Integrating Scrum Methodology into a Text Classification Project: Classifying Essays into LLM Generated or Human Generated Texts

Course:
Agile Project Management Methodologies
CRN-50046-202302

Group Members:
Dabas, Rashika
Dimaculangan, Roan Kathrina
Ip, Chak Leung Vincent
Suresh, Priyanka Yadhav

Table of Contents

I. Project Background and Objectives

Project Background

The rapid advancement and integration of Large Language Models (LLMs) such as GPT-3 into various areas have resulted in significant shifts, particularly in academic settings and the broader domain of information dissemination.

In education, the growing reliance on AI for academic tasks has emerged as a double-edged sword. While LLMs offer exceptional convenience in content creation, they simultaneously pose risks to the foundational pillars of educational values. Critical thinking, creativity, and original research are increasingly under threat as AI-generated content becomes more accessible and sophisticated. This trend, as discussed by Michael Perkins in his 2023 article in the Journal of University Teaching & Learning Practice, leads to a diminished student engagement with learning materials, potentially degrading the quality of learning and hindering the development of crucial academic skills. Moreover, the shift demands a redefinition of academic plagiarism in the AI era to establish new academic frameworks and policies.

Parallelly, in the domain of information dissemination, the impact of LLMs, underscored by the Center for Security and Emerging Technology, raises alarms in the digital age. The ability of LLMs like GPT-3 to produce highly articulate and persuasive content has significant implications for the spread of misinformation. The realistic nature of AI-generated content can be exploited in disinformation campaigns, as evidenced in the study "Truth, Lies, and Automation" by the Center for Security and Emerging Technology. This study revealed GPT-3's effectiveness in crafting narratives, such as climate change denial, and its proficiency in complex tasks like rewriting news articles to sway public opinion. These capabilities of LLMs underscore the potential risks in scenarios where misinformation could have profound social, political, or economic repercussions, necessitating robust countermeasures and strategies for content verification.

Objectives

The primary objective of this project is to develop a sophisticated AI-based system capable of accurately classifying essays as either generated by Large Language Models (LLMs) like GPT-3 or authored by humans. This initiative seeks to address the growing challenges posed by the integration of LLMs in academic settings and information dissemination, specifically:

- Safeguarding the integrity of academic work by distinguishing between AI-assisted content creation and human effort.
- Mitigating the spread of AI-generated misinformation by accurately identifying the source of texts in the broader context of media and information.

II. Project Scope

The scope of the project covers the following domains:

Data Collection

The project will focus on assembling a diverse collection of human-generated essays, covering a broad spectrum of emotional and cultural themes to ensure a comprehensive dataset. This includes essays characterized by:

- Joy, happiness, or excitement, to capture positive emotional expressions.
- Sadness or grief, to include narratives reflecting negative emotional states.
- Fear, to encompass essays that convey anxiety or apprehension.
- Historical or cultural topics, ensuring the inclusion of essays with rich cultural insights and backgrounds.
- Personal opinions on various subjects, to gather opinionated essays that reflect individual viewpoints.
- Texts containing slang words or idiomatic expressions, to cover informal and colloquial language use.

Cloud Data Access Setup

The project will establish a secure cloud storage solution for the dataset, ensuring data integrity and accessibility for processing and analysis. This will facilitate scalable storage and secure management of collected essays.

Data Preprocessing

The collected data will undergo thorough cleaning and preprocessing to eliminate noise and standardize formats across the dataset. This step is critical for preparing the essays for machine learning analysis, ensuring consistency and accuracy in model training.

Data Distribution

Visualization techniques will be employed to illustrate the data distribution between human-generated and LLM-generated essays. This will provide valuable insights into the diversity and balance of the dataset, facilitating informed decisions during the model design phase.

Model Design

The project will explore RNN and LSTM models enhanced with dropout layers among other techniques to prevent overfitting and ensure the model generalizes well to unseen data. Experimentation with different concepts, such as dropout, will be crucial in refining the model's ability to accurately classify texts without memorizing the training dataset.

Model Training and Validation

Models, including RNNs and LSTMs, will be trained using the preprocessed dataset, with a specific focus on validating their performance across a variety of texts. This phase is

crucial for assessing the models' effectiveness in distinguishing between different types of essays and their origin.

Evaluation

Model performance will be evaluated emphasizing personalization and mitigating overfitting. The iterative process based on feedback will guide subsequent refinements in model design and training methodologies, ensuring continuous improvement and accuracy in classification.

User Interface Development

The project will create multiple UI prototypes, with at least three enhanced versions developed for approval. This iterative approach to UI design aims to refine the user experience, ensuring an intuitive and efficient interface for interacting with the classification system.

III. Project Execution using Scrum Methodology Overview

The adoption of the Scrum methodology for our project, which aims at differentiating between essays generated by Large Language Models (LLMs) and those authored by humans, signifies a smart strategy to handle the project's development. The agility provided by Scrum, characterized by its systematic yet adaptable framework, is particularly suited for the dynamic environment of text classification.

The project demands a methodology capable of accommodating iterative development cycles and Scrum facilitates this through its emphasis on incremental progress, cross-functional collaboration, and regular feedback loops. By leveraging Scrum, we position the project to adaptively respond to emerging discoveries and refine our approach to text classification, ensuring that the methodologies and solutions remain advanced and relevant in the field of AI.

IV. Team Composition and Roles

Role	Resource Name	Tasks
Customer/ Product Owner	Priyanka Yadhav Suresh	Customer provides and discusses the high-level requirements of the product with the Product Owner
		Product Owner provides and discusses the Sprints 1 and 2 backlogs to the Scrum Master
		Product Owner checks the output/ implemented User Stories of the Scrum Team after 2 weeks of Sprint 1 and 2

		(either Accept or Reject) through Sprint Retrospective ceremony
Scrum Master	Rashika Dabas	Scum Master performs Sprint 1 and 2 Planning (User Story discussions and Story Points assignment) with the Scrum Team.
		Scrum Master performs the following with the Scrum Team during Sprint Planning: - discusses the Sprint 1 and 2 backlogs with the Scrum Team if doable for 2 weeks - trims the backlogs into User Stories and assigns User Story Points to each item
		Scrum Master uploads the User Stories for Sprint 1 and 2 in JIRA
		Scrum Master creates and tracks the progress of the Sprint 1 and 2 activities through JIRA and Burndown Chart
		Scrum Master performs Daily Scrum with the Scrum Team to discuss – (a) What has been accomplished (b) What is currently being implemented and (c) Bottlenecks or Issues of the Scrum Team
Data Scientist	Roan Kathrina Dimaculangan	Performs Sprint 1 and 2 tasks and updates the assigned User Stories in JIRA religiously
		Setups the CI/CD Pipeline in GitHub
		Performs review of the source code of the Machine Learning Engineer (Peer Review) prior pushing to final code to GitHub
		Performs QA to the output/ user stories implemented by the Machine Learning Engineer
Machine Learning Engineer	Chak Leung Vincent Ip	Performs Sprint 1 and 2 tasks and updates the assigned User Stories in JIRA religiously
		Performs review of the source code of the Data Scientist (Peer Review) prior pushing to final code to GitHub
		Performs QA to the output/ user stories implemented by the Data Scientist

