase the adaptability and user friendliness.

2.2 Specific Objectives

Objective 3: Build a solution to select most optimal team for the apparel industry

People clues will generate the most optimal team for apparel industry considering attributes related to the apparel 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.

Objective 4: Build a solution to select most feasible team for apparel industry

People clues will generate the most feasible team for apparel industry considering attributes related to the apparel 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.

Objective 5: Build a solution to select most optimal team for IT industry

People clues 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.

Objective 6: Build a solution to select most feasible team for IT industry

People clues 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.

Objective 7: Affordable price

Nowadays lot of apparel companies and IT companies spend a huge cost on consultation when forming teams. For a growing small scale or medium scale companies, these BI tools are too expensive and not affordable. But these companies can invest a small amount of investment and get affective results towards the future goals in the company. Our plan is to serve this community and help them to uplift their business with the help of **People Clues** for the immerging businesses.

Objective 8: Data Analysis

Generally a company which does not consider using a BI Tool for Team Dynamics has a downfall of 95% in the data analyzing section which can give productive information if the data has been analyzed. Our tool analyses the data within a company relevant to Team Dynamics and gives productive information in less amount of time which can lead to analysis of companies' performance and gives vital information in predicting the future of the business.

Objective 9: Business on the Go

People Clues can be accessed where ever you go. It can be accessed from the web with simple clicks and everything will be at the hands of the user, hence the predictions and decisions can be made easily.

Objective 10: Explore and Understand Modern IT techniques

We expect to explore and understand the modern IT techniques such as Team Dynamics, Business Intelligence, Predictive Model Building, Data Mining [3], and Big Data Analysis; and invent innovative solutions which can be adopted for the development of the Team Dynamic tool.

Objective 11: Competitions

Participate for IT competition NBQSA (Business category) and gain knowledge and experience.

3 RESEARCH METHODOLOGY

This section includes detailed descriptions about the techniques and mechanism employed to make **People Clues** 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.

