Revised High Level Design

Continuous Integration (CI) for Pothole Detection Model

- CI Pipeline Setup:

Automated Testing: automate unit testing of data preprocessing scripts, model training code, and deployment scripts.

Integration Testing: Use ClearML to manage different versions of models and datasets, ensuring compatibility and performance across various versions.

- Automated Model Validation:

Data Integrity Testing: Use Great Expectations to automate data validation, ensuring data quality and consistency before training.

Model Performance Testing: Automate the evaluation of model performance on a validation dataset, ensuring new changes do not reduce performance.

Continuous Training (CT) for Pothole Detection Model

- Automated Retraining Pipelines:

Trigger-Based Retraining: Set up automated retraining pipelines triggered by specific criteria such as new data availability or performance degradation. Use ClearML to manage these workflows.

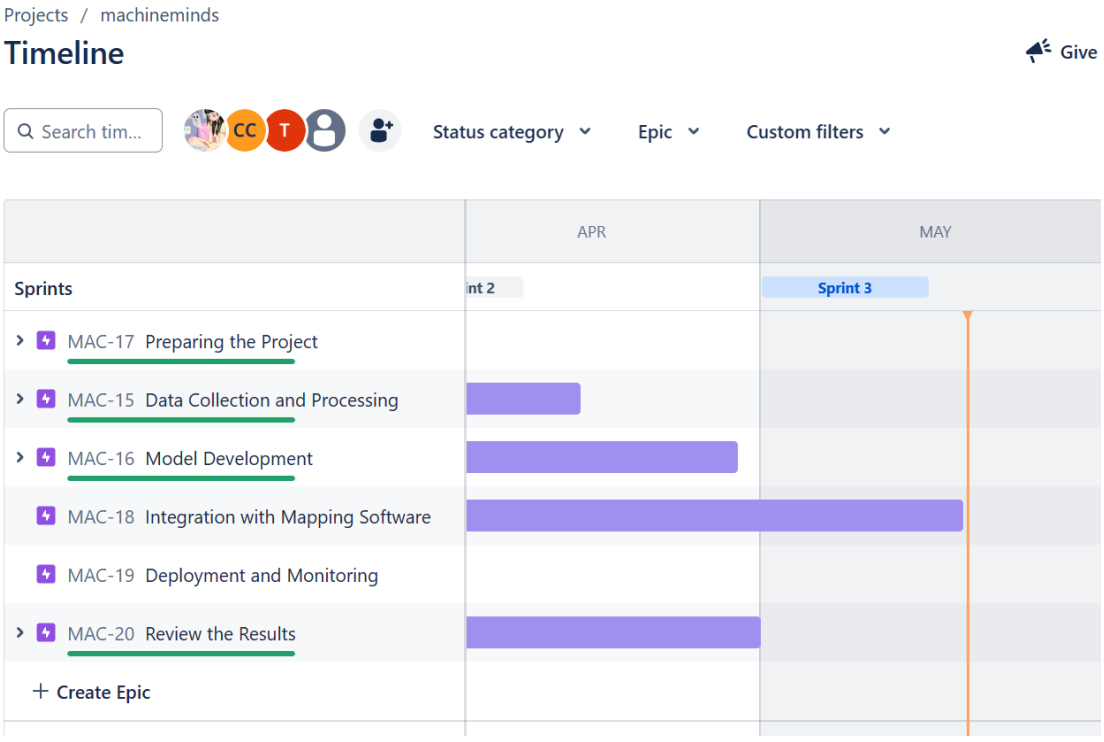
Data Preprocessing Automation: Automate data preprocessing steps to ensure efficient handling of new data and preparation for training.

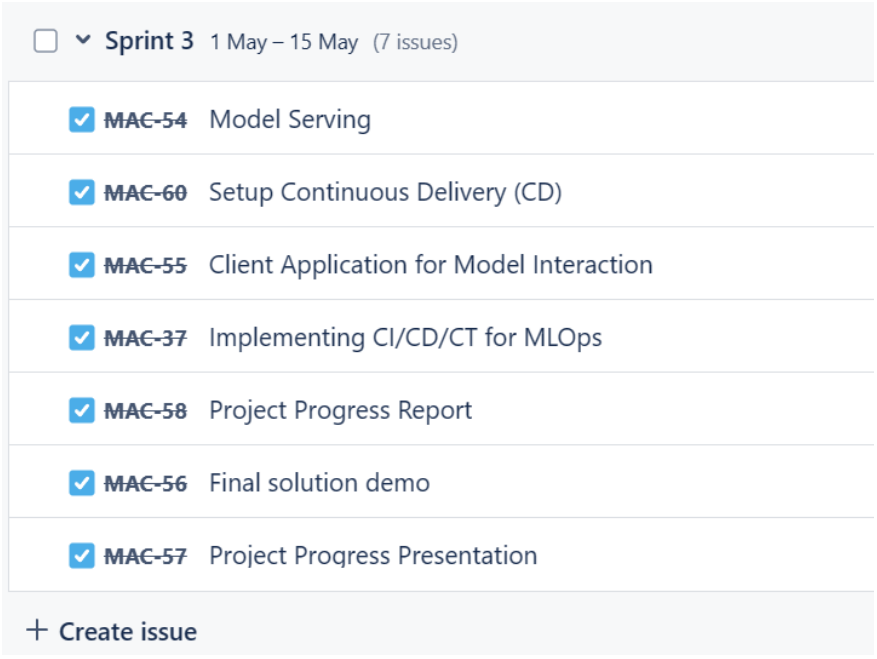
- Model Evaluation and Versioning:

