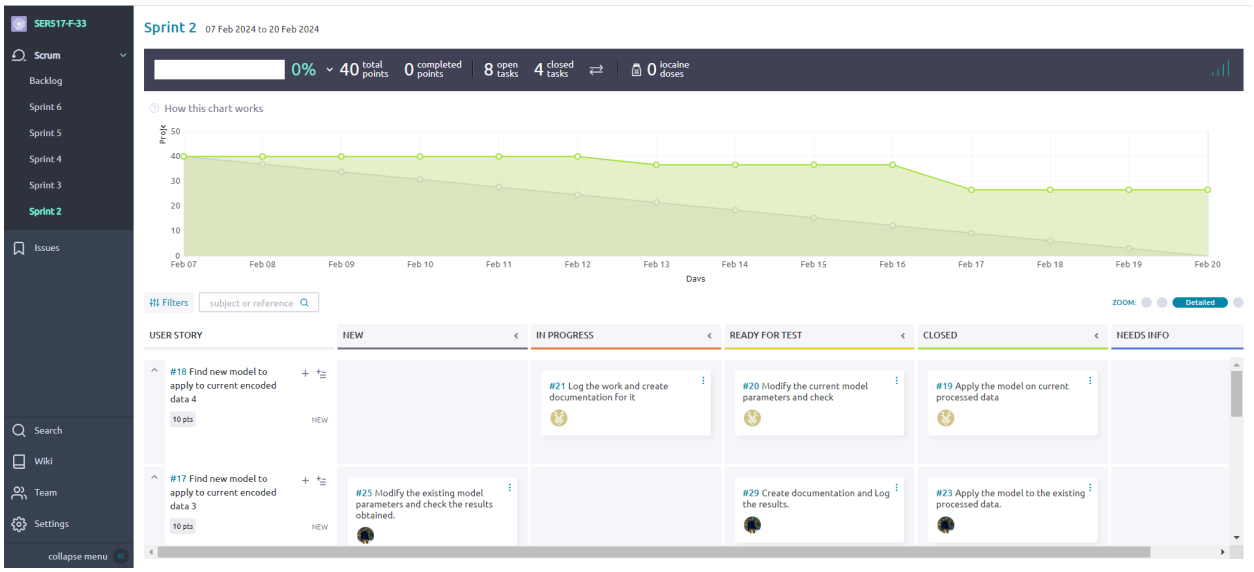
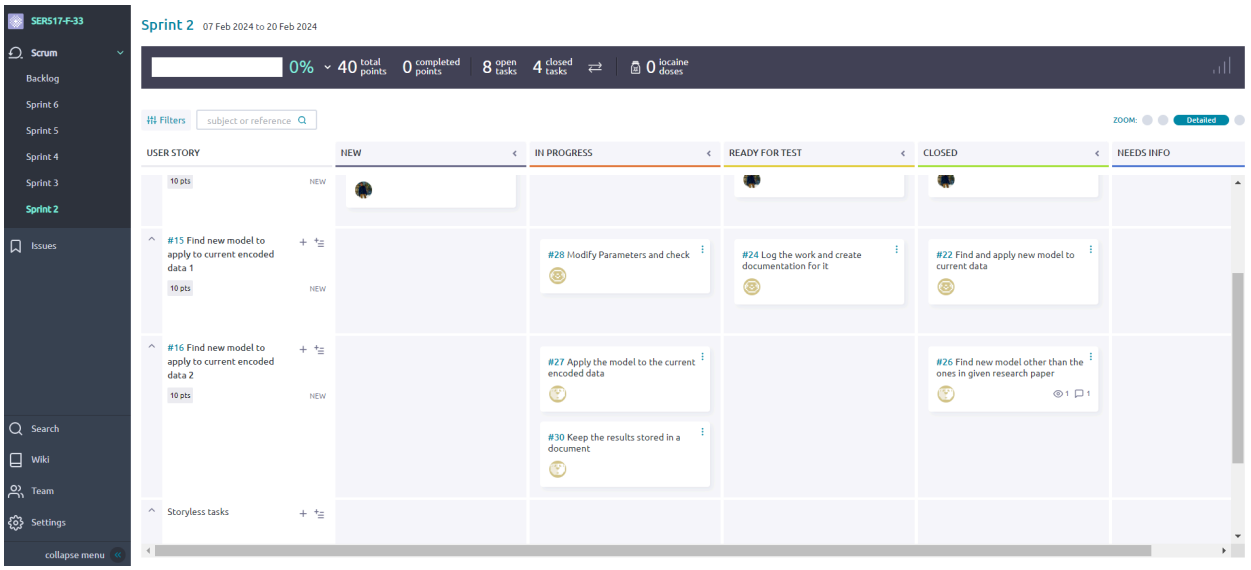
Report Group-33 Sprint-2