Team Jira Profiles

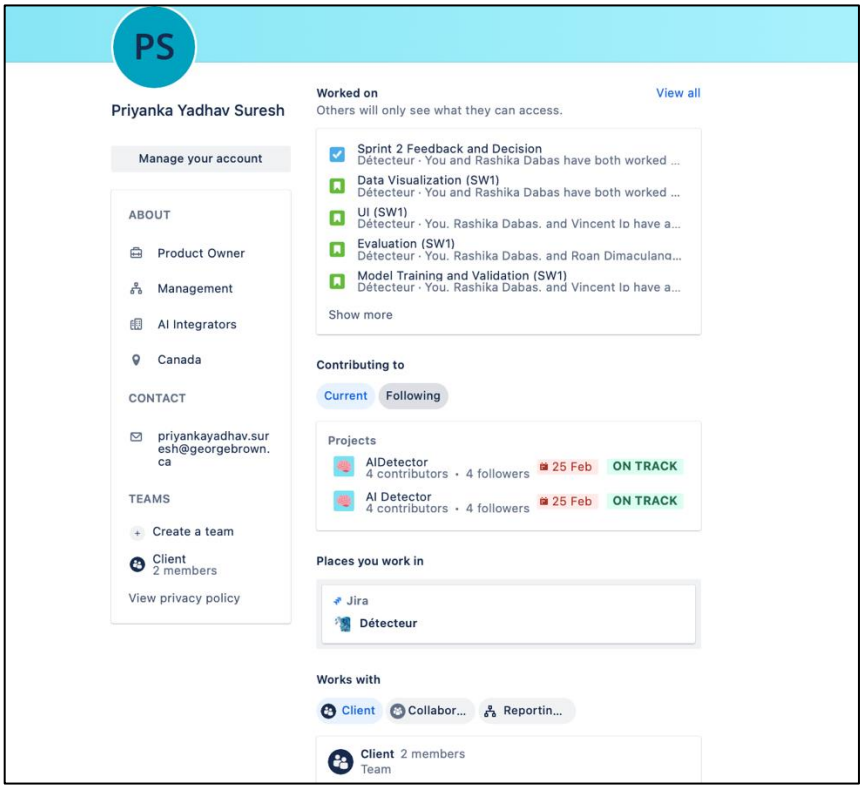


Figure 1 Product Owner Jira Profile

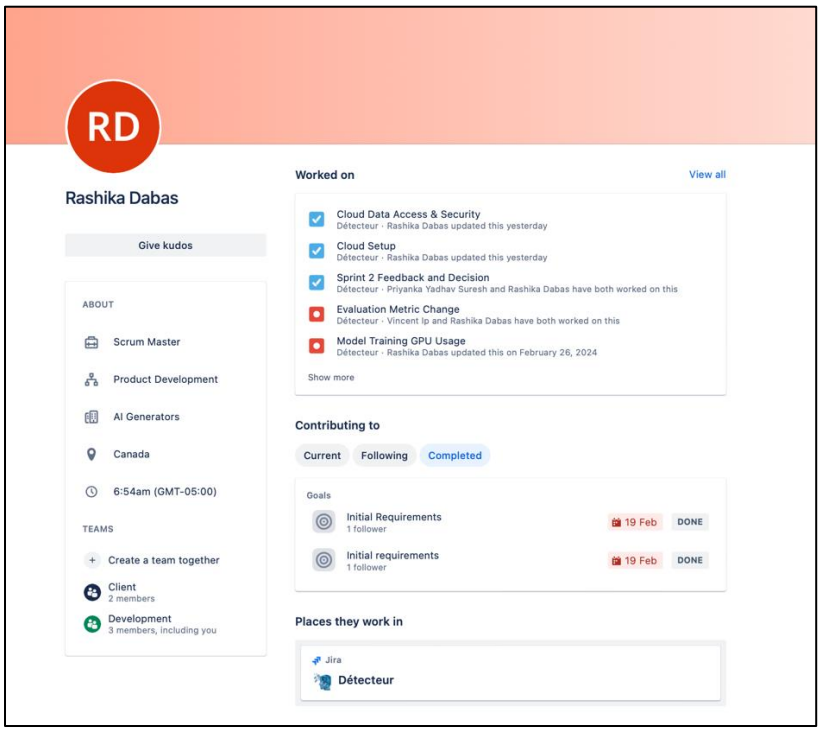


Figure 2 Scrum Master Jira Profile

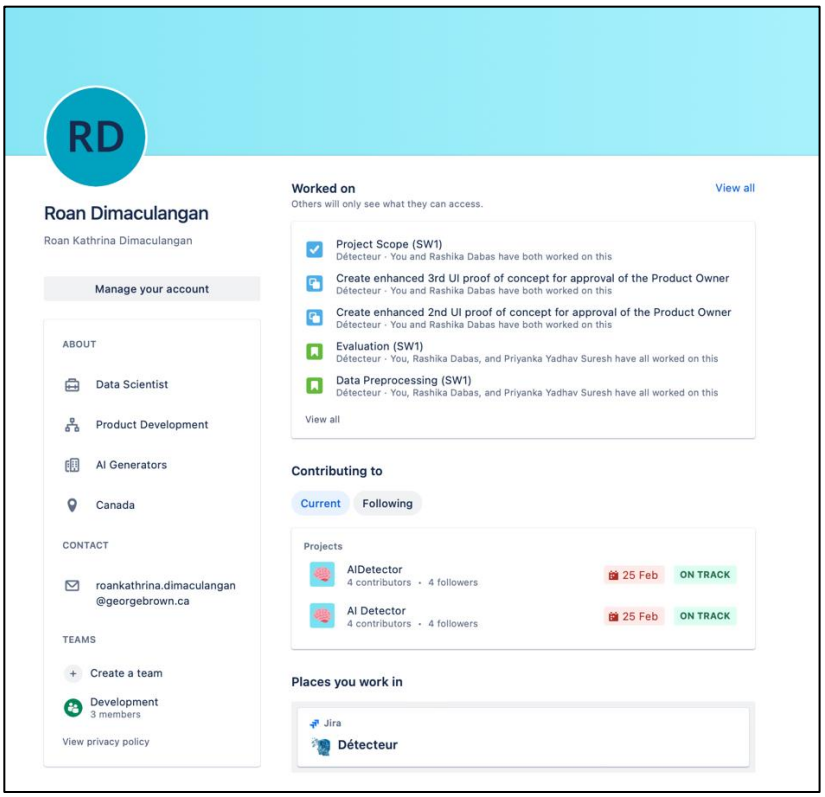


Figure 3 Data Scientist Jira Profile

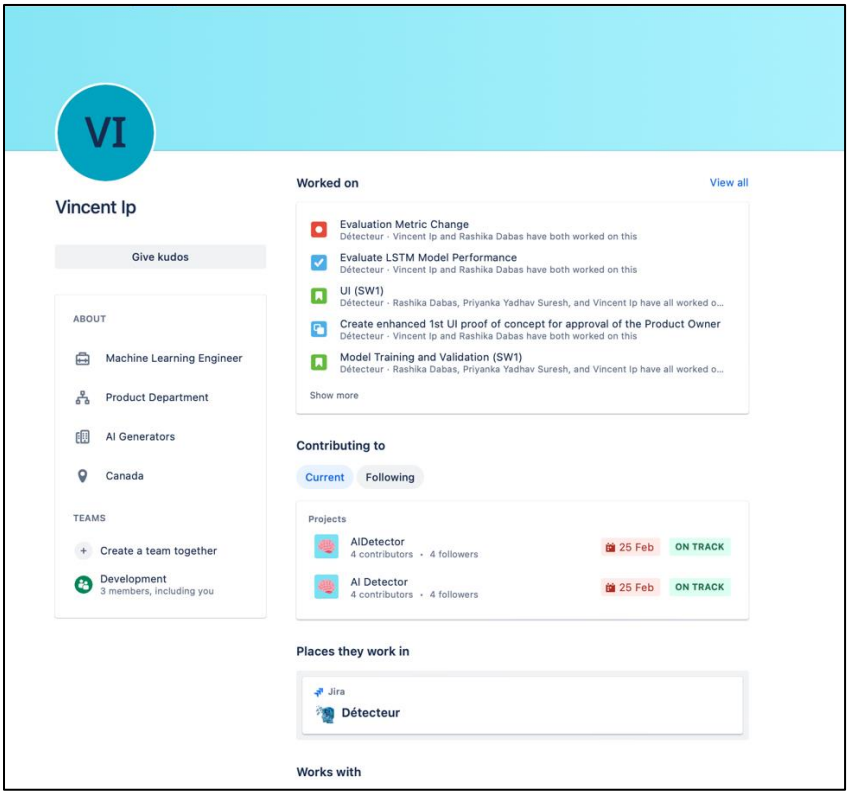


Figure 4 Machine Learning Engineer Jira Profile

V. Scrum Framework Components: Ceremonies, Backlog Management, and Visualization Tools

Jira Project (Détecteur) Link:

https://georgebrown-rashikadabas.atlassian.net/jira/software/projects/SCRUM/boards/1

A. Sprint 1: RNN Model Training and UI Enhancement Initiatives

Sprint Planning

- The objective of Sprint 1 was to do RNN Model Training and UI Enhancement Initiatives. The requirements were given by the Product Owner.
- Since the Product Owner requested for an RNN model to be initially developed, the Scrum Team started off with the same.

