BSc (Hons) Computing Course 2024/25

Level 6 Production Project

Name: Avishek Sah	**Student I.D.: 77356813**
Course: BSc (Hons) Computing	**Supervisor's Name: Rohit Raj Pandey**

Final Project Individual Aim & Objectives

Title of my Project: Prediction of Cardiovascular disease using transformer model

Aim of my Project: This project aims to develop an advanced predictive model that uses self-attention mechanisms in transformer architecture to accurately assess the risk of cardiovascular diseases (CVD) based on patient health data.

Objectives of my Project:

- Gain understanding on transformer model and its concepts.
- Develop a self-attention-based transformer model and understand the advantage over other models.
- Pre-process a dataset with features from patient data, such as age, cholesterol levels, and blood pressure.
- Test the model thoroughly against basic models to make sure it is accurate and reliable.
- Develop a website to take the input form the patient and show the prediction.

Specification of my Product:

This project combines research and product development. The system will function as:

- A publicly available cardiovascular disease dataset will be utilized for training and testing the model.
- The system will use web interface where users can input patient health data and receive predictions about CVD.
- A transformer-based model with self-attention mechanisms will be implemented using PyTorch.

- Libraries like Pandas and Scikit-learn will be used for data pre-processing tasks to prepare the input data for model training and prediction.

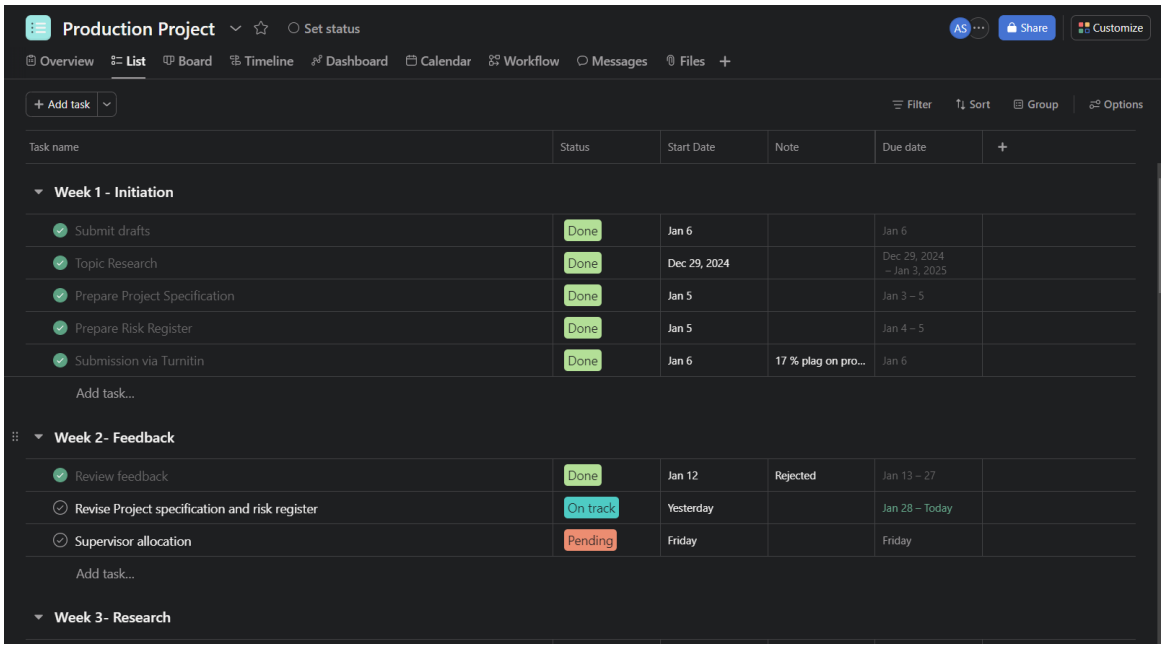
Research:

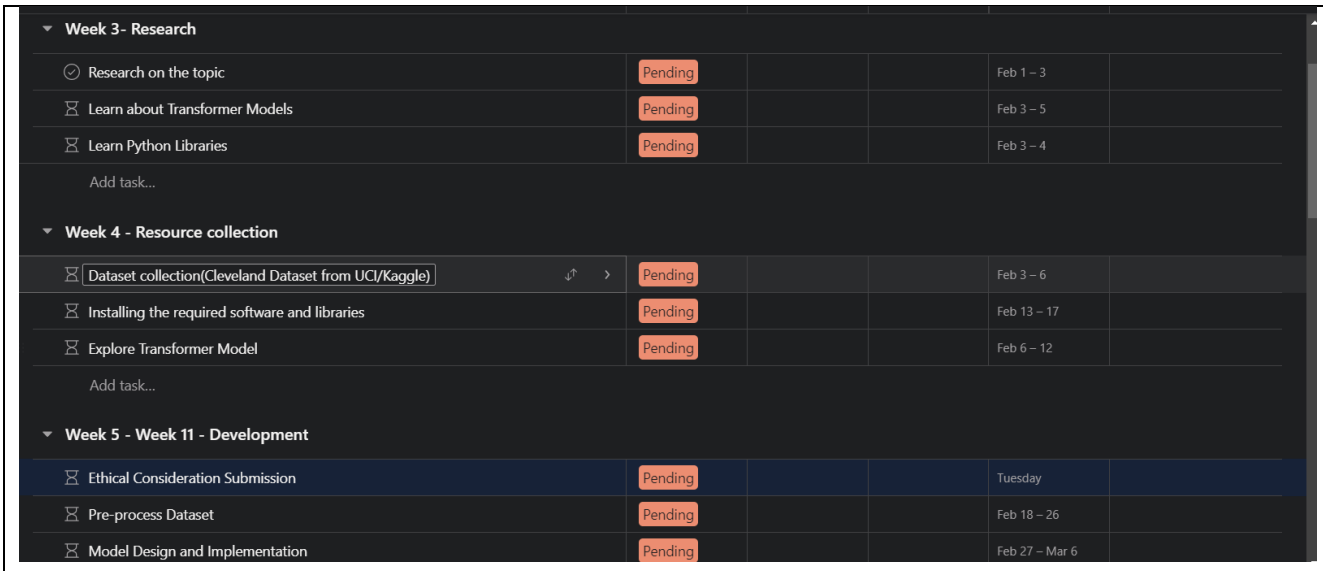
This project focuses on improving heart disease prediction using a self-attention-based transformer model, as proposed by Rahman et al. (2024). Unlike traditional models like CNNs and RNNs, which struggle with sequential data and interpretability, the transformer model leverages multi-head self-attention to capture complex patterns in patient health records, such as age, cholesterol, and blood pressure (Choi et al., 2017; Dutta et al., 2020).

The key difference lies in the model's ability to focus on relevant features and provide interpretable predictions, helping healthcare professionals understand decision-making processes (Rahman et al., 2024). Additionally, its parallelizable architecture ensures faster training compared to sequential models like RNNs (Shah et al., 2020). The project will pre-process the Cleveland dataset, implement the transformer model using PyTorch, and evaluate its performance against baseline models using metrics like accuracy and F1-score. By addressing the limitations of traditional methods, this project aims to deliver a robust and interpretable solution for early heart disease detection.

Project Planning & Methodology

Project Planning:



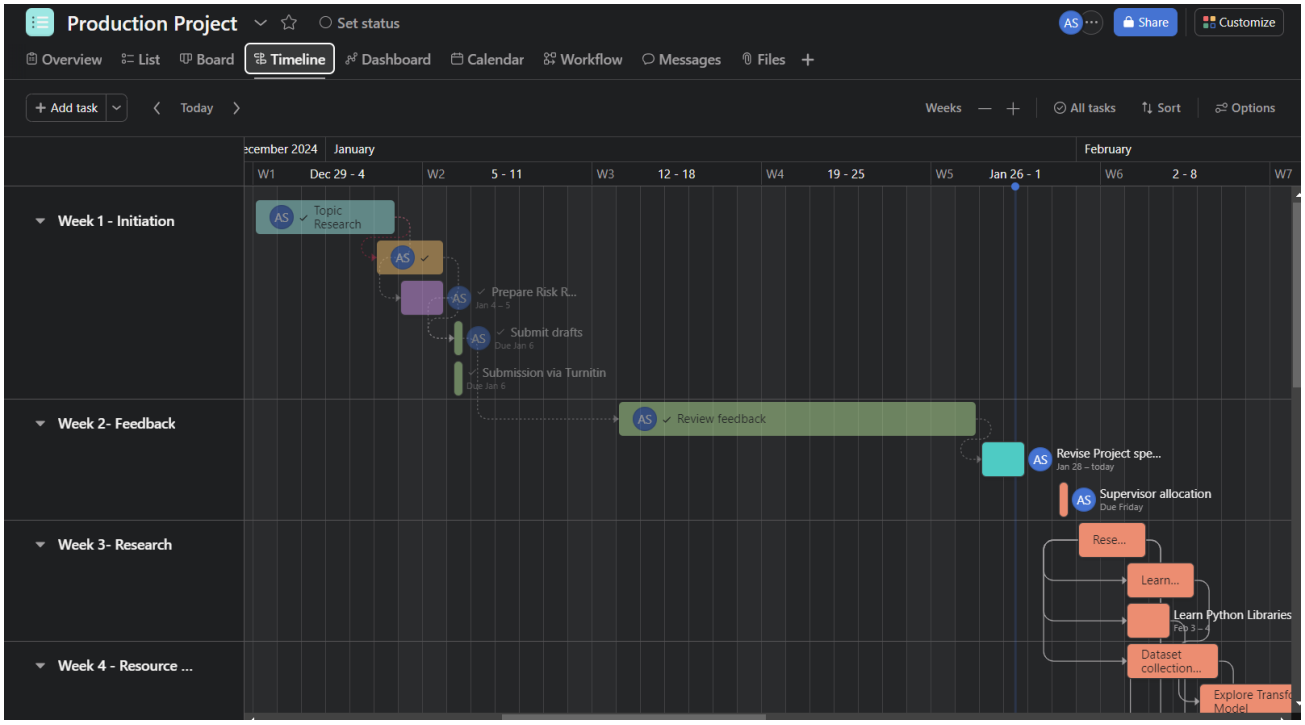


Week 3- Research					
Research on the topic	Pending			Feb 1 – 3	
Learn about Transformer Models	Pending			Feb 3 – 5	
Learn Python Libraries	Pending			Feb 3 – 4	
Add task...					
Week 4 - Resource collection					
Dataset collection(Cleveland Dataset from UCI/Kaggle)	Pending			Feb 3 – 6	
Installing the required software and libraries	Pending			Feb 13 – 17	
Explore Transformer Model	Pending			Feb 6 – 12	
Add task...					
Week 5 - Week 11 - Development					
Ethical Consideration Submission	Pending			Tuesday	
Pre-process Dataset	Pending			Feb 18 – 26	
Model Design and Implementation	Pending			Feb 27 – Mar 6	

Task name	Status	Start Date	Note	Due date	+
Model Design and Implementation	Pending			Feb 27 – Mar 6	
Model Training and Evaluation	Pending			Mar 7 – 14	
Build UI for system interaction	Pending			Mar 17 – 18	
Integrate the system to the UI	Pending			Mar 19 – 21	
Testing and debugging	Pending			Mar 24 – Apr 3	
Make Presentation slides	Pending			Apr 4	
Prepare for the presentation	Pending			Mar 22 – 23	
Add task...					
Week 12- Presentation					
Work in Progress Presentation	Pending			Mar 24	
Gather feedback and make changes to the plan	Pending			Mar 25 – 29	
Add task...					
Week 13- Refine System					
Optimize model and interface	Pending			Mar 31 – Apr 5	