Members:
Arnav Raviraj
Adit Sandeep Virkar
Vinay Kantilal Chavhan
Shivanjay Vilas Wagh

1. **Taiga**
   Link: https://tree.taiga.io/project/araviraj8-ser517-f-33/backlog

   Burndown for Sprint 2 & Sprint 2:
   Duration of Sprint 2: 7th Feb 2024 - 20th Feb 2024





Sprint 2 Review:
As our Sprint 2 is still in progress, the team is confident that we can complete all the assigned user stories and tasks before the end of Sprint 2 (20th Feb 2024) with the current team velocity.

User Stories Created for Sprint 2:

#15 - Find the new model to apply to current encoded data 1.
As a team member involved in model evaluation, I want to establish robust validation strategies designed explicitly for encoded data analysis.

#16 - Find the new model to apply to current encoded data 2.
As a team member involved in model evaluation, I want to establish robust validation strategies designed explicitly for encoded data analysis.

#17 - Find the new model to apply to current encoded data 3.
As a team member involved in model interpretation, I want to develop techniques for interpreting and visualizing the output of encoded data models So that stakeholders can gain actionable insights and understand the rationale behind model predictions.

#18 - Find the new model to apply to current encoded data 4.
As a team member responsible for data preprocessing, I want to optimize the encoding process to minimize information loss and maximize data representation.

Tasks created for User Stories in Sprint 2:

Tasks for User Story: #15:
#22: Find and apply a new model to the current data.
#24: Log the work and create documentation for it.
#28: Modify parameters and check.

Tasks for User Story: #16:
#26: Find new model other than the ones in given research paper.
#27: Apply the new model to the current encoded data.
#30: Keep the results stored in a document.

Tasks for User Story: #17:
#23: Apply the model to the existing processed data.
#25: Modify the existing model parameters and check the results obtained.
#29: Create documentation and Log the results.

Tasks for User Story: #18:
#19: Apply the model on current processed data.
#20: Modify the current model parameters and check.
#21: Log the work and create documentation for it.

Good unit test:
We had our second sponsor meeting with our sponsor, Dr. Abdallah Moubayed. We shared our insights on different models we understood as individuals and as a team.
The user stories and tasks created were based on Dr. Abdallah's milestones for this project.

2. **Google Organized Project Plan**
   Google Drive Link:
   https://drive.google.com/drive/folders/17fXt0y-jq5i_r6ynpcHEsISKF3EHksBE

   Youtube Video Link :https://youtu.be/qRhCHf-muAM

   Meeting-1: 10 Feb, 2024
   We discussed finding new models; some of the suggestions made were for xgboost, lightGBM, AdaBoost, etc.; we will further work on it.

   Meeting 2: 16 Feb, 2024
   In this meeting, we talked about the models we worked on, the challenges we faced, etc. xgboost has an overfitting problem, which we discussed. We also discussed CatBoost and AdaBoost models, which we used on current encoded data.

