

Project Management Plan

Analyzing Breast Cancer Datasets using Big Data Analytics

1. Project Summary

1.1 Project Overview

- BC (Breast Cancer) has become one of the most prevalent cancers diagnosed in women. If BC is identified at an earlier stage, it can help save lives of those affected by Breast Cancer. The tumors are classified into two types - Malignant and Benign. The cancerous tumors are Malignant and the noncancerous tumors are Benign. In the present scenario, it has become a challenging task to detect the occurrence of Breast Cancer tumors and the process takes several days to predict the same. By the use of proper algorithms pertaining to Big Data and Machine Learning this task can be achieved with much ease.

1.2 Project Scope

The research is being carried out with the following objectives:

(i) To study various big data and machine learning approaches for breast cancer diagnosis through their implementation.

(ii) To make a comparative study of the approaches.

For the purpose of research, I have considered only the classification task involved in such systems and used the existing feature space. The extraction of features from the mammographic images is not considered. The machine learning approaches that were considered here could be used for any other classification problem. I have focused mainly on breast cancer diagnosis, a medical domain problem.

1.3 Development Process

We follow the waterfall model of software development as it is simple and small.

1.4 Effort, Schedule and Team:

The team comprises of the following 3 persons:

Total Effort : 3 person-months

Project duration : 10 months

1.5 Assumptions made:

No major assumptions beyond what is stated in the SRS.

2. Detailed Effort and Schedule

A : ANURAG DEVAGIRI

B : POOJA GUNAGI

C : NIKITA JOSHI

#	Task	Estimated Effort (person - days)	Start Date (dd/mm/yyyy)	End date (dd/mm/yyyy)	Person
1	System design	9	Aug 10	Aug 20	C
2	Detailed design	13	Aug 21	Sep 5	A, B, C

3	Coding Input module	8	Sep 15	Sep 30	A, B, C
4	Coding Schedule module	8	Oct 5	Oct 15	A
5	Coding Output module	6	Oct 20	oct 30	C, B
6	Test planning	3	Nov 2	Nov 15	B
7	Testing and integration	5	Nov 20	Nov 30	A, C
8	Rework and final	3	Dec 5	Dec 12	A, B, C

The total estimated effort in person-days is: **53**

3. Team Organization

We will have a small team of three persons Nikita, Anurag and Pooja . We use a flat team structure of peers with one person having an additional role of project manager.

The assignment of tasks to each team member will be maintained in the detailed schedule, a high-level view of which is given in the table above The specific tasks are assigned to either a specific person or a combination of 2 or all 3 members depending upon the complexity and time required to complete the module.

4. Hardware and Software resources required

The only hardware resource required is:

- > 8GB RAM
- > Core i5 processor
- > 10GB free memory space
- > Hard Disk: 1 TB

The only software resource required is:

- > 64bit OS
- > Python 3 and above
- > OS: Windows 10
- > Spyder Editor

5. Quality Plan

The quality control process for this project will consist of the following:

- **SRS Review:** The SRS will be reviewed by a team.
- **Design Review:** Design document will be reviewed by the project team.
- **Unit Testing:** Each programmer is responsible for Unit Testing his module.
- **System Testing:** Will be done according to the system test plan, which will be reviewed.
- **Integrated Testing:** will be integrating all the components to check the final output and it will be reviewed by project coordinate and project guide.

6. Risk Management Plan

There are no risks with this project that might need any explicit mitigation.

7. Project Tracking

Three basic methods will be used for monitoring –

1. project logs

2. weekly meetings

3. reviews

As there is no timesheet system, each project member will record his activity in a project notebook and report the hours for each activity in the meetings.

Reviews will be held as per the quality plan.