Figure 1:To do task list

Gannt chart



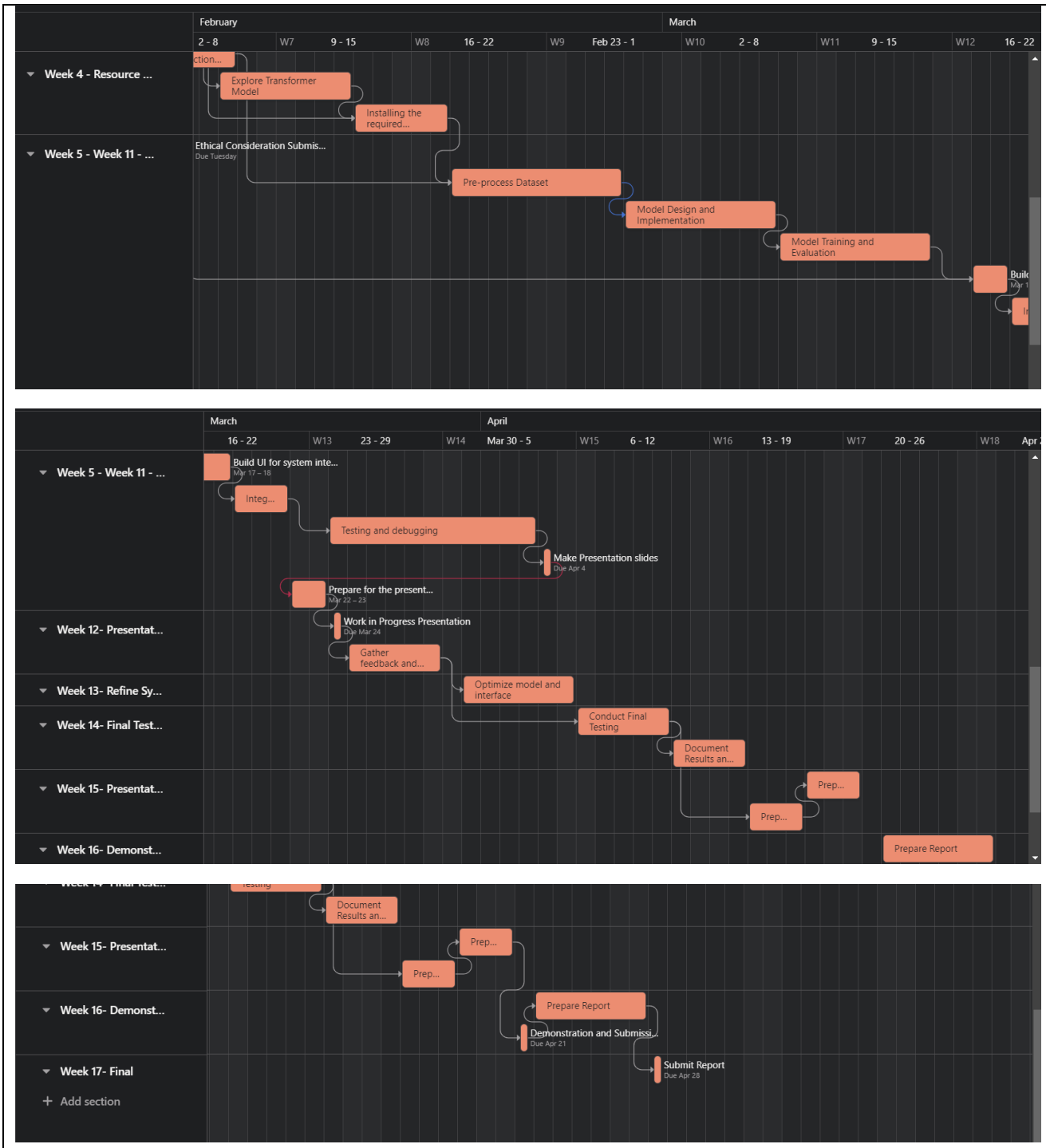


Figure 2: Gannt chart showing the task dependencies.