3.1 System Architecture

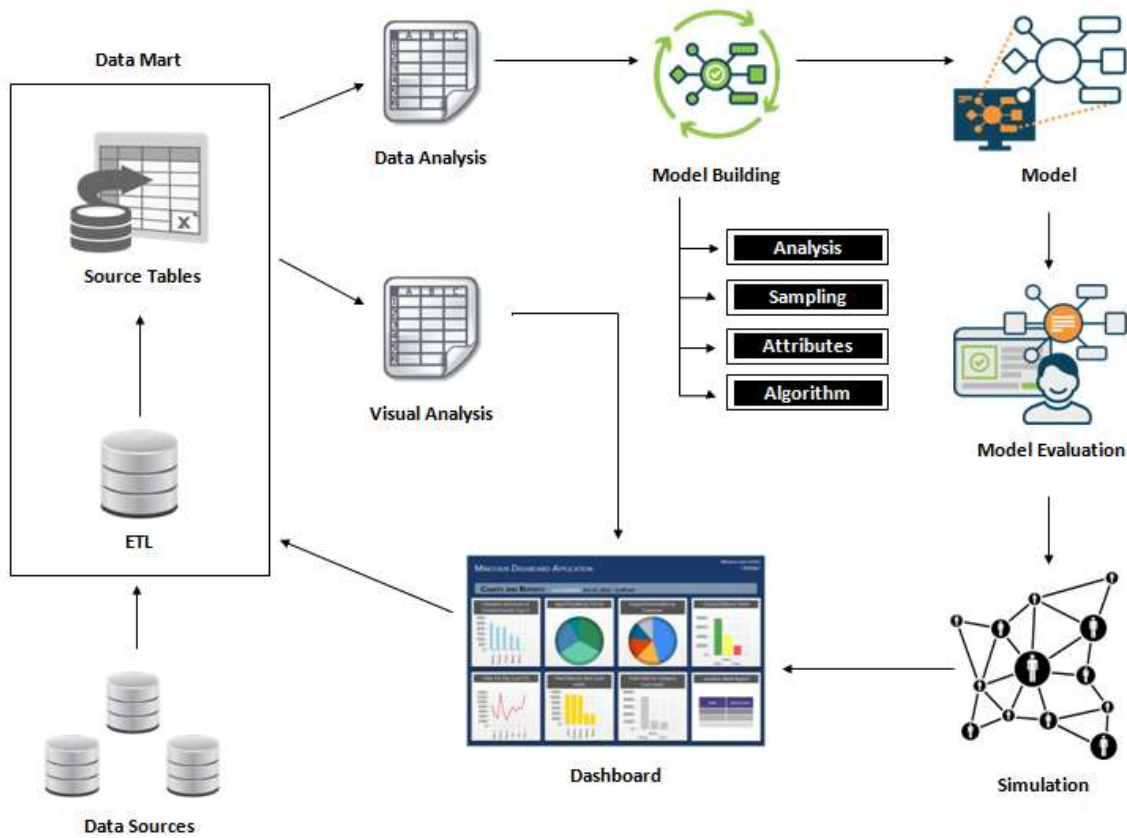


Figure 9 - People Clues system Architecture

In this research we have focused on implementing two arms of **People Clues** application.

1. Desktop Application

This includes core components of **People Clues** which are automated ETL tool, predictive model building and evaluation. These components will be displayed using controls in the main user interface.

2. Web Application

This includes another core component which is **People Clues** dashboard. It is used for visualization of data. It gives you an idea about the overall system in a glance.

There are four primary components that need to be handled in order to complete **People Clues**.

1. ETL Tool
2. Predictive Model Building
3. Predictive Model Evaluation
4. Dashboard Simulation

ETL (Extract Transform Load) Tool

Data in a business can be very useful if and only if analyzed properly which leads to take strategic business decisions. It is acceptable that the data can be in various formats and locations. Thus **People Clues** should be able to extract data automatically in whatever form it is available.

The user can select only required attributes to analyses the data hence **People Clues** will only consume the selected attributes and precede for predictive analysis. Furthermore **People Clues** will automatically detects odd data and remove it from processing. Hence it will result in high-accuracy and best quality output. We only focused on HR systems, Finance systems, Project tracking systems as data sources.

Basically ETL consists of 3 parts – Extract, Transform, and Load.

Extract includes 2 sub-parts, “identification of the data sources” and “extraction” of selected data automatically. In most of the organizations, there can be different data sources related to the system, stored in different locations in different formats such as xls,xlsx, csv, text, xml etc. So the system should be able to identify all these data sources and process them. After identification, necessary attributes has to be extracted and stored in a suitable way for local processing.

After that, data needs to be transformed to a more organized format. This is done in two steps, “data cleansing” and “data transformation”. Data cleansing is performed by detecting and removing and/or correcting a database's dirty data (i.e., data that is incorrect, out-of-date, redundant, incomplete, or formatted incorrectly). Data imputation techniques will be applied in order to cope up with missing data values and remove duplicate data so as to fit it to the models [4]. Then in the transformation step, data may need to be merged, aggregated, enriched, summarized, or filtered depending on the nature of the integration scenario.

Finally the cleansed and transformed data are loaded and stored in data mart for further processing. The data mart is a subset of the data warehouse that is usually oriented to a specific business line or team. “A data mart is a simple form of a data warehouse that is focused on a single subject (or functional area), such as Sales, Finance, or Marketing. Data marts are often built and controlled by a single department within an organization. Given their single-subject focus, data marts usually draw data from only a few sources. The sources could be internal operational systems, a central data warehouse, or external data [4].”

Predictive Model Building

People Clues mainly works with data mining [3], trend and forecasting. Once the data is collected and required, relevant data is selected, **People Clues** will make the statistical model which will generate predictions.

So in building the predictive model **People Clues** will follow the following steps;

- Hypothesis Testing [9].
- Data Sampling [5]-[7].
- Algorithm building; Classification, Association, Factor Analysis and Regression [4].

This will generate a team for a given project depending on whether it is optimal or feasible. This optimality or feasibility will be selected during the prediction of the predictive model. With the help of this predictive model, the business performance can be shown as how it has been in the past, present and what will happen in the future, thus business predictions can be made wisely and easily.

The predictive model should be made up of a number of predictors, which are variable factors that are likely to influence future behavior or results. First step is to sample the main dataset to obtain a representative, and statistically valid sample of the whole using a suitable sampling technique. A proper sampling technique such as randomized statistical sampling or probability sampling like cluster sampling and stratified sampling must be used for this purpose[6]. Using those techniques the system has to produce Test Set, Training Set and Holdout set in order to bench mark the models and find out the most accurate model.

