

Report Group-33 Sprint-3

Members:

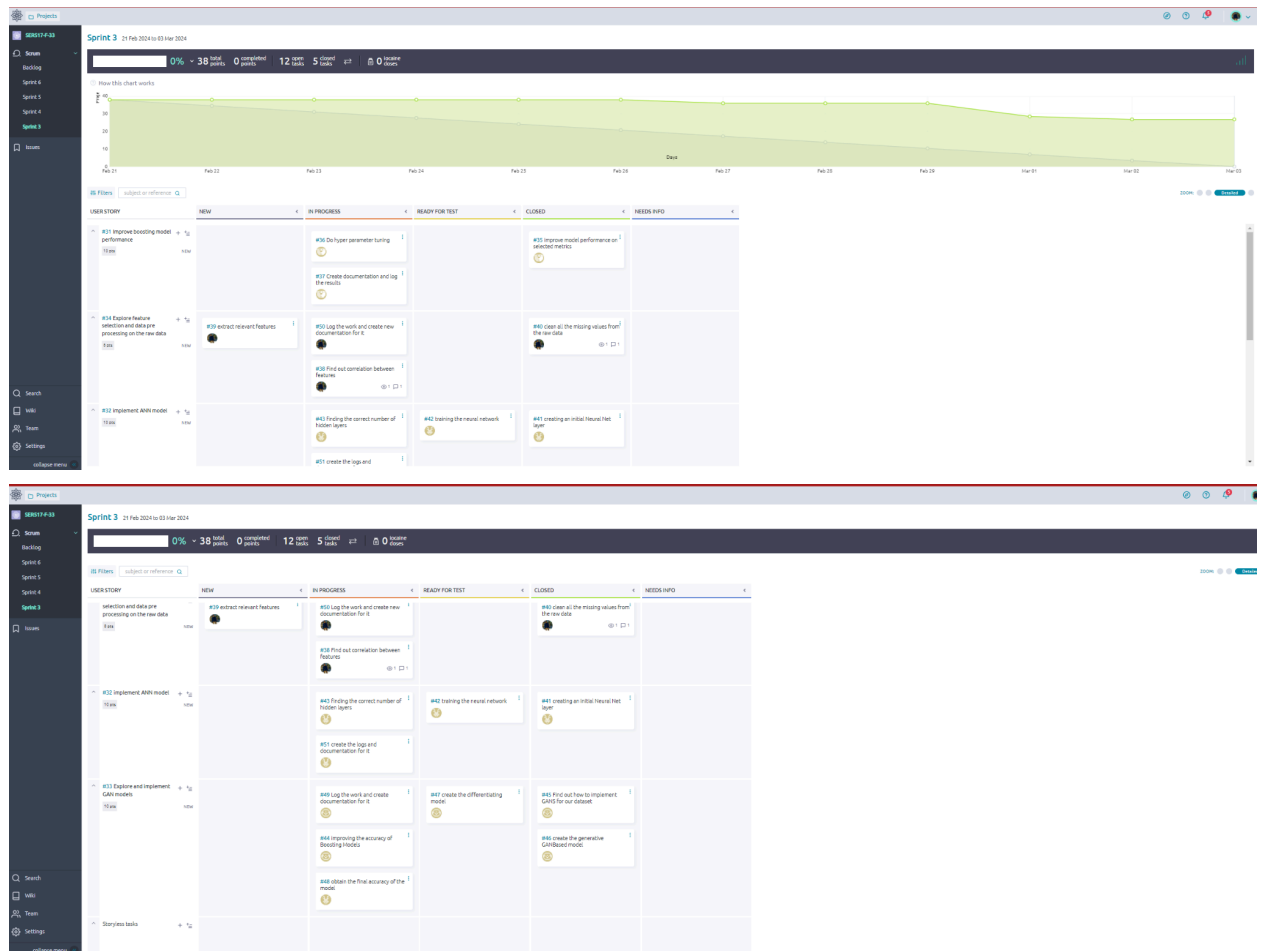
Arnav Raviraj
Adit Sandeep Virkar
Vinay Kantilal Chavhan
Shivanjay Vilas Wagh

1. Taiga

Link: <https://tree.taiga.io/project/araviraj8-ser517-f-33/backlog>

Burndown for Sprint 3:

Duration of Sprint 3: 21st Feb 2024 - 3rd Mar 2024



Sprint 3 Review:

As our Sprint 3 is still in progress, the team is confident that we can complete all the assigned user stories and tasks before the end of Sprint 3 (3 March 2024) with the current team velocity.

