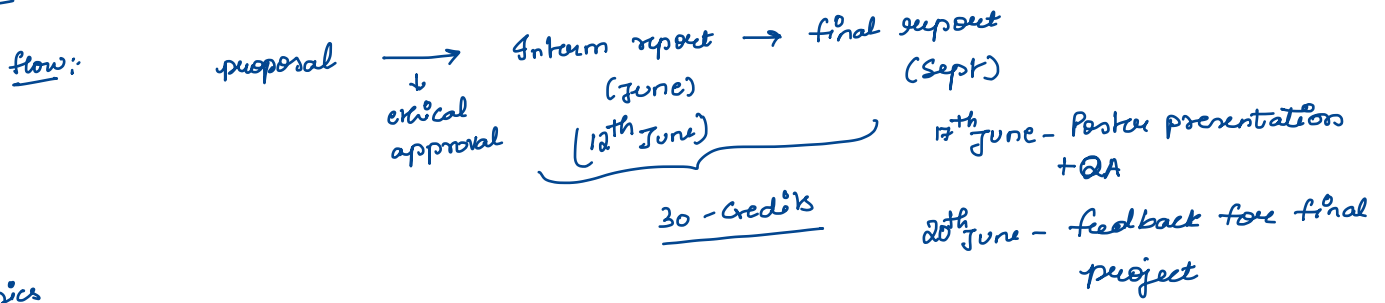




22/01/2025



Topics

- P1.3 1. Financial Forecasting with Generative AI.
(32) A generative AI-based project for predicting and analyzing financial trends.
- P1.4 2. Nutritional Analysis for Teenagers
(57) A DA project to assess and improve nutritional intake for teenagers.
- P1.2 3. SMS Spam Detection with Machine Learning.
(25) A machine learning model for detecting & filtering spam messages.
- P1.1 4. Predictive health Analytics for Early Detection
(13) A Tensor flow powered system to analyze health data for the early detection of diseases.

03/03/2025 finding Base Research Papers and Dataset

1. Performance Analysis of ML Algorithms in CKD prediction.

- Got dataset from M.L UCL repo
- Implemented 8 M.L models for CKD to evaluate their performance.

09/03/2025

Predicting heart disease



- = Method is different.
- = how that attribute is used.
- = simplicity training method.
- = different training method.
- = cost efficiency.
- = dataset

latest review papers

24/03/2025 Monday call

- * Create standard leaflet in Jyot hub.
- * Mention we are using existing dataset, we are not creating a new one in ethical form and submit.
- * Update MOM
 - decide on which model
 - work on 20 papers + summarize.
- *** update 20 papers + summarize
 - ✓ 5 Papers
 - * 5 Papers
 - * 5 Papers
 - * 5 Papers.

Topics to learn (Models/Algorithm)

- * KNN algorithm
- * Logistic regression
- * Naïve Bayes
- * Deep Neural Network
- * Deep learning techniques in Weekly topics

25/03/2025 Models we can use

1. KNN :- gives accurate and precise predictions.
 - the distance measures determine how accurate the predictions are

Points to remember while writing literature review

- * Write what techniques they used > what evaluation metrics they considered > what results they achieved.
- * We can also mention how they trained their base model - what accuracy they obtained.
- We can further mention what preprocessing techniques they used and what are the results {computational time, evaluation metrics}

Topics to learn

- * Soft voting.
- * hard voting.
- * ANOVA f-test.

07/04/2025.

— 12 papers of lit review.

* Call with Dr. Zohaib Ijaz.

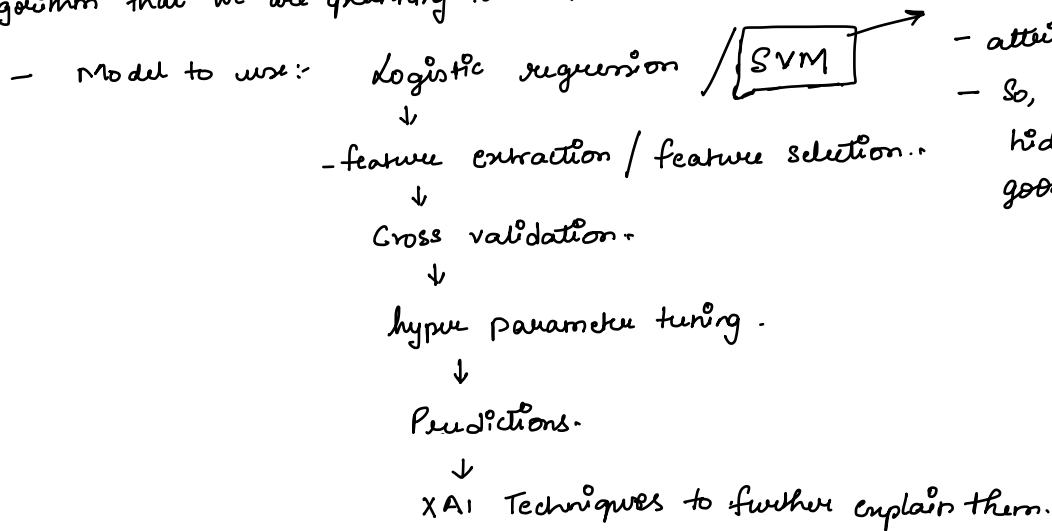
* create minutes of meeting.
update to GitLab

