

# Proposal for Research Methodology

PA2537: Research Methodology in Software Engineering and Computer science

Version NUMBER – April 23, 2015

Thesis	Tentative title	Improvising algorithms in descriptive data analytics related to Big Data
	Classification	Descriptive data analytics, Big Data, data analysis, cm-spade algorithm, cm-clasp algorithm
Student 1	Name	Siva Venkata Prasad Patta
	e-Mail	<a href="mailto:Sipa15@student.bth.se">Sipa15@student.bth.se</a>
	Social security nr	9312217137
Student 2	Name	Venkata Sesha Sowjanya Mudrageda
	e-Mail	<a href="mailto:Vemu15@student.bth.se">Vemu15@student.bth.se</a>
	Social security nr	9403295703
Supervisor	Name and title	Samuel A. Fricker
	e-Mail	<a href="mailto:Samuel.fricker@bth.se">Samuel.fricker@bth.se</a>

## 1. Introduction

Descriptive analytics is a part of big data which uses data aggregation and data mining techniques to summarize the raw data in order to interpret it. Generally the descriptive analytics deals with the past time which has already occurred. It is also useful for reporting the past events of a company's financials, sales, customers etc. The event which occurred even a minute ago also come under descriptive analytics. The descriptive data analytics provides a scope to study the past action of certain things and helps us to anticipate the future outcomes related to those things.

The Descriptive statistics are used to study the aggregate level of a company's market and give us suggestion to increase the sale of the company. It helps in completely studying each and every work that take place in a company right from its starting time. The raw data collected is interpreted to a meaningful data which helps in finding out the faults in the company and gives few counter steps to overcome the problems faced by the company.

## Related Works

A researcher, Philippine Fournier-Viger, an assistant professor in the University of Moncton has done extensive research in the field of Data Mining, especially about Descriptive Data Analytics. He has also proposed two algorithms called CM-clasp and CM-Spade which helped in retrieving data for the topic.

## 2. Aims and Objectives

Aim:

- To find the loopholes in the CM-clasp and CM-Spade algorithms.
- To solve the problems that are associated with Descriptive Data Analytics.

Objectives:

- Systematic literature review on “Descriptive data Analytics” is done to gain an idea of flaws in the existing algorithms in order to retrieve data.
- Analyze the existing methods for identifying the drawbacks in the algorithms.
- Identify the factors that determine the amount of information retrieved and trying to predict a method that could improvise the retrieval of data.

## 3. Research Questions

**RQ1:** Given a set of transactions, finding all the association rules with confidence and support greater than or equal to specified thresholds is still a problem. How to overcome this?

**RQ2:** Vertical algorithms generate a large amount of infrequent candidates hence pruning the unwanted candidates is another major issue with respect to the mining of sequential patterns. This has to be improvised in the existing algorithms. Finding the loopholes and proposing easy methods to sort this out has been a very important aspect for researchers to figure out recently.

## 4. Method

*Method used for answering RQ1:*

Literature review is done for answering RQ1 to get a better idea on the algorithms. Solve this by finding association rules for a given support, we mine the sequential patterns, by using experimentation rules.

*Method used for answering RQ2:*

To prune the unnecessary items we use co- occurrence mapping method, which could also be classified as experimental method.

## 5. Expected Outcomes

Descriptive analysis is a field that is still in its blooming stage. An algorithm, which can retrieve the required information from already existing raw material has been proposed. But that algorithm has some loop holes which we are expecting to overcome.

## 6. Risk management

The possible risks that might occur are as follows:

- **Convenient programming language selection and use:** Programming language could be chosen by the writer, according to his convenience, expertise and as supported by the software tools used for executing the experiment.
- **Unavailability of sufficient data:** As it would be a pilot study, we will not be able to give data in big quantity which could suffice the need of descriptive data analytics. The result may or may not vary when done in a real field.

## 7. Time Plan

The time plan for implementing is as follows:

Topics covered with in the time duration	Start Date	Duration	End Date
SLR submission	3/19/2015	1	3/19/2015
Start of Research proposal	3/20/2015	28	4/16/2015
Search string formed based on the selected topic	3/20/2015	3	3/22/2015
Search for related papers	3/23/2015	9	3/31/2015
Selecting suitable journals	4/1/2015	9	4/9/2015
In depth study of articles/papers/journals	4/10/2015	7	4/16/2015
Research Proposal	4/17/2015	10	4/26/2015
Initiation of Research Proposal	4/17/2015	5	4/21/2015
Evaluating the Paper	4/22/2015	4	4/25/2015
Submission of the proposal	4/26/2015	1	4/26/2015
Implementation	4/27/2015	30	5/24/2015
Model the planned proposal	4/27/2015	15	5/10/2015
Rechecking the proposal method	5/11/2015	6	5/15/2015
Document the finalized model	5/16/2015	5	5/20/2015
Final Document	5/21/2015	3	5/23/2015
Submission	5/24/2015	1	5/24/2015