User Stories Created for Sprint 3:

#31 - Improve boosting model performance.

As a team member involved in model improvement, I want to improve the model performance to reduce interpretation time and increase accuracy.

#34 - Explore feature selection and data pre-processing on the raw data.

As a team member responsible for data preprocessing, I want to select relevant features and data-cleaning techniques to feed the model high-quality data.

#32 - Implement the ANN model.

As a team member involved in researching new models, I want to develop and explore ANN models To gain insights into the feasibility of neural networks for our dataset.

#33 - Explore and implement GAN models.

As a team member researching new models, I want to develop a GAN model so that we can train the model using adversarial techniques.

Tasks created for User Stories in Sprint 3:

Tasks for User Story #31:

#35: Improve model performance on selected metrics.

#36: Do hyperparameter tuning.

#37: Create documentation and log the results.

Tasks for User Story: #34:

#40: Clean all the missing values from the raw data.

#38: Find out the correlation between features.

#39: Extract relevant features.

#50: Log the work and create new documentation for it.

Tasks for User Story: #32:

#41: Creating an initial Neural Net layer.

#42: Training the neural network.

#43: Finding the correct number of hidden layers.

#51: Create the logs and documentation for it.

Tasks for User Story: #33:

#44: Improving the accuracy of Boosting Models.

#45: Find out how to implement GANS for our dataset.

#46: Create the generative GAN-based model.

#47: Create the differentiating model.

#48: Obtain the final accuracy of the model.

#49: Log the work and create documentation for it.

Good unit test:

We had our third sponsor meeting with our sponsor, Dr. Abdallah Moubayed. We shared different ML models we implemented as individuals and as a team. We also shared our analysis based on these models' accuracy, training time, and prediction time. 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_r6ynpcHEslSKF3EHksBE

Youtube Video Link:  SER 517 Capstone Team F 33 Sprint 3 Review 3

Meeting 1: 23 Feb 2024

During the meeting, we initiated Sprint 3 and formulated user stories along with their respective tasks. A productive discussion ensued regarding the assignment of story points and effort estimation for these stories. The overarching objective of the sprint we mapped out is to accomplish the implementation of at least one Artificial Neural Network (ANN) and GAN-based model.

Meeting 2: 26 Feb 2024

In the meeting, we addressed the blockers and progress of our ongoing work. A focal point of discussion revolved around applying Artificial Neural Networks (ANN) and GAN-based models to our processed data. We delved into strategies for modifying the parameters of the boosting algorithm, as well as explored the potential benefits of implementing various feature selection techniques to enhance the overall model performance.

Meeting 3: 28 Feb 2024

During the meeting, we engaged in a comprehensive discussion regarding our implemented models and delved into their architectures. We exchanged insights on the optimal selection of hidden layers tailored to specific model requirements. Our discourse extended to data preprocessing techniques and exploratory analysis, particularly examining feature correlation to discern the strongest relationships with the label column. Through the collaborative sharing of results and ideas, we fostered a deeper understanding of our models and the intricate dynamics within our dataset.

GitHub

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

GitHub commits for Sprint 3:

Commits

main

All usersAll time

Commits on Mar 1, 2024

added encoded data	AdiVikar committed now	Verified	439fa87		<>
added ann tensorflow	AdiVikar committed 31 minutes ago	Verified	5618f30		<>
Delete Ann_template.py	shivanjay-wagh committed 50 minutes ago	Verified	5ae812d		<>
research paper on ANN	AdiVikar committed 51 minutes ago	Verified	a05e6de		<>
simple ann class architecture.	shivanjay-wagh committed 52 minutes ago	Verified	2c86a1b		<>
simple ann class	shivanjay-wagh committed 1 hour ago	Verified	f397b78		<>
organized all the files	AdiVikar committed 1 hour ago		2b4e677		<>
Applied XGBoost classifier using GAN Based model	VinayChavhan committed 7 hours ago		f55cc2f		<>
Merge branch 'main' of https://github.com/Araviraj8/SER_517_F_33	VinayChavhan committed 7 hours ago		0d549ed		<>
added sprint-3 log	AdiVikar committed 5 hours ago	Verified	e0825e6		<>