— XAI Techniques.

— Evaluation.

* Daily 2 research papers.

Algorithm that we are planning to use:-



dataset is small

- attributes are very crucial
- So, SVM helps identifying hidden relations and gives good performance.

16/07/2025

Journal

Commits

July

June 30th - July 4th - Sick

7th - July 11th - Data set merging + EDA on merged data set

14th - July 18th - Data pre-processing.

21st - 25th -

28th - Aug 1st -

16th,

August

Aug 4th - 8th -

11th - 15th -

18th - 22nd -

25th - 29th -

Report

(8000 - 10,000) words

Abstract

+ Background

+ Introduction.

→ Create Base format

→ Abstract

→ Clear and concise research questions

→ Problem definition.

→ Project relevance in field of data analytics.

→ Research methodologies used.

→ Structure of the report.

→ Literature review

→ Data and Methods

→ Results, discussion & findings

→ Conclusion & future work

19/07/2025 EDA on combined dataset + Data preprocessing on combined dataset

* EDA Completed on Combined dataset. → Committed final version to GIT.

* uploaded the combined dataset. → Committed final version to GIT

* Created weekly journal in GIT lab. → Created individual directory from week 1-10

* Updated the weekly Summary of week 1,2,3.

Next Steps :-

* Data pre processing + filling NULL values using regression techniques.

21/07/2024

* Performed additional bi-variate analysis

10/08/2025.

* Apply Random forest M.L model → Get Evaluation metrics.

* Apply M.L Techniques → Get Evaluation metrics.

* Apply XAI Techniques → Get final conclusions.

16/08/2025

Applying Random forest model

- * We applied RF model \rightarrow 58.9% accuracy
 \rightarrow calculated accuracy
- * We performed hyper parameter tuning
 \rightarrow calculate "best-param"
 \rightarrow "best-score" \rightarrow 65.1% accuracy.

26/08/2025

We apply Bayesian Search.

- It works on Exploration and exploitation principles
- \rightarrow Exploration \rightarrow It explores new regions and gets the scores
 - \rightarrow Exploitation \rightarrow It tries the values around the value which has given better score.
- \rightarrow It finds good hyper parameter values in fewer trials.
 \rightarrow It is smarter than grid search & Random search.

27/08/2025

Requirements - {scikit-optimize}

data pre-processing :-

THAT, slope, serum-cholesterol, num-major-vessels.

\rightarrow since we are having skewed distribution, we shall use KNN Imputer to fill the rows.

- * Instead of using mean, median or mode we shall use KNN Imputer where the missing values are filled based on similar rows.

06/09/2025

	<u>Word count</u>	<u>Previous</u>	<u>Current</u>	
1.	Introduction	432	1671	
2.	Literature review	1352	800 - State of art 1352 - prev lit review 782 - additional 5 papers	} (2874)
3.	Exploration of dataset and methods	706	1541	
4.	Machine Learning Techniques	573	1092	} 7527
5.	Explainable AI Techniques	278	704	
6.	results	451	451	

(5,280 + 800 (SOA))

7. Evaluation

268

2213

8. Conclusion
+ future work

202

4262

552

11092 words — including words in tables.

Email to Zohaib Sir

- * changed report format from research paper \rightarrow thesis
- Changes made
 - * Updated Introduction with detailed matter mentioned in guidelines
 - * updated lit review with SOA tables + 5 additional papers
 - * updated Exploration of dataset and methods with univariate EDA graphs + Interpretations.
 - * updated evaluation sections with outputs and detailed interpretations.