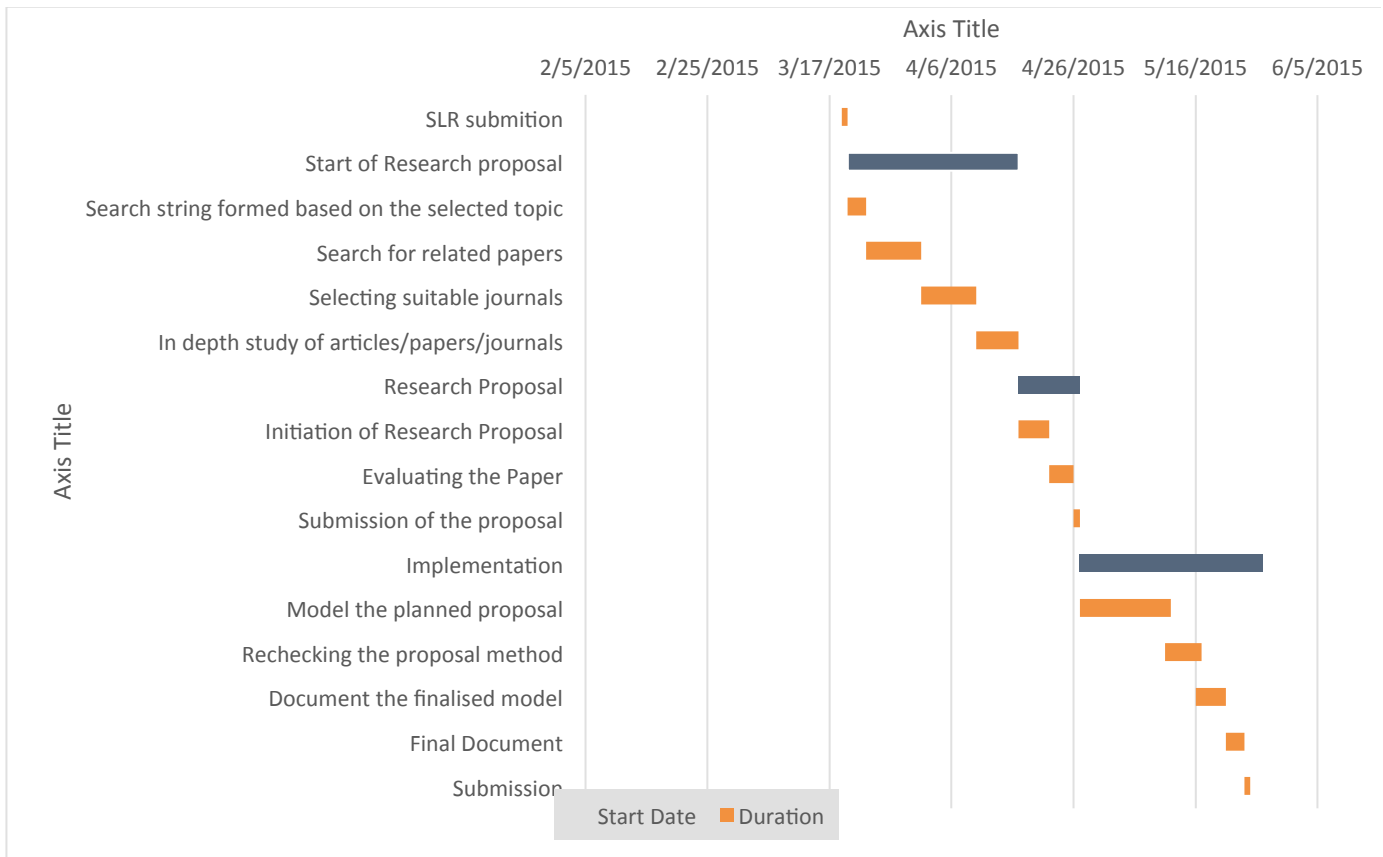


Figure 1: Gantt chart

## References

- [1] Pura, M.-L. (Dept. of Mil. Inf. & Math., Mil. Tech. Acad., Bucharest, Romania); Patriciu, V.-V. "Security analysis of robust user authentication protocol", *Proceedings 2010 8th International Conference on Communications (COMM)*, p 457-60, 2010
- [2] Joseph Bonneau, Cormac Herley, P.C. van Oorschot, Frank Stajano, "The Quest to Replace Passwords: A Framework for Comparative Evaluation of Web Authentication Schemes", *2012 IEEE Symposium on Security and Privacy*, p 553-67, 2012.
- [3] S.Ghandali and M. Ebrahimi Moghaddam, "Off-Line Persian Signature Identification and Verification based on Image Registration and Fusion", *Journal of Multimedia*, Vol 4, No 3 (2009), 137-144, Jun 2009.
- [4] Meenakshi S Arya, and Vandana S Inamdar, "A Preliminary Study on Various OffLine Hand Written Signature Verification Approaches". *International Journal of Computer Applications* 1(9):55–60, February 2010.
- [5] Indrajit Bhattacharyaa, Prabir Ghoshb, and Swarup Biswasb, "Offline Signature Verification Using Pixel Matching Technique", *Procedia Technology* Volume 10, 2013, Pages 970–977, International Conference on Computational Intelligence: Modeling Techniques and Applications, (CIMTA) 2013.