find the correlation between features	Araviraj8 committed 6 hours ago	5f97dff		
added Spearman correlation	Araviraj8 committed 6 hours ago	e96f96e		
added Pearson correlation	Araviraj8 committed 6 hours ago	f244f7e		
model weights dict	shivanjay-wagh committed 6 hours ago	Verified 379be09		
train_ann ^{new}	shivanjay-wagh committed 6 hours ago	Verified 68f988f		
train_ann changes ^{new}	shivanjay-wagh committed 6 hours ago	Verified 48fcd7e		
ann training module ^{new}	shivanjay-wagh committed 6 hours ago	Verified ea7ca9f		
Merge branch 'main' of https://github.com/Araviraj8/SER_517_F_33	VinayChavhan committed 7 hours ago	fa5ff5b		
Updated GAN Based code	VinayChavhan committed 7 hours ago	a2f0ea2		
Added GanBased study paper	VinayChavhan committed 9 hours ago	5efdb88		
Merge branch 'main' of https://github.com/Araviraj8/SER_517_F_33	VinayChavhan committed 9 hours ago	2977412		
Updated GANBased code	VinayChavhan committed 9 hours ago	7d5cccc		
correlation	Araviraj8 committed 9 hours ago	fece5e2		

vinaychavhan committed 2 hours ago		
correlation	Araviraj8 committed 9 hours ago	fece5e2 <>
Commits on Feb 29, 2024		
GANBased code	VinayChavhan committed yesterday	08a50ed <>
ann 2 layer model	AditVirkar committed yesterday	Verified dc38a70 <>
ann_initial	shivanjay-wagh committed yesterday	Verified e1a2b07 <>
Added pre-processed data	Araviraj8 committed yesterday	022ff2e <>
Delete 5G-NIDD Dataset. directory	Araviraj8 committed yesterday	Verified eb31363 <>
Delete 5G-NIDD_Dataset	Araviraj8 committed yesterday	Verified bfddbc6 <>
Commits on Feb 27, 2024		
Create clean Data	Araviraj8 committed 3 days ago	Verified f7ca709 <>
Create 5G-NIDD_Dataset	Araviraj8 committed 3 days ago	Verified 8cc0520 <>
Commits on Feb 19, 2024		
added sprint-2 log	AditVirkar committed 2 weeks ago	Verified 40e284f <>
Commits on Feb 16, 2024		

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. 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.

Sponsor Meeting 3:

Date: Feb 20, Tuesday, 1:30 pm

During the meeting, we engaged in a comprehensive discussion with the Sponsor regarding the application of our model. We presented our analysis, focusing on the utilization of boosting models. The Sponsor suggested exploring the potential of GAN-based and ANN-based algorithms, prompting us to consider various classification techniques for the test set and to experiment with different data processing methods.

Furthermore, we discussed the possibility of applying alternative data processing techniques to the raw data, including various feature selection strategies, to enhance performance. The Sponsor guided us to prioritize fixing one model and implementing ANN and GAN-based models before our next meeting, scheduled after spring break.

4. Plan, Retro, Review, Sprint meeting

● Sprint Plan:

★ **Goals and Objectives:** The goal of Sprint 3 was to work on GANs, ANN, and Boosting models and apply them to currently processed data.

