

Project Management Plan

Project Goal: Build a classification model to detect whether a person has cancer or not.

Duration: 6 Weeks (3 Sprints × 2 Weeks)

Team Size: 6 people

Methodology: Agile with Scrum framework

◆ Sprint 1 (Week 1–2): Data Preparation & Project Setup

Objective: Set up environment, collect/clean data, explore dataset, and prepare for modeling.

Tasks:

1. Project Setup

- Define project scope, success metrics (accuracy, precision, recall, F1, AUC).
- Set up Git repo, project tracking (Jira/Trello), shared storage.
- Assign roles.

2. Data Collection & Understanding

- Gather datasets (medical/public cancer datasets or hospital-provided).
- Perform exploratory data analysis (EDA).
- Handle missing values (continuous & categorical).
- Understand class imbalance (if many "no cancer" vs few "cancer").

3. Preprocessing

- Feature engineering (scaling, encoding categorical variables, handling outliers).
- Data split into train/validation/test.

Deliverables:

- Cleaned, well-documented dataset
- System Requirement Specification draft
- Preprocessing pipeline ready

Roles:

- Data engineer
 - Data scientists
-

◆ **Sprint 2 (Week 3–4): Model Development & Training**

Objective: Train baseline models, optimize, and compare performance.

Tasks:

1. Baseline Modeling

- Train simple models (Logistic Regression, Decision Tree).
- Evaluate performance (precision, recall, ROC curve).

2. Advanced Modeling

- Train more complex models (Random Forest, XGBoost, SVM, Neural Networks).
- Handle class imbalance (SMOTE, class weighting).

3. Model Evaluation

- Compare models across metrics (accuracy isn't enough — emphasize recall & F1 because false negatives are critical in cancer detection).
- Perform cross-validation.

4. Hyperparameter Tuning

- Use GridSearch / RandomSearch or Bayesian optimization.

Deliverables:

- Trained candidate models
- Model comparison report (with metrics)
- Best model identified

Roles:

- Data scientists
- Machine learning engineer

◆ Sprint 3 (Week 5–6): Deployment, Testing & Final Report

Objective: Finalize best model, test robustness, deploy, and prepare project documentation.

Tasks:

1. Model Finalization

- Select best-performing model.
- Test on holdout test set.
- Check fairness & bias (avoid false negatives).

2. Deployment

- Build API/Flask/Django app for model inference.
- Containerize with Docker (optional).
- Prepare demo UI (simple web form for input + output).

3. Project Documentation & Reporting

- Prepare final report (dataset description, methods, results, conclusion).
- Create presentation deck.
- Assign team members to present.

Deliverables:

- Final cancer detection model (with reproducible code)
- Deployed demo (API/web app)
- Final report & presentation

Roles:

- Quality test engineer
 - ML op's
-

◆ Project Timeline (Summary)

Sprint	Week	Focus Area	Key Deliverables
Sprint 1	1–2	Data collection, cleaning, preprocessing	Clean dataset, EDA report, preprocessing pipeline
Sprint 2	3–4	Modeling & evaluation	Trained models, comparison report, best model
Sprint 3	5–6	Deployment & documentation	Final model, deployed app, final report & presentation

◆ Roles in Project

- Project manager – Rohit Singh
- Data engineer-Dhrithi MV
- Data scientists - Lavanya K Gowda
- Machine learning engineer- Kusumitha K P
- Quality test engineer - Sourabh Rajendragouda Doddagoudar
- ML op's – Keerthana V and Likhitha D