Model Serving: Employ ClearML's model serving functionalities to manage the entire deployment lifecycle within the same environment that maintains model tracking and versioning.

Automated Comparison: Use ClearML to automatically log training metrics and facilitate model comparisons, guiding decisions to promote models to production based on performance.

Revised Project Agile Workspace





Sprint Review Report and Board

Sprint Period: May 1 to May 31

Team: Choonsik, Sarah, Tran

Achievements

- Model Optimization: Significant improvements in model accuracy and recall through advanced hyperparameter tuning and ensemble methods.
- CI/CD Pipeline Automation: Automated key parts of the CI/CD pipeline, reducing manual intervention and increasing deployment efficiency.
- Infrastructure Scalability: Successfully scaled the infrastructure to handle increased loads using Kubernetes and cloud-based solutions.

Challenges

- Addressing performance bottlenecks in data and model pipelines.
- Managing the complexity of the expanded infrastructure and ensured smooth operation.
- Balancing the time required for innovation with maintaining current systems.

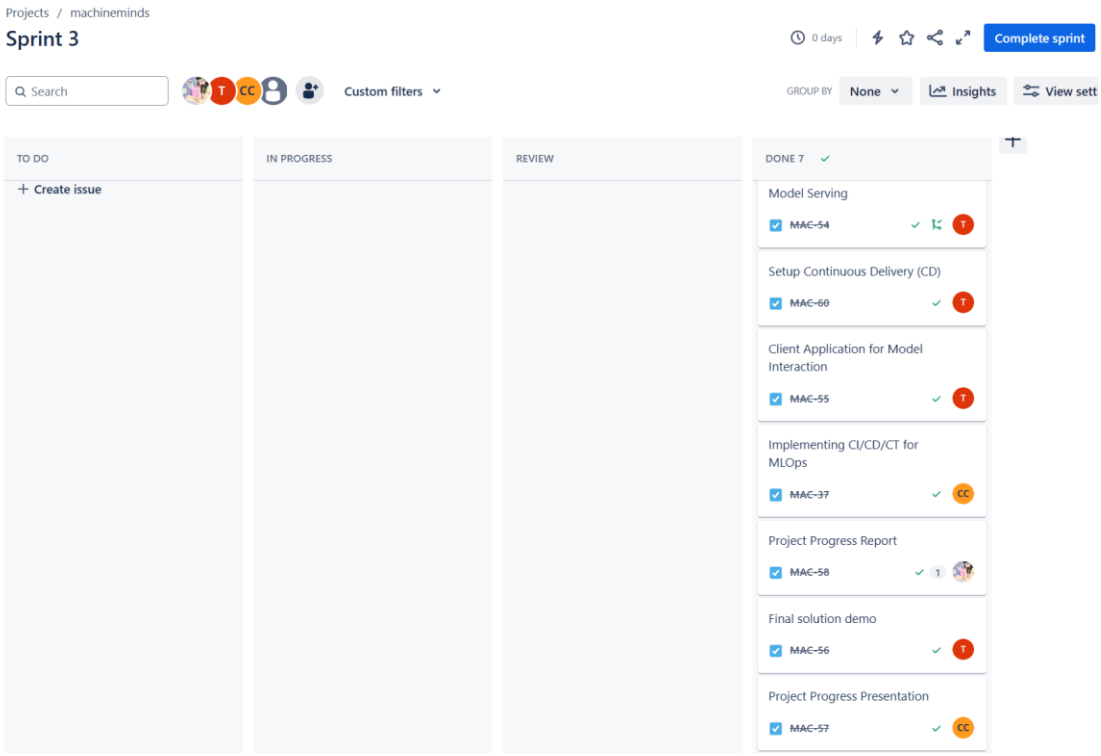
Reflection on Sprint 3

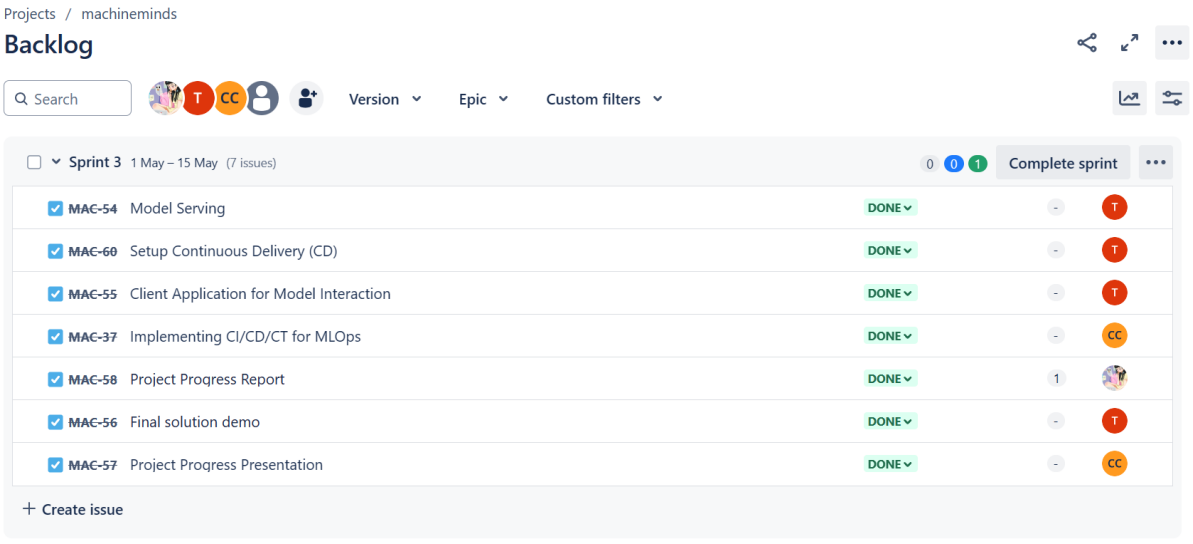
Sprint 3 aims to significantly enhance the MLOps workflow for the pothole detection model, focusing on advanced automation, model optimization, and infrastructure scalability. By refining CI and CT processes, the team will create a robust, efficient, and scalable machine learning pipeline capable of handling high-scale deployments and complex workflows.

Planned Actions for Next Sprint

- Automated Retraining Pipelines: Refine automated retraining pipelines based on performance metrics and new data.
- Advanced Monitoring Systems: Implement predictive alerts and automated incident response mechanisms.
- High-Scale Deployment Preparation: Prepare for large-scale deployments in production environments.

Revised Product Backlog





Jira

statusCategory: Done
Sorted by: Updated descending
1–29 of **29** as at: **19/May/24 7:13 PM**

T	Key	Summary	Assignee	Reporter	P	Status	Resolution	Created	Updated