Next, a suitable algorithm needs to be selected, based on the type of prediction required. This can be classification, association, factor analysis or regression. As each of these algorithms have different behaviors and produce different results, selection of proper algorithm has to be done with proper justifications.

Finally, selected algorithm needs to be customized to suit current scenario based on available data and produce final prediction model.

Predictive Model Evaluation

Predictive Model Evaluation is carried out by using 3 techniques - Classification Table (Confusion Matrix), ROC (Receiver Operating Characteristic) and Model Accuracy [8].

Classification Table shows the number of correct and incorrect predictions made by the classification model compared to the actual outcomes (target value) in the data.

ROC curve is a graphical plot that illustrates the performance of a binary classifier system as its discrimination threshold is varied. The curve is created by plotting the true positive rate against the false positive rate at various threshold settings.

Model Accuracy is measured by the ratio of correct predictions to the total number of cases evaluated.

Dashboard Simulation

Once the data is analyzed the Dashboard of **People Clues** will present it in a format which will make the user understand the distinction between raw data and predictive data in a meaningful way that would be able to depict current status of the entity. This data representation will help top management to take decisions wisely. **People Clues** dashboard will present data in an interactive way using graphs, charts, hierarchies and tables. Dashboard will provide summary of the results and key points regarding analyzed data. This approach enables user to generate an optimal or feasible team with their specific roles for a given characteristics of a project. This is not just a static dashboard where user is presented with a set of predefined datasets every time it loads. The system is expected to provide better interaction with the user through interactive

dashboard. While user can view high level result of the prediction, user may also trace selected data points and explore much deeper to find more detailed information of them.

Research area 1: Business Intelligence - To improve dynamic team formation process

Business intelligence (BI) is a technology-driven process for analyzing data. The proposed system enables the organizations to collect data from internal systems, prepare it for analysis, develop and run necessary queries against the data. According to the results team hierarchy is generated for a given project.

Research area 2: Data mining - To identify major factors and impact of attributes for team dynamics

Data mining is primarily useful for analyzing huge set of data from different perspectives and summarizes them into useful information [5]. Our system **People Clues** enables to determine relationships among "internal" factors such as key performance indicators (KPI), team sizes, job roles, project characteristics (deadlines , estimated effort, cost) or staff skills and employee historical data. It enables users to determine the suitable team for a given project characteristics. Finally, it enables them to drill down into summary information to view detail data.

Research area 3: Predictive Analytics – To predict the optimal or feasible team

People Clues extracts information from existing data sets in order to determine optimal or feasible team for a given project. In the process different predictive models has to be developed in order to decide optimal or feasible team in IT and Apparel industry. Depending on the optimality or feasibility a team will be formed with high accuracy.

3.2 Tools and Techniques

Tools:

- PyCharm
- DBMS Tools
- Adobe Dreamweaver
- IBM SPSS
- Hadoop

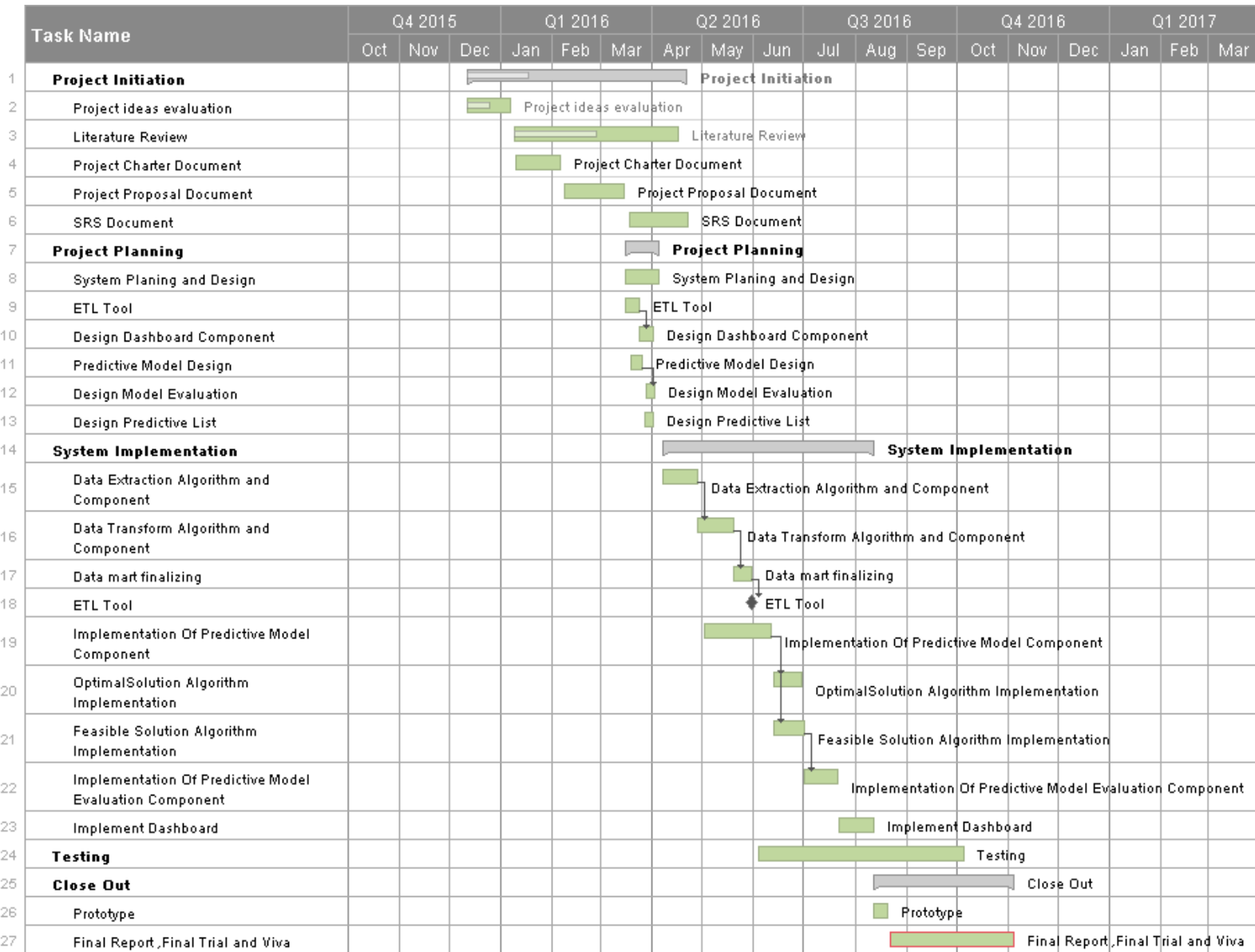
Technologies:

- Python
- MySQL, MS SQL
- HTML5
- Java
- Php

3.3 Testing

- Unit Testing – Each unit of the system will test by the group member who is developing that particular unit and will produce a defects free unit of coding.
- Component Testing – Several bug free units will combined together and test. Each member will combine their tested units together and will test them.
- Integration Testing – In this testing level users are responsible to test whether the relationships and communication between tested components are working as expect.
- System Testing – All the components from each group member will combine together and test the whole system to verify the functionality and the performance of it.
- Beta Testing – After Integrate the system, complete project will be release as beta version to the selected IT and apparel companies, ongoing test the whole system to verify the functionality and the performance of it, and finally get user feedbacks.