   Github
   Link: https://github.com/Araviraj8/SER_517_F_33

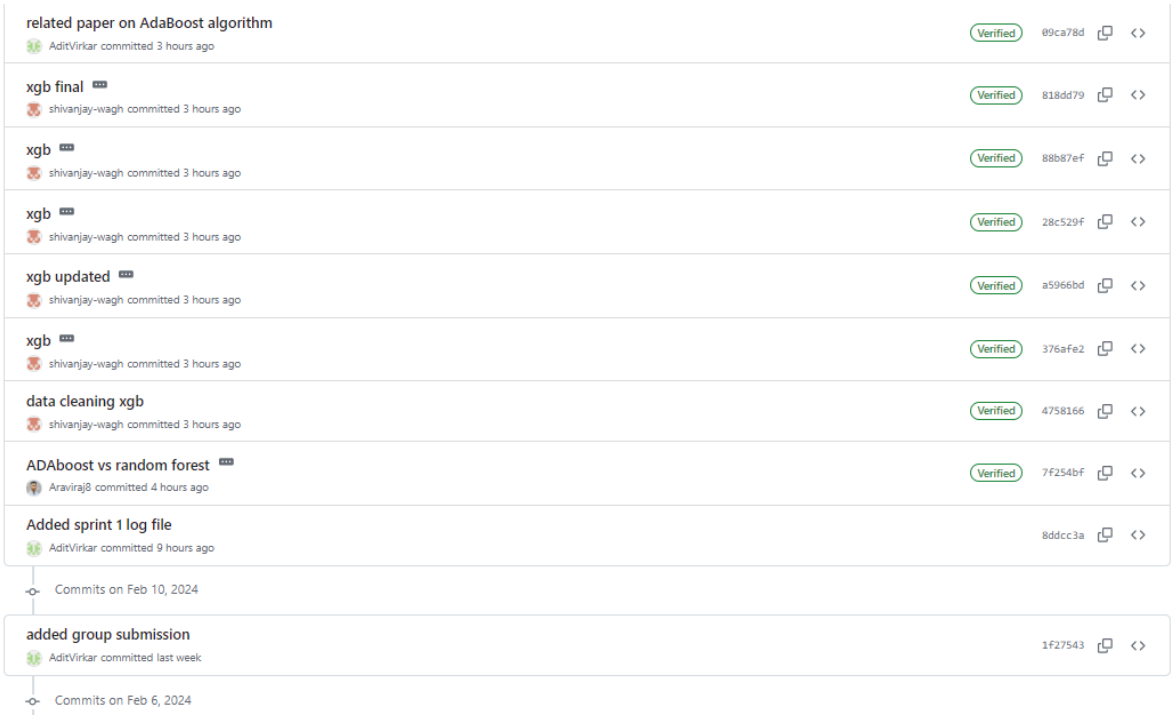
   GitHub commits for Sprint 2:

Commits

main ▾ | All users ▾ | All time ▾

Commits on Feb 16, 2024

final code AdaBoost Araviraj8 committed 1 minute ago		f303a68
added code for adaBoost Araviraj8 committed 2 minutes ago		a48c435
organized all files Araviraj8 committed 3 minutes ago		1f4ffe3
final data preprocessing AdaBoost AditVirkar committed 23 minutes ago		00c9f4c
initial data preprocessing AditVirkar committed 24 minutes ago		56ccd89
Added CatBoost algo used paper for reference VinayChavhan committed 4 hours ago		3242acb
CatBoost Final code VinayChavhan committed 4 hours ago		a4056a6
Added confusion matrix VinayChavhan committed 4 hours ago		e5cbe1a
Added Catboost code VinayChavhan committed 4 hours ago		62eb751
update the date cleaning catBoost VinayChavhan committed 4 hours ago		69dce55
Data clean file for CatBoost VinayChavhan committed 4 hours ago		ef87e44
related paper on AdaBoost algorithm AditVirkar committed 3 hours ago	Verified	09ca78d
xgb final shivanjay-wagh committed 3 hours ago	Verified	818dd79



3. **Sponsor Meeting**

   Sponsor Meeting 1:

   Date: Jan 23, Tuesday, 1:30 pm

   Our sponsor gave us a research paper entitled '5G-NIDD: A Comprehensive Network Intrusion Detection Dataset Generated over 5G Wireless Network'. They introduced the problem statement, described the dataset's generation process, and highlighted its importance. Our task was to investigate new machine-learning models. In the following meeting, we were assigned to conduct a literature review, familiarize ourselves with the project and the paper, and so forth. The overall objective was to enhance the models examined in the paper and conduct an analysis.

   Sponsor Meeting 2:

   Date: Feb 6, Tuesday, 1:30 pm

   We handed in our literature review and received feedback on it. During our discussion, we focused on the XGBoost model, known for its superior performance compared to Random Forest (RF) and Artificial Neural Networks (ANN) examined in the original

paper. Our suggestion involves leveraging XGBoost to enhance the model's performance across various metrics such as accuracy, inference time, and training time. Additionally, we delved into topics like adversarial machine learning and federated learning (FL), the latter being a distributed approach for deploying intrusion detection models. Furthermore, we addressed concerns regarding noise attacks and poisoning attacks during the conversation.