☑	MAC-60	Setup Continuous Delivery (CD)	trongnghia.tran	Choonsik Cho	=	DONE	Done	01/May/24	19/May/24
☑	MAC-55	Client Application for Model Interaction	trongnghia.tran	Choonsik Cho	=	DONE	Done	01/May/24	19/May/24
☑	MAC-54	Model Serving	trongnghia.tran	Choonsik Cho	=	DONE	Done	01/May/24	19/May/24
☑	MAC-37	Implementing CI/CD/CT for MLOps	Choonsik Cho	Choonsik Cho	=	DONE	Done	30/Mar/24	19/May/24
☑	MAC-56	Final solution demo	trongnghia.tran	trongnghia.tran	=	DONE	Done	01/May/24	19/May/24
☑	MAC-58	Project Progress Report	Sarah Altalhi	trongnghia.tran	=	DONE	Done	01/May/24	19/May/24
☑	MAC-57	Project Progress Presentation	Choonsik Cho	trongnghia.tran	=	DONE	Done	01/May/24	19/May/24

Release Notes

Overview:

Sprint 3 focused on elevating the MLOps workflow for the pothole detection model by enhancing model optimization, pipeline efficiency, and system scalability.

Achievements:

- Model Performance: Achieved significant improvements in model performance through advanced hyperparameter tuning and ensemble methods.
- Pipeline Automation: Established highly automated workflows within the CI/CD process, ensuring efficient model testing, validation, and deployment.
- Infrastructure Scalability: Enhanced infrastructure to support high-scale deployments using Kubernetes and cloud-based solutions.

Progress and Next Steps:

- Continue refining models using insights gained from extensive A/B testing and advanced evaluation metrics.
- Explore additional ensemble techniques and parameter tuning to further boost performance metrics.
- Implement more granular monitoring of model drift and data changes to ensure model accuracy over time.
- Advance monitoring systems with predictive alerts and automated incident response mechanisms to preemptively address potential issues.
- Integrate deeper analytics capabilities to provide more detailed insights into model performance and system health.
- Prepare for high-scale deployments in production environments, ensuring robust infrastructure and optimized performance.
- Conduct extensive load testing and performance benchmarking to validate system readiness for large-scale operations.