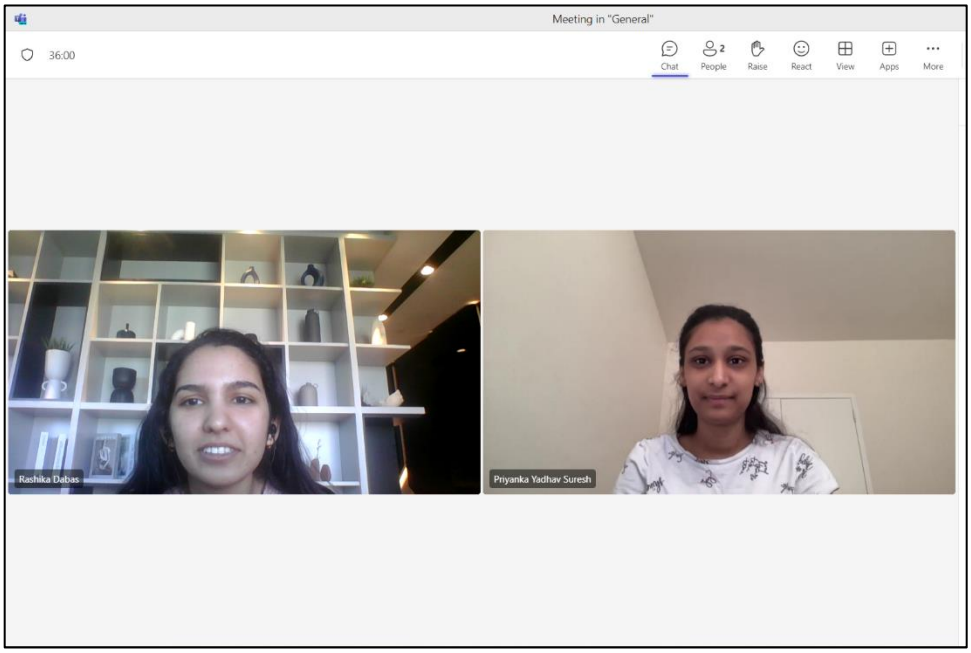


Figure 5 Product Owner and Scrum Master Meeting for Sprint 1 Requirements

Requirement Flow Chart

1. Objective
   To build a detector product that detects whether the given sentence is AI generated or manually entered by human.
2. Scope
   The channel that is subjected to this sentimental analysis are essays that are submitted by the students on their college portals.
3. Target Audience
   Students are the main target audience for this product, in the later stages we can make it generic depending on the need.
4. Dataset
   We can provide companies dataset for the final stage but as of now the team can use any sample data that is relevant to the need.
5. Model to implement.
   I want the team to build me a RNN model initially and I also require a UI.
6. Functional Requirement
   It should contain a textbox to copy paste the text to detect the originality of the text.

Figure 6 Sprint 1 Requirements from the Product Owner

Daily Scrum Meetings

- The Scrum Team performed two Daily Scrums to discuss (1) What has been completed (2) What is currently being performed and (3) Bottlenecks or issues, if any from below tasks:
  - Collect a diverse set of human-generated essays.
  - Clean and preprocess the data to remove noise and standardize formats.
  - LSTM models incorporating dropout layers
  - Train models using the preprocessed data and validate their performance.
  - Evaluate model performance in terms of personalization and overfitting. Iteration: Based on feedback, iterate on the model design and training process.
- Summary and key points from two selected Sprint 1 Daily Scrum are as follows:
  - The Data Scientist reported a possible delay for the User Story: Collection of a diverse set of human-generated essays. The acquisition of human-generated essays appeared to be more difficult than expected. As a mitigation plan for possible delay, the Scrum Team agreed to help in the acquisition if the Data Scientist cannot find the day before the target deadline.
  - The prototype of RNN model was built but overfitting of data was observed. Further work will be performed to resolve overfitting and improve prediction accuracy.

Figure 7 Sprint 1 Daily Scrum Meeting listed as Task in Jira

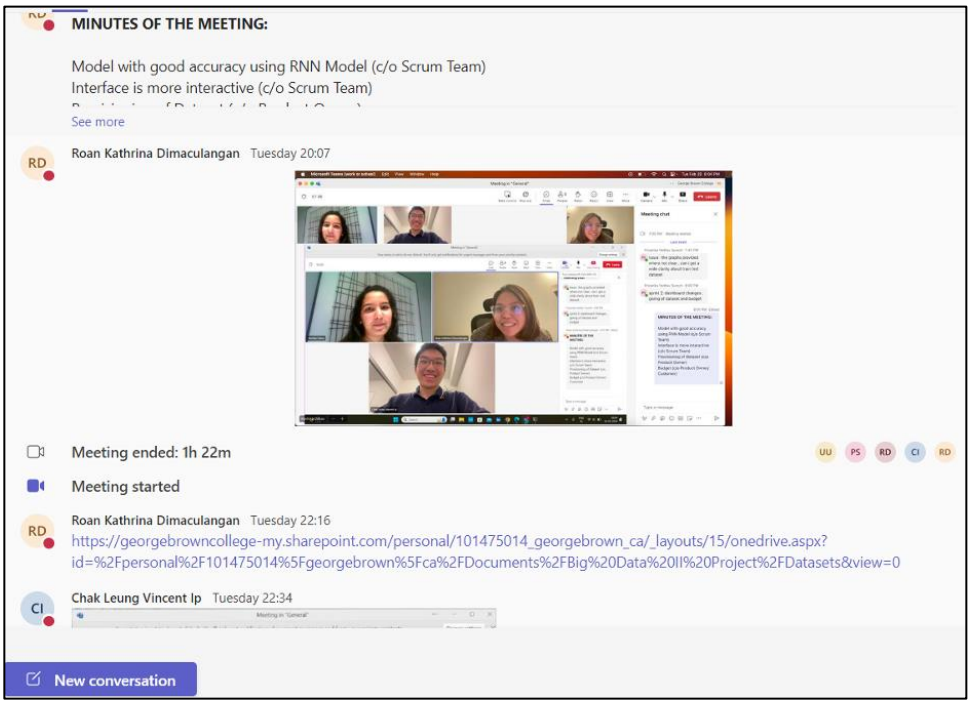


Figure 8 Scrum Team performing Daily Scrum during Sprint 1

Sprint Review

During the Sprint 1 Review, the Scrum Team presented the accomplished Sprint 1 User Stories to the Product Owner via Teams call. The Product Owner accepted all the accomplished User Stories and especially expressed her satisfaction with the RNN model accuracy.

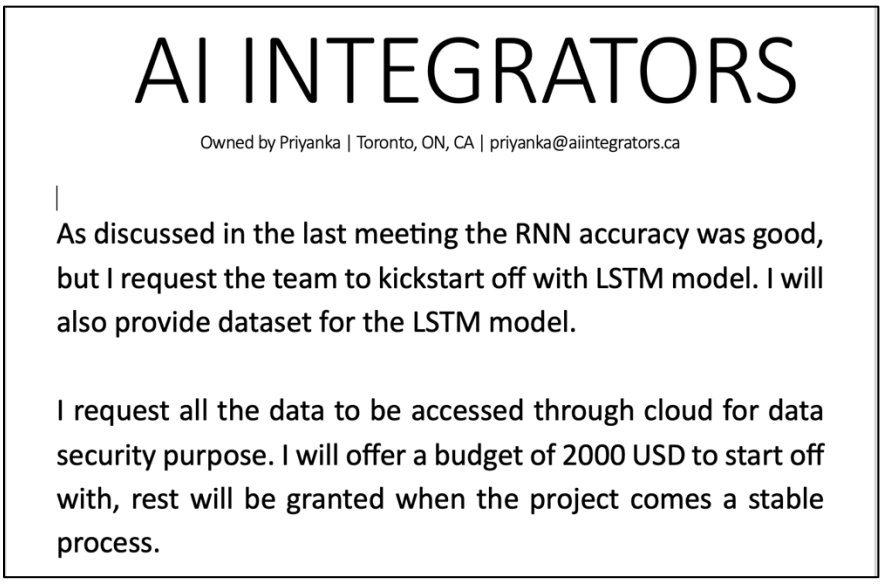


Figure 9 Product Owner's Feedback on Sprint 1 Review

Sprint Retrospective

During the Sprint 1 Retrospective, the team acknowledged several accomplishments and areas for improvement:

What worked during Sprint 1?