3.4 Gantt chart



3.5 Work Breakdown Structure

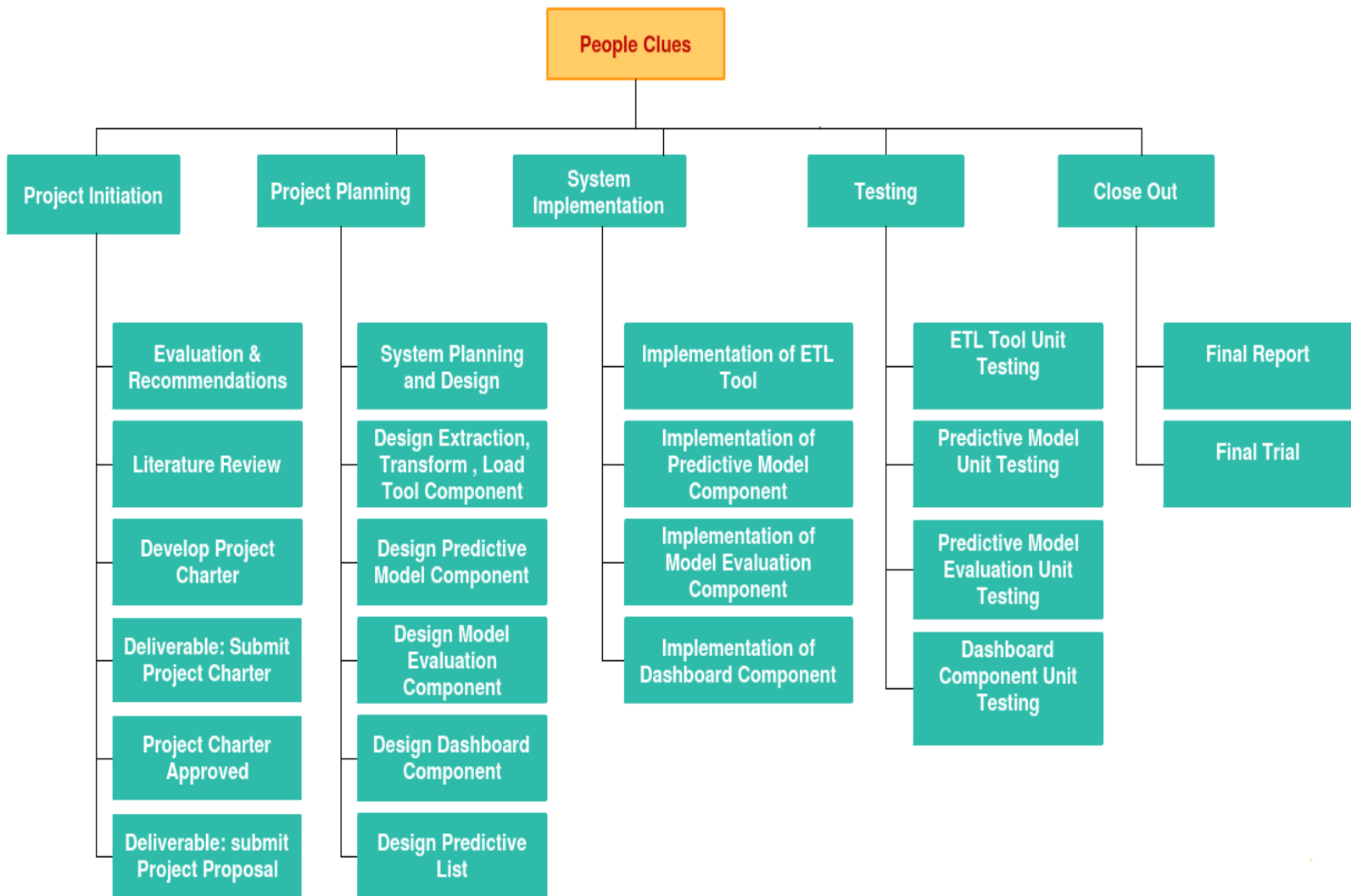


Figure 10 - Work Breakdown Structure

4 DESCRIPTION OF PERSONAL AND FACILITIES

Member	Component	Task
Perera B.G.A IT13082598		<p>(All the tasks are related to IT Industry)</p> <ul style="list-style-type: none"> • Identification of attributes with the use of relevant data from different sources. • Cleaning unorganized data (Data cleansing) • Extracting and saving the selected data and in a proper format for further processing. • Transforming data into different formats as needed • Loading cleansed and transformed data into database. • Creation of the Data Warehouse/Data Mart • Data quality management • Develop the desktop application including ETL (Using Hadoop) and Prediction model. • Sampling the data using a proper sampling technique and data mart. • Identification of the most influential factors • Developing algorithms and techniques for prediction making using above identified factors. • Implementing the predictive models using the data in the data ware house. • Developing algorithms for feasible solution. • Monitor the accuracy of the prediction models and evaluate the model. • Automate the model. • Publish the model in cloud platform • Implementation of the dashboard of the application • Develop dynamic simulation for users. • Documenting • Testing
Perera L.N.P IT13009700		<p>(All the tasks are related to IT Industry)</p> <ul style="list-style-type: none"> • Identification of attributes with the use of relevant data from different sources. • Cleaning unorganized data (Data cleansing) • Extracting and saving the selected data and in a proper format for further processing. • Transforming data into different formats as needed • Loading cleansed and transformed data into database. • Creation of the Data Warehouse/Data Mart • Data quality management • Develop the desktop application including ETL (Using Hadoop) and Prediction model.