Timeline

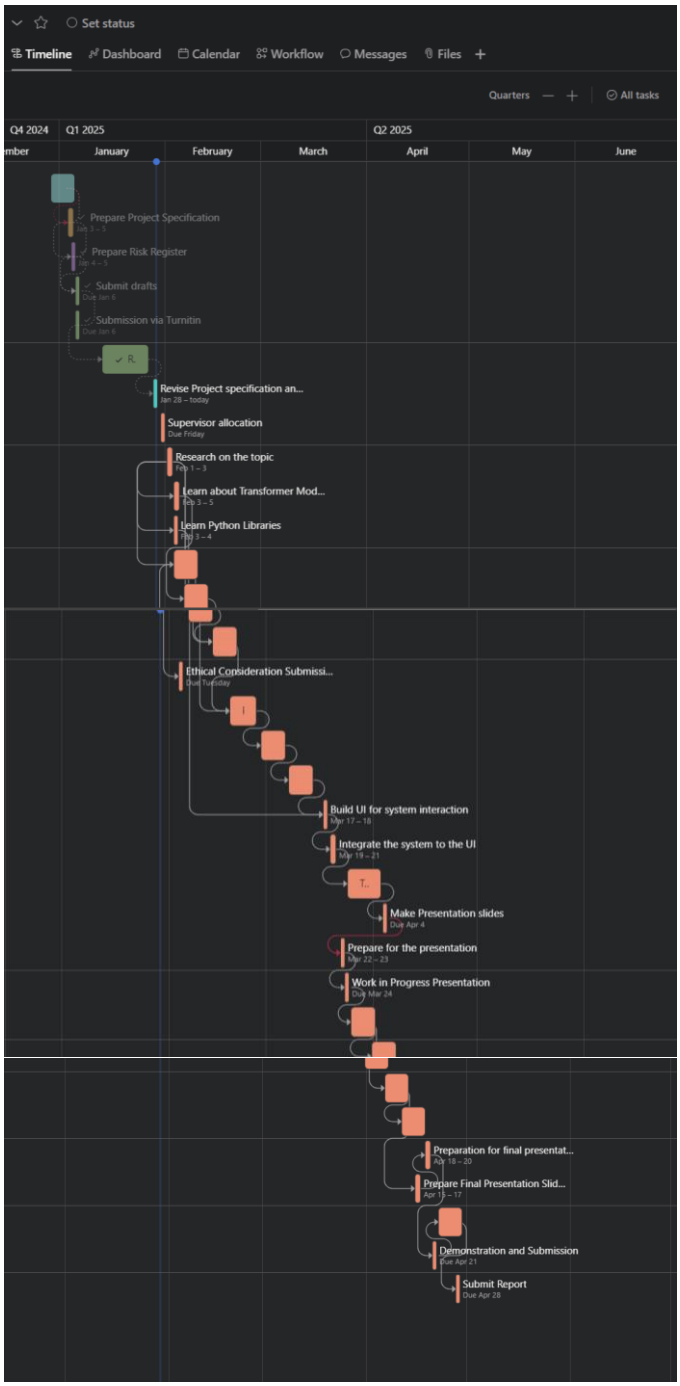


Figure 3: Timeline of the project.

Methodology:

The methodology focuses on achieving accurate predictions of cardiovascular disease risk using transformer model. This approach combines quality research and quantitative analysis to validate results. The methodology includes:

Finding the gap in existing method and understanding the role of transformer model.

Gather datasets publicly available (Kaggle, Cleveland Heart Disease dataset).

Implement the self-attention-based transformer model and evaluate model.

Design a website and test functionality.

Documentation and Reporting.

Evaluation:

The evaluation of this project aims to gain the performance and effectiveness of the proposed model which shall be measured using quantitative metrics such as accuracy, recall, AUC-ROC and F1-score. An analysis comparing the self-attention-based transformer model benefit over traditional dataset will be highlighted.

Product Specification:

Divided the product specification into functional and non-functional requirements by using MoSCow approach.

Functional Requirements	MoSCoW
Implement the self-attention-based transformer model	M
Dataset integration and pre-processing	M
Heart disease risk prediction	M
Model evaluation and performance metrics	M
User-friendly interface for input/output	S
Comparative analysis with baseline models	C
Real-time prediction capability	S
Advanced visualization of results	C
Integration with healthcare systems	W

Non- Functional Requirements	MoSCow
Developed using non-paid software tools	M
Rigorous testing and validation	M
Easy to use by the user	S
Fast response time for prediction	C
Compatibility across another platform	W

Resources

The hardware and software I require to complete my Project successfully:

Hardware

- Dell vostro 5502

Software

- Ms Word
- Ms Excel
- Chrome, Ms Edge
- Visual studio code
- Python – TensorFlow, NumPy, Pandas, PyTorch, Scikit-learn
- Laravel, php
- Bootstrap
- MYSQL / MongoDB
- OpenCV
- GitHub
- Jupyter Notebook
- Asana
- Kaggle
- Canva

Human Resource

Name	Role
Rohit Raj Pandey	Supervisor
Avishek Sah	Researcher/Programmer

I am working on my Project with the following people

Name: Avishek Sah	**Role:**
	Module Leader: Rohit Raj Pandey Supervisor

Initial Bibliography

Bibliography

Rahman, A.U. *et al.* (2024) *Enhancing heart disease prediction using a self-attention-based transformer model*. https://www.nature.com/articles/s41598-024-51184-7#citeas (Accessed: January 28, 2025).

Choi, E. *et al.* (2017) *Using recurrent neural network models for early detection of heart failure onset*. https://academic.oup.com/jamia/article/24/2/361/2631499 (Accessed: January 28, 2025).

Dutta, A. *et al.* (2020) 'An efficient convolutional neural network for coronary heart disease prediction,' *Expert Systems With Applications*, 159. https://www.sciencedirect.com/science/article/abs/pii/S0957417420302323?via%3Dihub.

Shah, D., Patel, S. and Bharti, S.K. (2020) 'Heart Disease Prediction using Machine Learning Techniques,' *SN Computer Science*, 1(6). https://doi.org/10.1007/s42979-020-00365-y.

Effective heart disease prediction using hybrid machine learning techniques (2019). https://ieeexplore.ieee.org/document/8740989 (Accessed: January 28, 2025).