- **UI Visualization:** The team successfully developed an appealing and functional UI. This was a significant win, as the visual aspect of the project is working great and has met the intended design standards of the Product Owner.
- **Data Preprocessing/Exploratory Analysis**: The thoroughness of the data preprocessing and exploratory analysis activities laid a solid foundation for model training. This step was executed with commendable attention to detail, ensuring the data was well-prepared for the next stages.

What did not work during Sprint 1?

- **Model Training Time:** It was noted that the limited time allocated for model training affected the model's accuracy. The team concluded that the time constraints prevented the achievement of a higher accuracy level.
- **Dataset Acquisition Complexity:** The task of acquiring datasets was initially assigned too low of a complexity score. The time and effort required to find suitable datasets for model training were underestimated, which almost led to delay.

What can be improved for the next Sprints?

Based on the discussions and the challenges faced in Sprint 1, the team has planned several improvements for Sprint 2:

- **Parameter Optimization:** The team plans to experiment with different model parameters, such as dropout rates and the number of epochs, to enhance model accuracy and optimize the use of training resources.
- **Refinement of User Story Estimation:** The team will perform a more rigorous initial check of the user stories or requirements before assigning story points. This process aims to ensure that the estimates are more accurate and reflective of the actual effort and complexity involved.

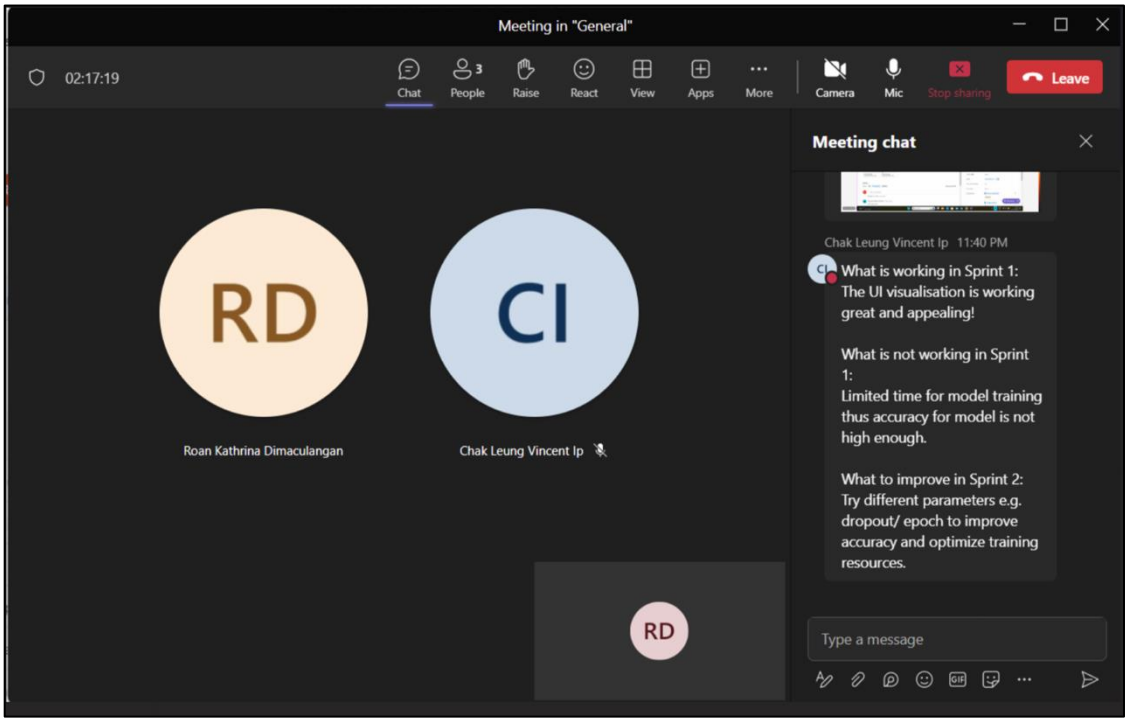


Figure 10 Scrum Team performing Sprint 1 Retrospective via Teams

Sprint Backlog Management

Below are the logged User Stories in Jira during Sprint 1. Note that User Stories were no longer in Jira after closing the Project.

Task	User Story	User Story Points	Sprint No. / Week No
Collect a diverse set of human-generated essays.	Collect human-generated essay that are joy, happiness or excitement focused	2	SW 1-1
	Collect human-generated essay that are sadness or grief focused		

	Collect human-generated essay that are fear focused		
	Collect human-generated essay that are historical or culturally inclined		
	Collect human-generated essay that discusses personal opinions on topics (opinionated essays)		
	Collect human-generated essay that are comprising of slang words or idiomatic expressions		
Clean and preprocess the data to remove noise and standardize formats.	Clean and preprocess the data to remove noise and standardize formats	1	SW 1-1
Show Data Distribution of human-generated and LLM-generated data in a visualization	Show Data Distribution of human-generated and LLM-generated data in a visualization	1	SW 1-1
LSTM models incorporating dropout layers	Experiment on incorporating different concepts to avoid model from memorizing (i.e. dropout, etc.)	3	SW 1-2
Train models using the preprocessed data and validate their performance.	Train using RNN model using the preprocessed data and validate their performance.	3	SW 1-2
Evaluate model performance in terms of personalization and overfitting. Iteration: Based on feedback, iterate on the model design and training process.	Evaluate model performance in terms of personalization and overfitting. Iteration: Based on feedback, iterate on the model design and training process.	3	SW 1-2
Create enhanced UI for the application	Create enhanced 1st UI proof of concept	2	SW 1-2

	for approval of the Product Owner		
	Create enhanced 2nd UI proof of concept for approval of the Product Owner		
	Create enhanced 3rd UI proof of concept for approval of the Product Owner		

Below are some of the issues that were handled during the first sprint.

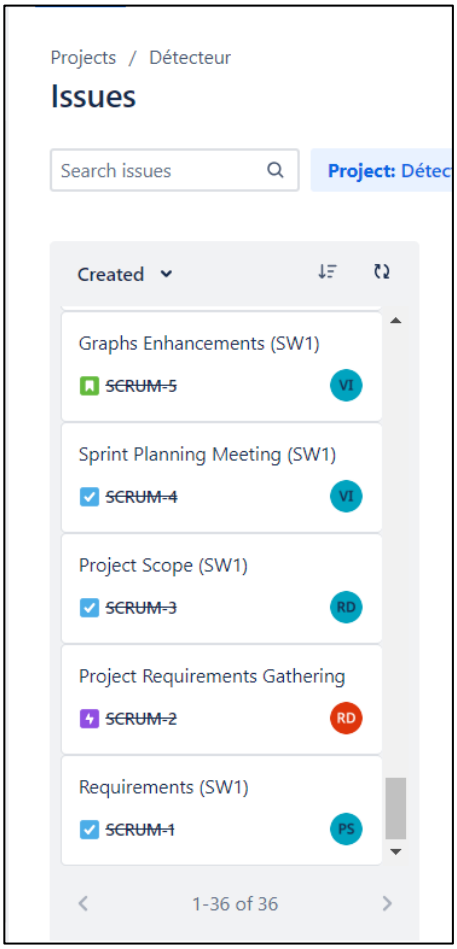


Figure 11 Sprint 1 Issues

Scale followed: 4 days = 2 weeks due to time constraints

Artifacts and Charts

The figure below shows that Scrum Team was able to complete all the logged User Stories for Sprint 1 within the estimated timeline:

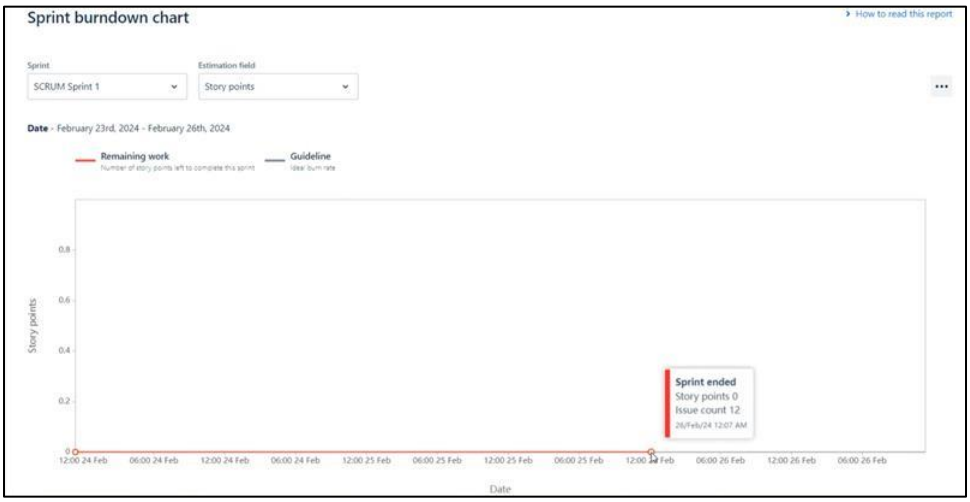


Figure 12 Sprint 1 Burndown Chart

B. Sprint 2: LSTM Model Training and Secure Cloud Storage Setup

Sprint Planning

- Objectives and goals for Sprint: To develop LSTM model and to relocate data input pipeline to cloud for data security purpose.

- Summary of Sprint 2 Meeting between the Product Owner and the Scrum Master: The product r thanked the contributions from Sprint 1. The predication accuracy of RNN model used in Sprint 1 was discussed. It was also requested by the Product Owner to explore the use of the LSTM model. As an additional requirement, all the data to be used in training, testing and validation should be uploaded and accessed through the cloud facilities for data security purposes. An additional budget of 2000 USD was granted for the development of Sprint 2.

AI

INTEGRATORS

Owned by Priyanka | Toronto, ON ,CA | priyanka@aiintegrators.ca

As discussed in the last meeting, I requested the team to kickstart with the LSTM model. After the final presentation, I am impressed by the team's effort in this application. The dataset usage is amazing.

I approve this project. I will also release the funds upon discussing with other stakeholders and will provide every technical help needed by the team. Good Job Team!!!

Figure 13 Sprint 2 Requirements provided by the Product Owner

- Summary of Sprint 2 Planning of the Scrum Team: Scrum Team organized a Sprint 2 Planning with the Development Team to discuss the requirements provided by the Product Owner. The team further break down the requirements into User Stories and assigned members (Data Scientist and Machine Learning Engineer) provided the User Story Points as they deemed appropriate for the complexity of the task.

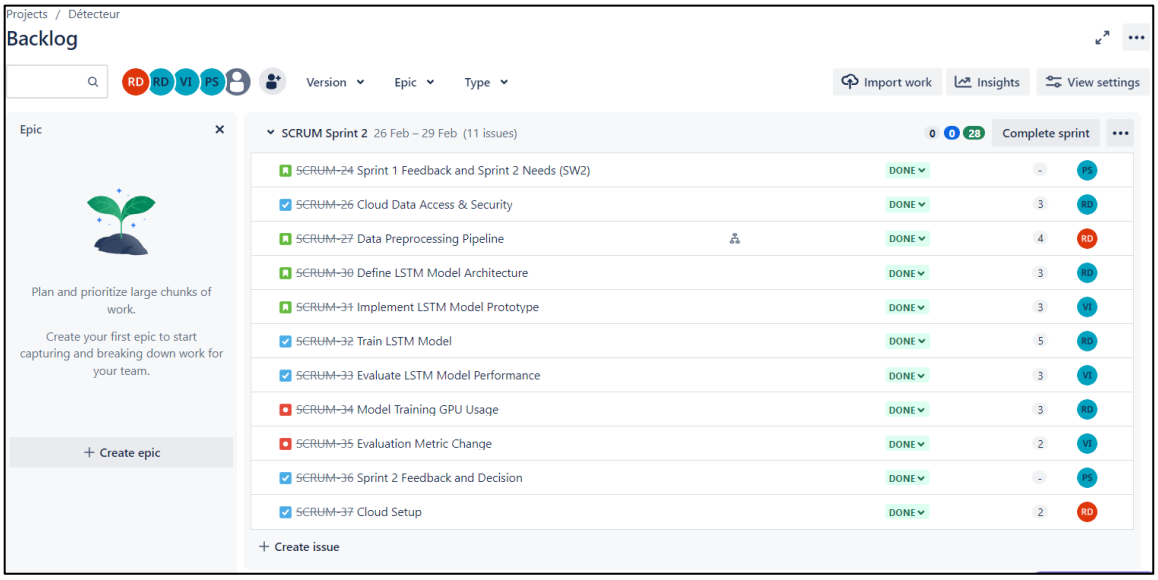


Figure 14 Sprint 2 Backlog Items (Tasks, Bugs and User Stories)

Daily Scrum Meetings

- The Scrum Team performed two Daily Scrums to discuss (1) What has been completed (2) What is currently being performed and (3) Bottlenecks or issues, if any from below tasks:
    - Issues raised from the Sprint 2 meeting
    - Development progress of the LSTM model
    - Set up of cloud infrastructure and dataset loading into the Cloud pipeline
- Summary and key points from two selected Sprint 2 Daily Scrum are as follows:
    - The prototype of LSTM model was built but overfitting of data was observed. Further work will be performed to resolve overfitting and improve prediction accuracy.
    - Cloud infrastructure was sourced, and the initial set-up of the data pipeline was initiated. Further work for completion of pipeline setup and loading of data was expected.

Sprint Review

- Summary of the Sprint 2 User Stories and updates presented by the Scrum Team to the Product Owner:
    - The LSTM model was developed seeing a 4% accuracy improvement from its counterpart RNN model.
    - Cloud platform was set up with data pipeline established. Data input was diverted from cloud platform into the model. All data outputs were also diverted to store in cloud server. All data is now contained in cloud platform only. Data security policy in place.
    - Only USD 2000 was used for Sprint 2, the sprint was completed within budget.
    - Sprint 2 was completed within the deadline set by Product Owner.
- Feedback received from Product Owner for the Sprint 2:
    - Both stakeholders presented significant satisfaction with the achievements and deliverables of Sprint 2.
    - Product Owner commented the deliverable is beyond her expectation, and the accuracy in prediction attained the need of her business model.

