
Waterfowl Travel: Team Iteration 4 Report



3/31/2025

Prepared by the students of CSC 4615 – Spring 2025

Jacob Sullivan

Tania Perdomo Flores

Revel Etheridge

Kenny Adams

Drew Burkhalter

Breanna Woosley

Under the direction of

Dr. William Eberle (Professor)

Stories Identified:	3
Plan For Next Iteration:	4
Burnup Chart:	5
Retrospective Summary:.....	5
Team Temperature:	7
Showcase:.....	8
Appendix:	8

Stories Identified:

Authors: Revel Etheridge, Drew Burkhalter

Contributors: Revel Etheridge, Drew Burkhalter, Jacob Sullivan, Kenny Adams, Breanna Woosley, Tania Perdomo Flores

During this iteration, developers focused on improving their models in terms of efficiency, ability to pull weather data from NOAA, and accountancy for weather covariances. All models were finalized by their respective development member(s), including RNN, Random Forest, LSTM, and GNN. Model prediction capabilities were compared, with the LSTM and GNN models showing the highest level of prediction accuracy and efficiency. As a result, these two models were selected to be the focus of our resources in terms of time and development moving forward. The LSTM model adopted a k-d tree to automate selection of the nearest weather data point per duck to allow for more efficient predictions. The GNN model also incorporated an edge weighting prototype with NetworkX to allow for weather covariances to gain individual factor weights in its nearest neighbor calculations.

- **Stories completed**

- As a developer, I want to optimize my machine learning model and predictions so that I can finalize migration prediction results for the LSTM model. (Tania Perdomo Flores, Breanna Woosley) [5pts]
- As a developer, I want the map to be further developed with features like location data, weather, and more thorough duck data so that I can have a cleaned-up map for us to discuss with the client/stakeholder. (Jacob Sullivan) [5pts]
- As a developer, I want to refine my machine learning model and predictions so that I can finalize migration prediction results for the RNN model. (Drew Burkhalter) [8pts]
- As a developer, I want to be able to extract weather information from NOAA so that I can incorporate it with my ML model. (Jacob Sullivan, Tania Perdomo Flores, Breanna Woosley, Kenny Adams, Revel Etheridge, Drew Burkhalter) [5pts]
- As a developer, I want to refine my machine learning model and predictions so that I can finalize migration prediction results for GNN. (Revel Etheridge) [8pts]
- SPIKE: As a developer, I want to be able to access and utilize the HPC efficiently so that I can work faster on the project. (Jacob Sullivan, Tania Perdomo Flores, Breanna Woosley, Kenny Adams, Revel Etheridge, Drew Burkhalter) [3pts]
- As a developer, I want to refine my machine learning model and predictions so that I can finalize migration prediction results for Random Forest Model. (Kenny Adams) [8pts]
- As a developer, I want to implement a k-d tree within my LSTM model to identify the nearest NOAA weather data location so we can enhance the model's spatial accuracy by aligning predictions with the most relevant local weather conditions. (Tania Perdomo Flores, Breanna Woosley) [8pts]
- As a developer, I want to determine boundaries of latitude/longitude for the flyway so that we can have an automated validation system for pulling weather points. (Tania Perdomo Flores, Revel Etheridge) [8pts]
- As a developer, I want to create an edge weighting prototype based off covariances so that weather data can be accounted for in migration predictions. (Revel Etheridge) [8pts]
- As a user, I want to explore implementing data vectorization so we can enable batch computation and improve the efficiency and scalability of the model's training and inference processes. (Breanna Woosley) [5pts]

Stories Dropped

- As a user, I want to adjust prediction parameters (e.g., tailwind effect or barometric pressure thresholds) so I can explore how changes impact migration patterns. (Breanna Woosley, Tania Perdomo Flores, Kenny Adams) [5pts]
- As a developer, I want to refine my machine learning model and predictions so that I can finalize migration prediction results for Convolutional LSTM. (Tania Perdomo Flores) [8pts]

Plan For Next Iteration:

Author: Kenny Adams, Drew Burkhalter

Contributors: Revel Etheridge, Drew Burkhalter, Jacob Sullivan, Kenny Adams, Breanna Woosley, Tania Perdomo Flores

Iteration 5 will involve our switch towards a focus on two models instead of the five models we had before. Due to time constraints, we came to the decision that choosing two models will allow us to create both an effective model and have a way to compare the data with the predictions. For this purpose, we will be using a GNN and LSTM model. We will use a k-d tree, which will be useful due to the need for spatial assignment. Additional improvements to the prediction algorithm will involve improved confidence intervals and making the covariances of the environment more integral to the prediction. The environmental factors will be the most difficult to implement due to their massive variety and volume. During the last iteration we had difficulty implementing NOAA and will continue to improve our usage of it while being open to other methods of data collection.