		<ul style="list-style-type: none"> • Sampling the data using a proper sampling technique and data mart. • Identification of the most influential factors • Developing algorithms and techniques for prediction making using above identified factors. • Implementing the predictive models using the data in the data ware house. • Developing algorithms for optimal solution. • Monitor the accuracy of the prediction models and evaluate the model. • Automate the model. • Publish the model in cloud platform • Implementation of the dashboard of the application • Develop dynamic simulation for users. • Documenting • Testing
Samarasingha H.K.L IT13068974		<p>(All the tasks are related to Apparel Industry)</p> <ul style="list-style-type: none"> • Identification of attributes with the use of relevant data from different sources. • Cleaning unorganized data (Data cleansing) • Extracting and saving the selected data and in a proper format for further processing. • Transforming data into different formats as needed • Loading cleansed and transformed data into database. • Creation of the Data Warehouse/Data Mart • Data quality management • Develop the desktop application including ETL and Prediction model. • Sampling the data using a proper sampling technique and data mart. • Identification of the most influential factors • Developing algorithms and techniques for prediction making using above identified factors. • Implementing the predictive models using the data in the data ware house. • Developing algorithms for feasible solution. • Monitor the accuracy of the prediction models and evaluate the model. • Automate the model. • Publish the model in cloud platform • Implementation of the dashboard of the application • Develop dynamic simulation for users. • Documenting • Testing

Senaratna P.M. IT13024314		<p>(All the tasks are related to Apparel Industry)</p> <ul style="list-style-type: none"> • Identification of attributes with the use of relevant data from different sources. • Cleaning unorganized data (Data cleansing) • Extracting and saving the selected data and in a proper format for further processing. • Transforming data into different formats as needed • Loading cleansed and transformed data into database. • Creation of the Data Warehouse/Data Mart • Data quality management • Develop the desktop application including ETL and Prediction model. • Sampling the data using a proper sampling technique and data mart. • Identification of the most influential factors • Developing algorithms and techniques for prediction making using above identified factors. • Implementing the predictive models using the data in the data ware house. • Developing algorithms for optimal solution. • Monitor the accuracy of the prediction models and evaluate the model. • Automate the model. • Publish the model in cloud platform • Implementation of the dashboard of the application • Develop dynamic simulation for users. • Documenting • Testing
--	--	---

5 BUDGET (If Any)

Description	Cost Rs. (Approximately)
Microsoft Azure App Service and SQL Database	10000.00
Other Expenses (Stationaries)	2000.00
Total project cost estimate	12000.00

6 REFERENCES

- [1] Shuib Basri, and Rory V. O'Connor, "Software Development Team Dynamics in SPI: A VSE Context," In Proc. Asia-Pacific Software Engineering Conference 19 th'04, 2012, pp.1-2.
- [2] Markus Waibel, Member, IEEE, Laurent Keller, and Dario Floreano, Senior Member IEEE, "Genetic Team Composition and Level of Selection in the Evolution of Cooperation," IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION, VOL. 13, NO. 3, pp. 649, JUNE 2009.
- [3] [Online]. Available: <http://www.bisoftwareinsight.com/business-intelligence-stats/> [Accessed: March 01, 2016].
- [4] Elena M. Hernández-Pereira, Diego Álvarez-Estévez and Vicente Moret-Bonillo, "Automatic Classification of Respiratory Patterns Involving Missing Data Imputation Techniques", Biosystems Engineering, 138.March 2016 ,2015, 65–76 .
- [5] Oracle Documentation, "Oracle® Business Intelligence Standard Edition One Tutorial," Oracle, E10312-01. [Online].
Available: http://docs.oracle.com/html/E10312_01/dm_concepts.htm. [Accessed: March. 02, 2016].
- [6] Jeffrey D. Wells and others, "An Evaluation of Sampling Methods Used to Produce Insect Growth Models for Postmortem Interval Estimation", International Journal of Legal Medicine, 129.2 ,2015, 405–10
- [7] [Online]. Available: <http://web-assets.domo.com/blog/wp-content/uploads/2013/11/2013-Data-Driven-Marketing-Survey-Report-Final.pdf>[Accessed: March 7, 2015]
- [8] Keith W Buffinton, Kathryn W Jablow and Kathleen a Martin, "Project Team Dynamics and Cognitive Style", Operations Management, 14.3 ,2002, 570–77
- [9] [Online]. Available: <http://www.312analytics.com/analysis-types-reporting-hypothesis-testing-data-mining/>[Accessed: March 01, 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] Nahid Saberi, Mohsen Mahvash, and Marco Zenati, Member, IEEE, "An artificial system for selecting the optimal surgical team,"