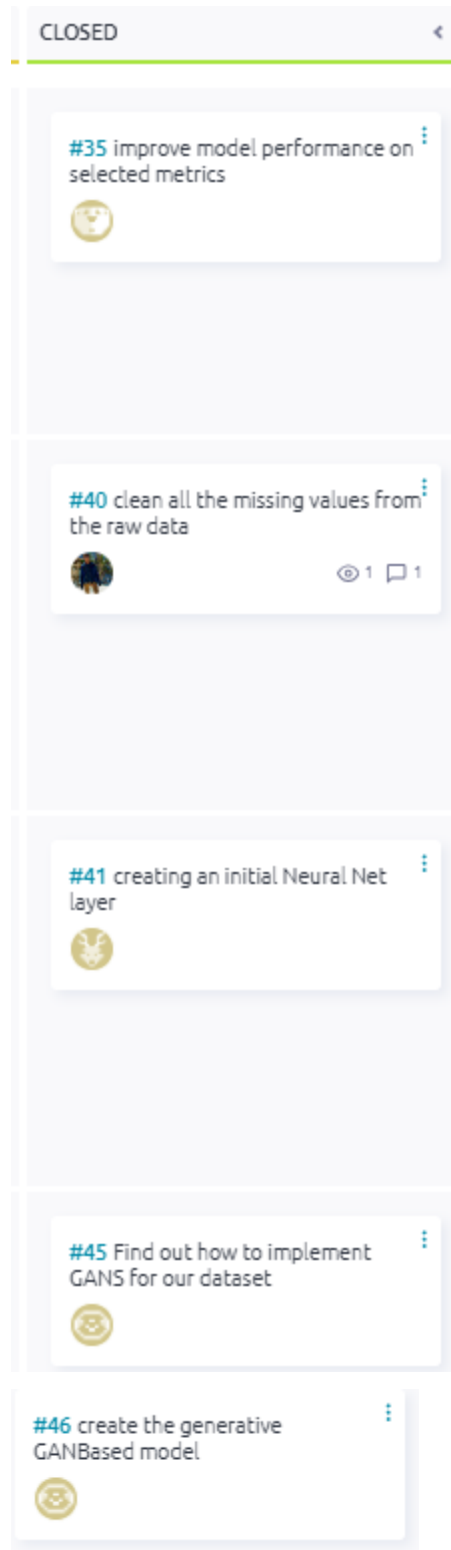
User Stories: The user stories are as follows:

- **Story 1** (Assignee: Adit): As a team member involved in model improvement, I want to improve the model performance to reduce interpretation time and increase accuracy.
- **Story 2** (Assignee: Arnav): As a team member responsible for data preprocessing, I want to select relevant features and data cleaning techniques to feed the model high-quality data.
- **Story 3** (Assignee: Shivanjay): As a team member involved in researching new models, I want to develop and explore ANN models So that we can gain insights into the feasibility of neural networks for our dataset
- **Story 4** (Assignee: Vinay): As a team member researching new models, I want to develop a GAN model to train the model using adversarial techniques. Stakeholders can gain actionable insights and understand the rationale behind model predictions.

★ **Priorities:** Team members worked on improving all three models, exploring each, and documenting the results.

- ★ **Sprint Backlog:** Each user story had two types of tasks. One was to implement an initial model architecture. The Second was to improve the model parameters and check the accuracy, prediction, and training time.
- ★ **Resource Allocation:** Each team member got an equal weightage of story points assigned to them.
- **Sprint Review:**
 - ★ **What do you think is the value you created in this deliverable?**
We now have three proper models implemented. Now, we can analyze each of them and improve them further. We can also compare them to each other on various critical parameters.
 - ★ **Do you think you worked enough and did what was expected?**
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 met with our sponsor/customer and discussed the work. The sponsor was satisfied with our work and looking forward to completing our next milestone.
- **Sprint Retrospective:**
 - ★ **What Went right?:** All models and their architecture were implemented.
 - ★ **What went wrong?:** Group coordination could have been better. Dependencies on other tasks meant there was idle wait time in between
 - ★ **What Could Be Improved:** We could improve our group coordination so that tasks sequentially dependent on each other are completed promptly.
 - ★ **How can we improve? :** Make dependent tasks more sequential, assign them in order of priority, and maintain that priority order.
 - ★ **Action Items:** Implement three different model architectures and improve their performance.
- **Review:**
 - ★ **Completed Work:** The Sprint is ongoing and will be completed on 3rd March 2024. Most of the tasks were completed.

Closed tasks for Sprint 3:



★ **Demo:** We discussed all three models and their implementation.

★ **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 3 was to implement all three model architectures and improve the accuracy by training the model.

★ **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 implementing the model and understanding its architecture.