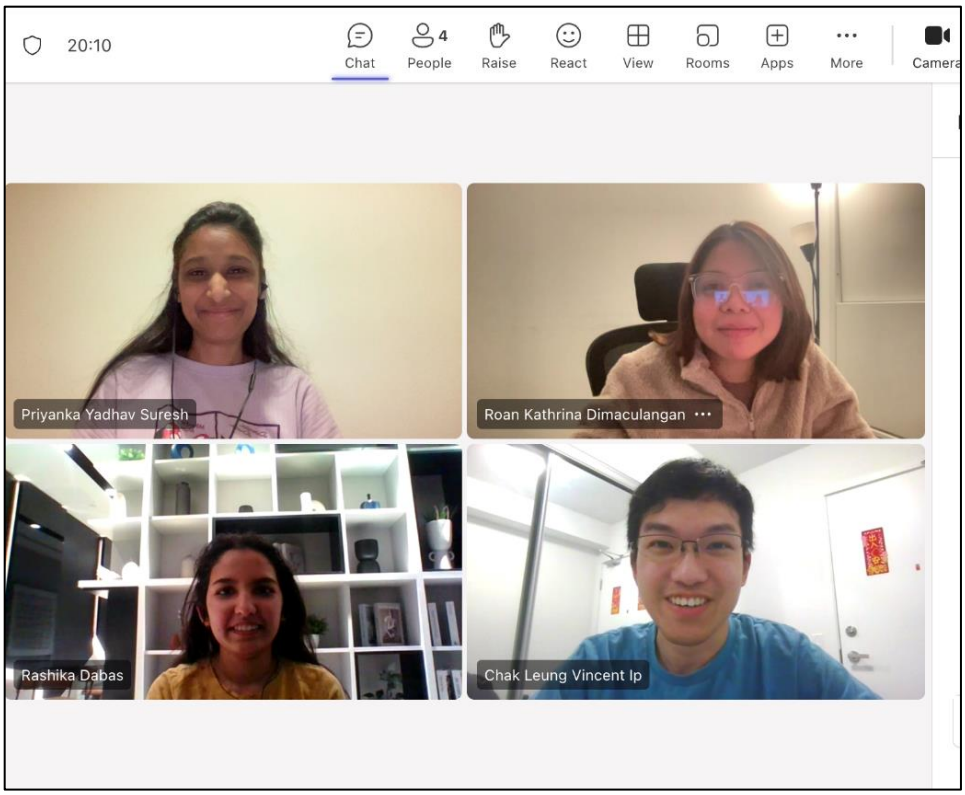


Figure 15 Sprint 2 Review via Teams (Scrum Team with the Product Owner)

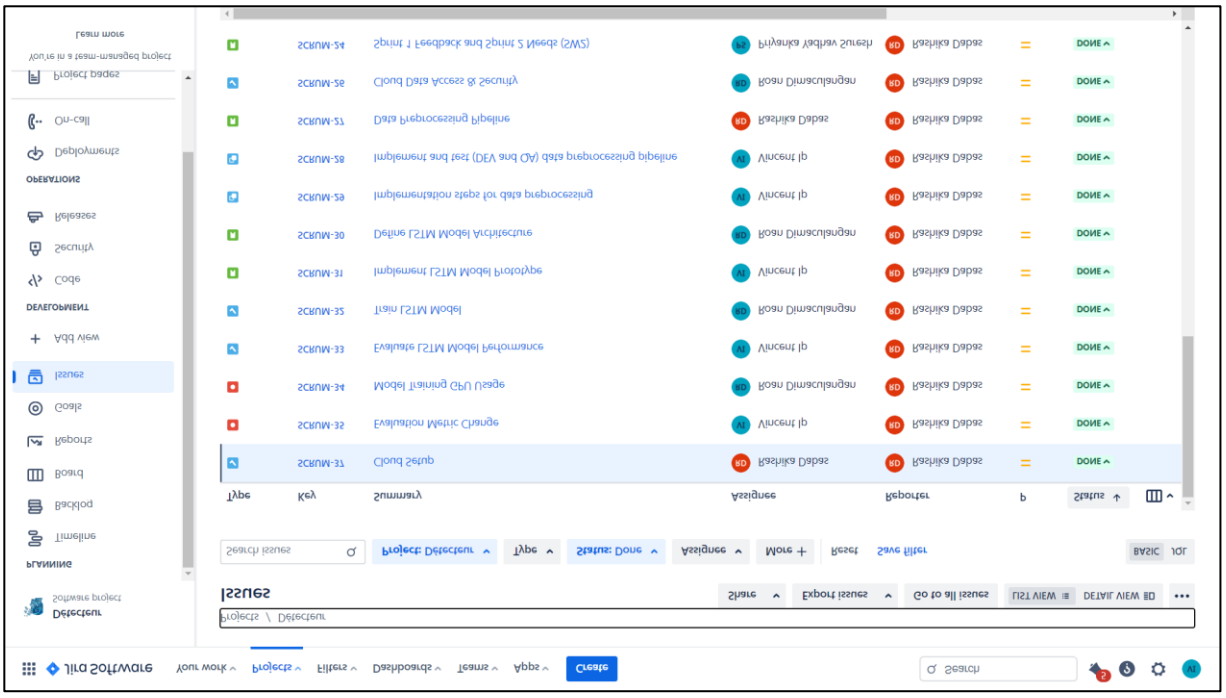


Figure 16 Screenshot of the completed Sprint 2 Backlog items

Sprint Retrospective

During the Sprint 2 Retrospective, the team acknowledged several accomplishments and areas for improvement:

What worked during Sprint 2?

- **Cloud as Data Storage**: The team's decision to use cloud storage for datasets was validated as a good practice, enhancing data security and accessibility.
- **Data Scientist's Flexibility**: The Data Scientist's adaptability in assuming tasks typically assigned to the Machine Learning Engineer was a significant advantage. This inter-role flexibility and comprehensive understanding of the project lifecycle allowed for smooth task reallocation in response to unforeseen circumstances.

What did not work during Sprint 2?

- **High User Story Points:** Assignments with high User Story Points (such as 5) were too ambitious, leading to a redistribution of tasks by the Scrum Master to ensure the completion of all committed Sprint 2 User Stories.

What can be improved for the next Sprints?

Based on the discussions and the challenges faced in Sprint 2, the team has planned several improvements for Sprint 3:

- **Refinement of User Story Points:** There is a need to reassess how User Story Points are assigned to ensure they accurately reflect the complexity and effort required. This will help prevent bottlenecks and overcommitment.
- **Task Breakdown:** Complex User Stories, such as those involving LSTM Model Training, should be broken down into smaller, more manageable tasks. This approach will improve task completion rates and enhance the quality of work within the two-week sprint cycle.

Sprint Backlog Management

Issues discussed during sprint 2 are as follows:

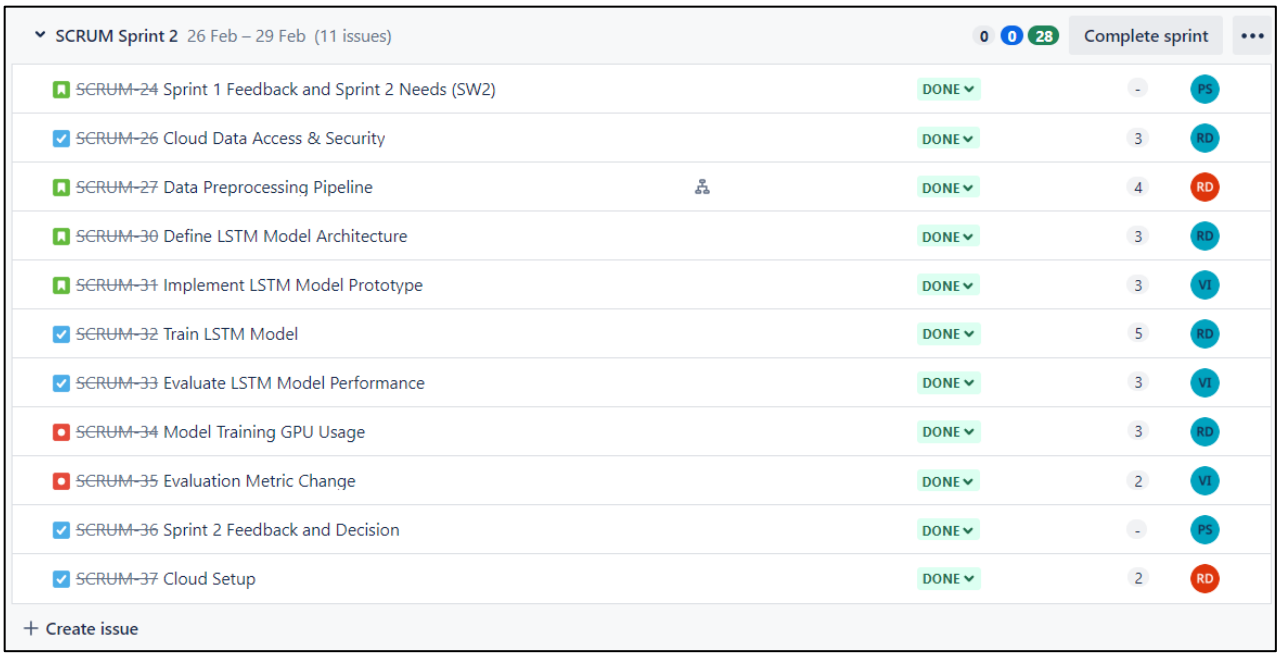


Figure 17 Sprint 2 Backlog

Scale followed: 4 days = 2 weeks due to time constraints

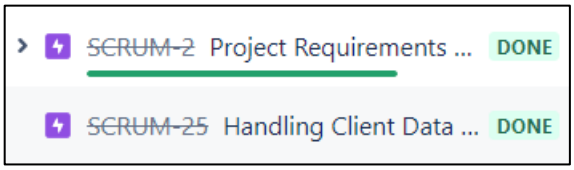
Artifacts and Charts

The figure below shows that Scrum Team was able to complete all the logged User Stories for Sprint 2 within the estimated timeline:

Figure 18 Burndown Chart of Sprint 2

VI. Analysis

We followed the below epics for our Project:



Also, to make the project goal-oriented and meet specific objectives, clear goals and subgoals were defined via Atlas platform of Atlassian.