- [12] Zhongcheng Zhang, Xuefan Yang and Weikai Liu, "Optimization Model for Team and Member Selection in Mathematical Contest in Modeling," IEEE International Conference on Wireless Communications' 2010, pp.60 - 63.
- [13] Waibel, Markus, Keller, Laurent, Floreano, Dario, "Genetic Team {C}omposition and {L}evel of {S}election in the {E}volution of {M}ulti- {A}gent {S}ystems," {IEEE} {T}ransactions on {E}volutionary {C}omputation, vol. 13, no. 12, July, 2005. [Online serial]. Available: <http://infoscience.epfl.ch/record/111745/files/Waibel09-IEEE.pdf>. [Accessed Mar. 2, 2016].
- [14] Agustín-Blas, Luis E.,Salcedo-Sanz, Sancho, Ortiz-García, Emilio G.,Portilla-Figueras, Antonio,Pérez-Bellido, Ángel M.,Jiménez-Fernández, Silvia, "Team formation based on group technology: A hybrid grouping genetic algorithm approach Computers & Operations Research, vol. 38, issue. 2, p. 484-495, 2011. [Online serial]. Available: sciencedirect, <http://www.sciencedirect.com/science/article/pii/S0305054810001371>. [Accessed March 06, 2016].
- [15] E. Semsar-Kazerooni, K. Khorasani, "A game theory approach to multi-agent team cooperation" American Control Conference, 2009. ACC '09.' 2010, pp. 4512 - 4518.
- [16] Chen, Shi Jie Gary, Lin, Li, "Modeling team member characteristics for the formation of a multifunctional team in concurrent engineering", vol. 51, issue. 2, p. 111-124, 2004.[Online serial]. Available: researchgate.net, https://www.researchgate.net/profile/Shi_Jiegary_Chen/publication/3076819_Modeling_Team_Member_Characteristics_for_the_Formation_of_a_Multifunctional_Team_in_Concurrent_Engineering/links/53f227f40cf2f2c3e7fcc723.pdf. [Accessed March 06, 2016].
- [17] Xu, Y, Ma, J, Guo, X, "Modeling researchers' characteristics for the formation of research team", 2012. [Online serial]. Available: pacis-net.org, [http://www.pacis-net.org, http://www.pacis-net.org/file/2012/PACIS2012-182.pdf](http://www.pacis-net.org/file/2012/PACIS2012-182.pdf). [Accessed March 06, 2016].
- [18] Becchetti, Luca,Castillo, Carlos, "Online Team Formation in Social Networks", p. 839-848, 2012. [Online serial]. Available: research gate, https://www.researchgate.net/profile/Carlos_Castillo/publication/262240293_Online_Team_Formation_in_Social_Networks/links/548aeb130cf214269f1d7243.pdf. [Accessed March 06, 2016].
- [19] [Online]. Available: <http://www.softwareadvice.com/bi/> [Accessed: March 07, 2016]

- [20] [Online]. Available: <https://www.oracle.com/solutions/business-analytics/business-intelligence/index.html> [Accessed: March 07, 2016]
- [21] Elena del Val, Juan Miguel Alberola, Victor Sanchez-Anguix, Alberto Palomares, Ma Dolores Teruel, “A Team Formation Tool for Educational Environments””, vol. 293, p. 173-181, 2014.[Online serial]. Available: http://link.springer.com/chapter/10.1007/978-3-319-07476-4_21#page-1

7 APPENDICES (If Any)