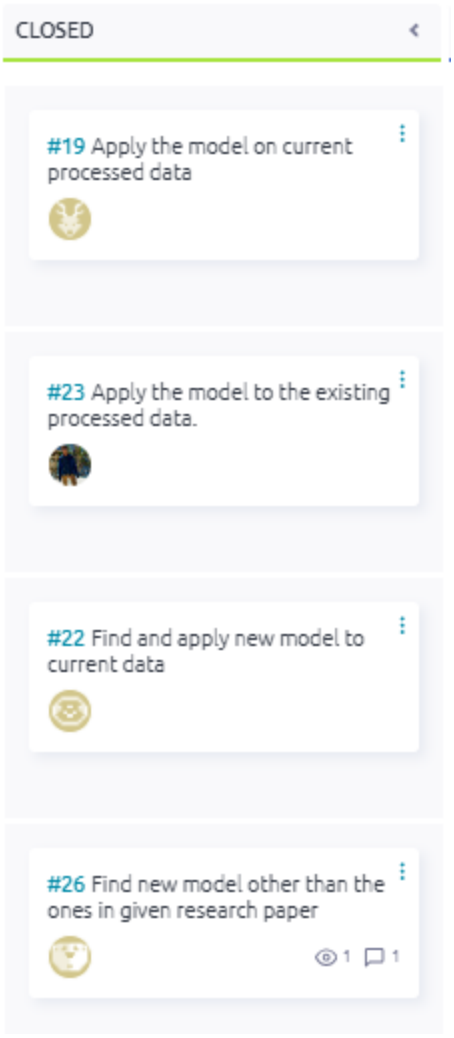
4. **Plan, Retro, Review, Sprint meeting**

- **Sprint Plan**:
  - ★ **Goals and Objectives**: The goal of Sprint 2 was to find a new model and apply it to currently processed data.
    User Stories: The user stories are as follows:
    - **Story 1:** Find a new model to apply to current encoded data (Assignee: Vinay Chavhan) As a team member involved in data analysis, I want to leverage a new model for analyzing encoded data So that we can improve the accuracy and effectiveness of our data analysis tasks.
    - **Story 2** (Assignee: Shivanjay Wagh) As a team member responsible for data preprocessing, I want to optimize the encoding process to minimize information loss and maximize data representation.
    - **Story 3** (Assignee: Adit Virkar) As a team member involved in model evaluation, I want to establish robust validation strategies designed explicitly for encoded data analysis.
    - **Story 4** (Assignee: Arnav Raviraj) As a team member involved in model interpretation, I want to develop techniques for interpreting and visualizing the output of encoded data models So that stakeholders can gain actionable insights and understand the rationale behind model predictions.
  - ★ **Priorities**: Team members worked on finding a new model with the same story points 10.
  - ★ **Sprint Backlog**: There were three types of tasks for each user story. One was to find a new model and apply it to current data; Second was to modify the model parameters and check the accuracy, prediction time, and training time. The third task was to keep a log of the work that everyone was doing and keep the documentation.
  - ★ **Resource Allocation**: Each team member was assigned an equal weightage of story points.

- **Sprint Review:**
  - ★ **What do you think is the value you created in this deliverable?**
    We created value by applying the newer model and analyzing it. This helped us understand what happens when we apply different models to our data, which will help us do the research for our next milestone.
  - ★ **Do you think you worked enough and did what was expected of you?**
    Yes, as discussed by our sponsor in our second meeting, we completed all the milestones, and our sponsor will verify them in the upcoming meeting.
  - ★ **Would you say you met the customers' expectations? Why, why not**
    Yes, we had a second meeting with our sponsor/customer and discussed the work, The sponsor was very happy and looking forward to the completion of our next milestone.

- **Sprint Retrospective**:
  - ★ **What Went right?:** Valuable work has been achieved in this sprint.
  - ★ **What went wrong?:** The Sprint is incomplete, and we are still working on completing tasks. The one thing that went wrong last time was that we did not dedicate a fixed time to keep meeting. Most meetings are going dynamic, which is going wrong in this sprint.
  - ★ **What Could Be Improved:** We could improve by dedicating ourselves to one task at a time, which will help us work faster.
  - ★ **How can we improve? :** We should be splitting up the work more and keeping track of it; that way, we will not have large tasks, and each task will be easy to tackle.
  - ★ **Action Items**: The main action item was to find the new model and apply it to the processed data, which was done by the team.
- **Review**:
  - ★ **Completed Work**: The Sprint is ongoing and will be completed on 20th Feb 2024. Most of the tasks were completed.

Closed tasks for Sprint 2:



★ **Demo**: We discussed finding a new model and how everyone applied it to the current data and showed it to the team.

★ **Stakeholder Feedback:** Stakeholders gave feedback in a second meeting to go ahead and work on future items, such as starting to apply models.

- **Sprint Meeting**:

★ **Sprint Goals**: The goal of Sprint 2 was to find a new model and apply it to currently processed data.

★ **Progress Updates:** The team progressed toward the sprint and completed the most valuable tasks. The Sprint is still ongoing.

★ **Impediments:** There were no such impediments; initially, the team faced a problem finding the new models, but with the team's efforts and support, we found models like XGBOost, CastBoost, and AdaBoost.