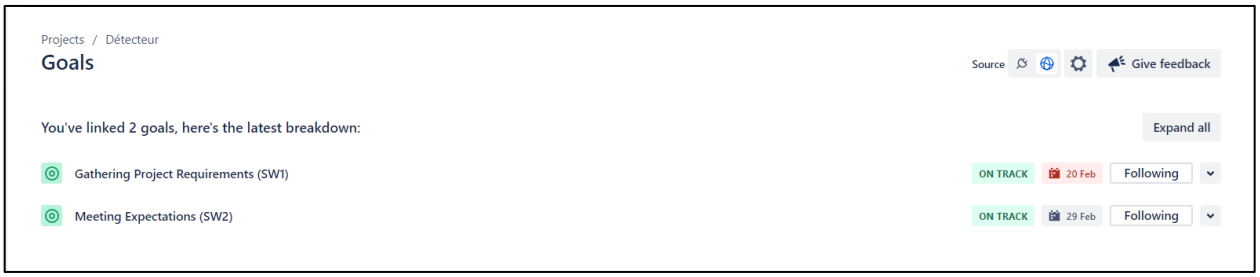


Figure 19 Goals

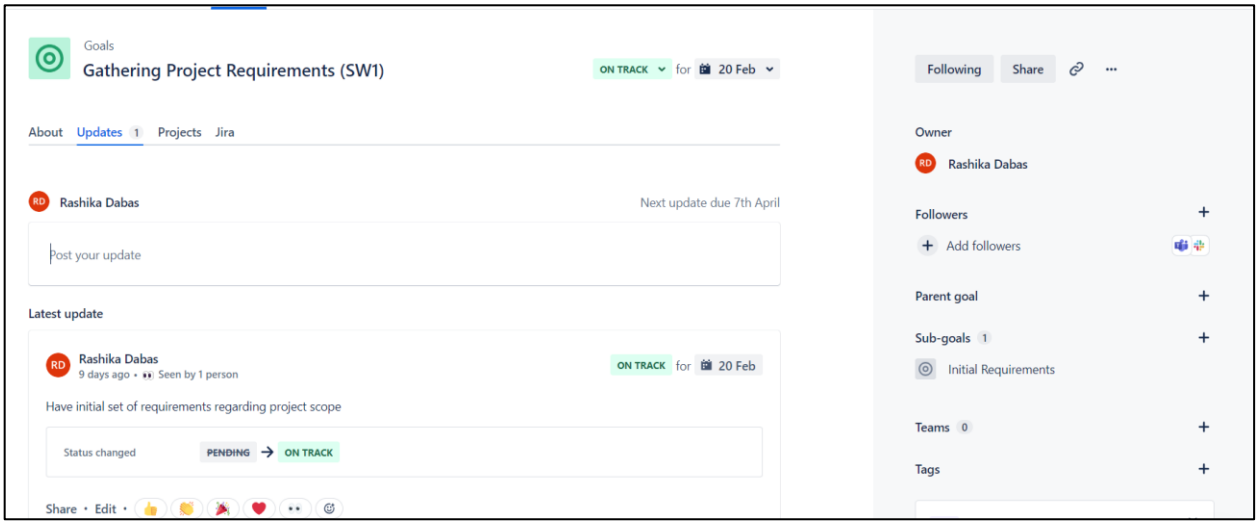


Figure 20 Goal 1

Figure 21 Goal 2

Following are the additional reports based on our Sprints 1 and 2:

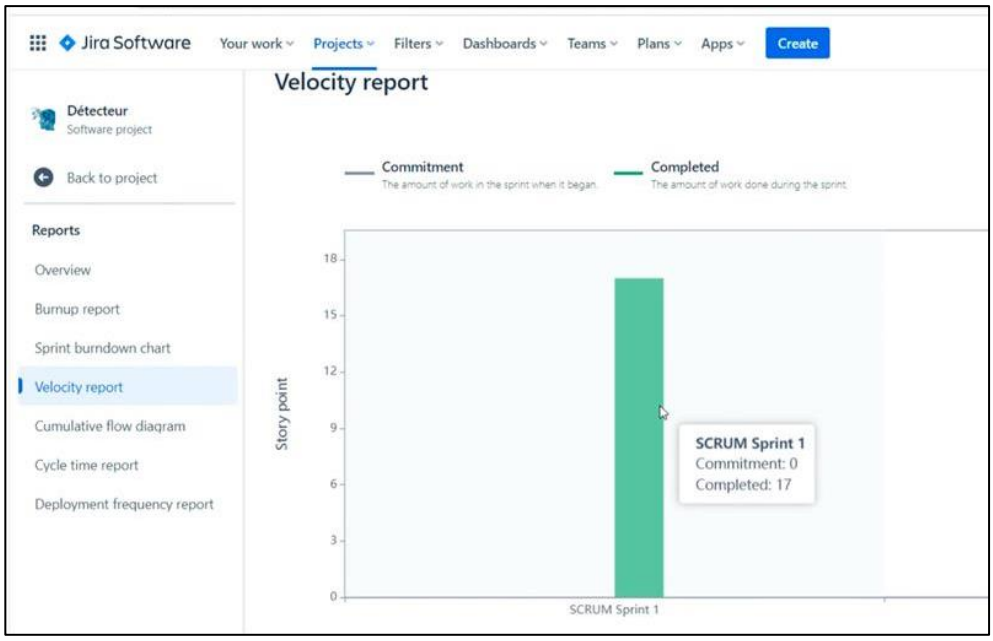


Figure 22 Sprint 1 Velocity Report

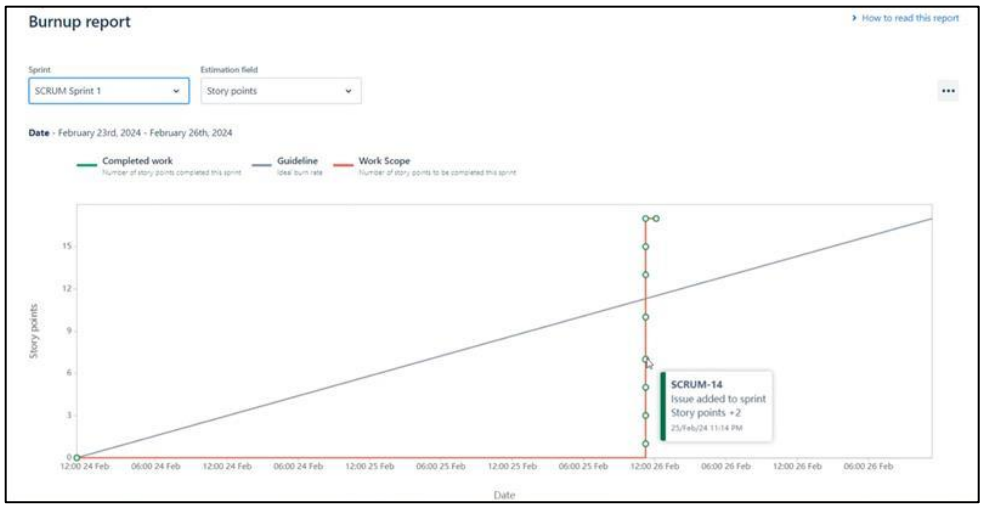


Figure 23 Sprint 1 Burnup Report

Figure 24 Cumulative Flow Diagram after Sprint 1

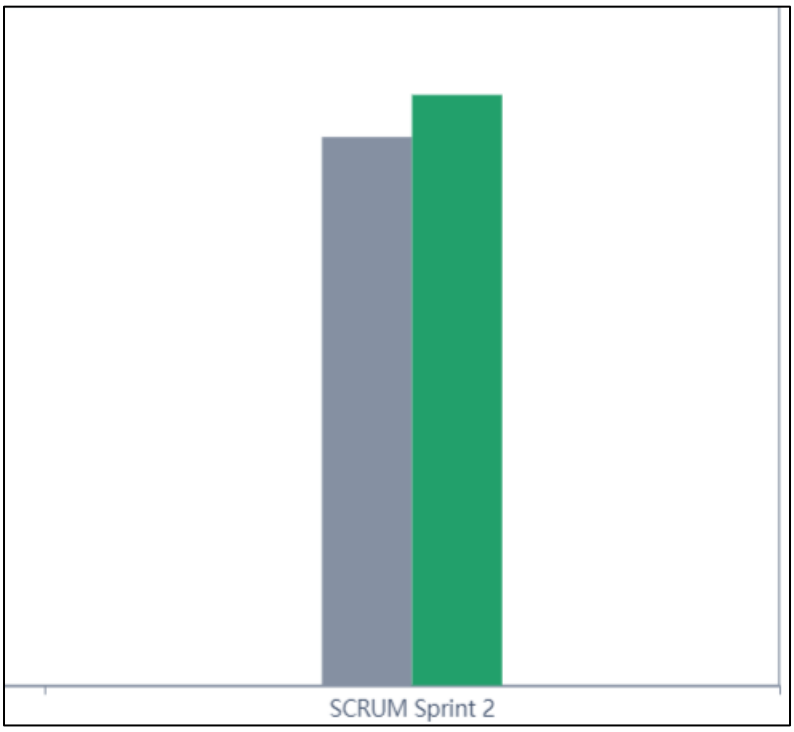


Figure 25 Sprint 2 Velocity Report (Commitment 26 and Completed 28)

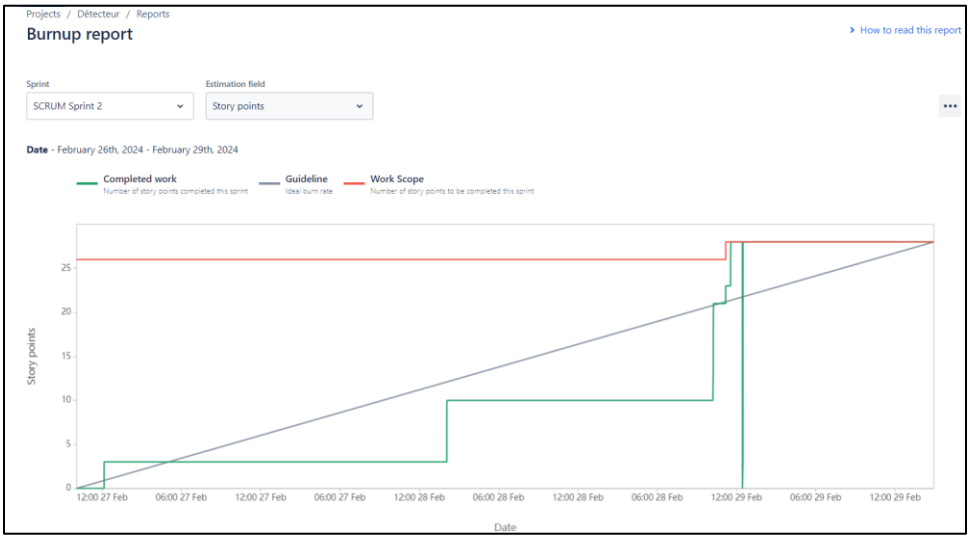


Figure 26 Sprint 2 Burnup Report

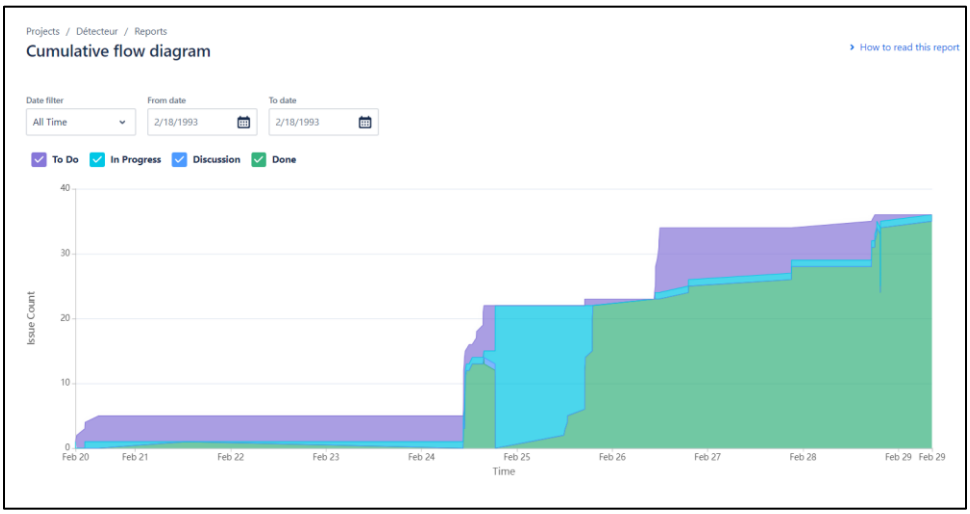


Figure 27 Cumulative Flow Diagram after Sprint 2

VII. Conclusion

Our project, which aimed at distinguishing between texts generated by Large Language Models (LLMs) and those written by humans, has successfully met its objectives through the integration of Agile Scrum methodology.

Below is the summary of the Scrum practices followed based on our Project Outcome:

- **Sprint Backlog Selection and User Story Adaptation**
  Our approach to managing the sprint backlog and adapting user stories has been the foundation of our project's success. Through iterative refinement across Sprints (Sprints 1 and 2), we have gradually learned the process of assigning user story points, ensuring that each sprint is both ambitious and achievable.

- **Collaboration and Open Communication**
  Daily stand-ups and sprint retrospectives became the venue for discussing the team's issues and bottlenecks. These sessions helped in quickly identifying and resolving challenges encountered by each member, which is essential in minimizing potential project delays and maintenance of the integrity of our classification system.

- **CI/CD Continuous Integration and Continuous Delivery/ Continuous Deployment**
  Embracing CI/CD, our team maintains an unwavering dedication to seamless software delivery. Through meticulous integration of code changes into a shared repository multiple times a day, we ensure that any conflicts are swiftly identified and resolved. This process is complemented by automated testing suites that guarantee the stability and reliability of our software with each iteration. Our commitment to Continuous Delivery/Continuous Deployment allowed us to push validated changes into production promptly, enabling rapid feedback loops and facilitating agile responses to evolving user needs. This iterative approach, rooted in collaboration and automation, empowered us to consistently deliver high-quality

features and enhancements, driving tangible value for our stakeholders and distinguishing our outputs in the competitive landscape.

- **Continuous Feedback and Iterative Improvement**
  The output/ deliverables feedback which is established through sprint reviews has been essential for our text classification methodologies' continuous refinement. Through the regular engagement with the Product Owner during sprint reviews (every sprint), the team was able to gradually grasp and align with the project's goals and/or requirements.

- **Value-Driven Development and Alignment with Objectives**
  Our team's commitment to value-driven development ensured that our efforts were consistently aligned with our project's core goals. The clear understanding of the Product Owner's requirements/ backlogs, refined through continuous sprint planning and sprint reviews, ensured that every sprint delivered is of tangible value towards distinguishing between AI-generated and human-generated content.