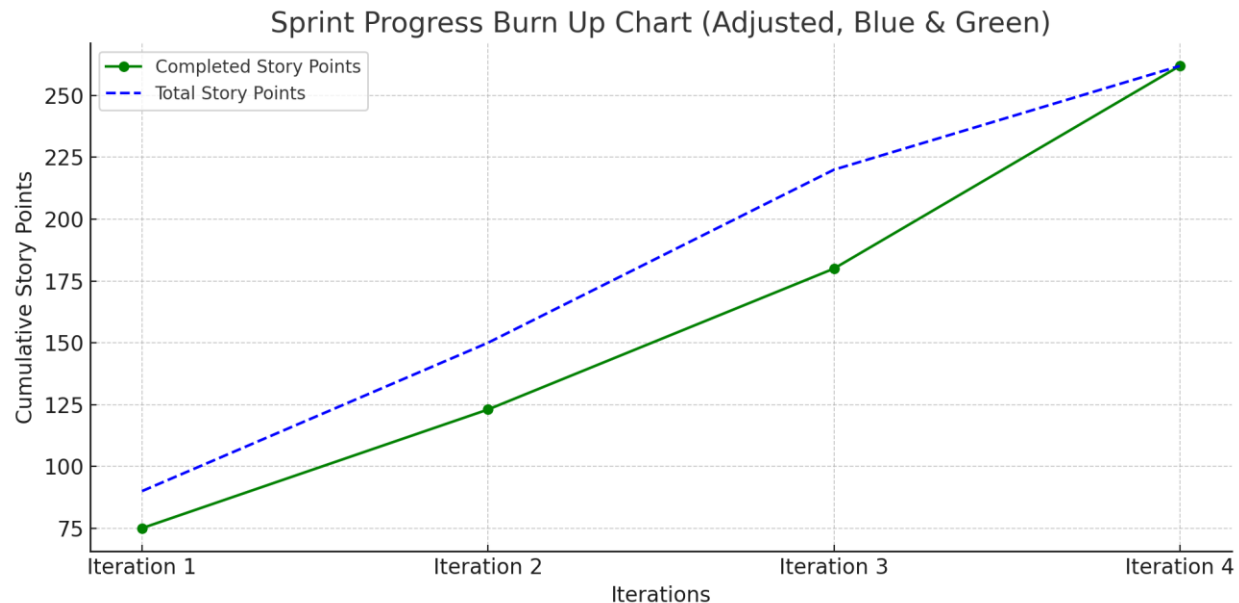
Additionally, due to the increase of scale, Wix will no longer serve our purposes; therefore, we will need to transfer to another tool and are currently looking at incorporating Flask. Complications could arise with this, and we will need to communicate with Dr. Cohen to determine what path is the most effective.

- As a user, I want to see the probability/direction of ducks leaving specific areas, along with the covariance of environmental factors (e.g., weather, temperature) that influence these departures, to better understand migration triggers. (Jacob Sullivan) [5pts]
- As a user, I want confidence intervals on predictions to gauge reliability and variability. (Tania Perdomo Flores, Breanna Woosley, Kenny Adams, Revel Etheridge, Drew Burkhalter) [5pts]
- As a developer, I want to implement the k-d tree to the GNN model to round off duck locations so that we give a more generic prediction. (Revel Etheridge) [5pts]
- As a developer, I want to finalize the GNN model and integrate a k-d tree for efficient spatial querying and NOAA weather parameter verification, so that I can model duck migration as a dynamic graph and make accurate, environmentally informed predictions. (Revel Etheridge, Kenny Adams, Drew Burkhalter) [13pts]
- As a developer, I want to finalize the LSTM model and integrate the k-d tree, data vectorization, and NOAA weather parameter verification so that I can generate accurate, location-aware predictions of duck migration patterns. (Tania Perdomo Flores, Breanna Woosley) [13pts]
- As a developer, I want to move away from Wix and explore more robust hosting options, such as Flask, to better support our data needs and enable a complete redesign of our frontend and user interface. (Jacob Sullivan) [13pts]

Burnup Chart:

Author: Jacob Sullivan

Contributors: Revel Etheridge, Drew Burkhalter, Jacob Sullivan, Kenny Adams, Breanna Woosley, Tania Perdomo Flores



We began Iteration 4 with a total of 81 story points, successfully completing 68 of these points, while 13 points were dropped due to the upcoming showcase deadline, as certain machine learning models proved less effective during testing. This iteration involved significant refinement of our migration prediction models, bringing our algorithm closer to a Minimum Viable Product (MVP), expected by the next iteration. All team members faced challenges integrating weather data into the models, causing a slowdown, however this period allowed us to make substantial progress toward successful implementation. Additionally, after evaluating multiple machine learning models, we've narrowed our focus down to two models demonstrating the most promising results. This combining of effort from five to two models is anticipated to increase our overall development progress significantly. Lastly, minor updates were made to enhance the UI map visualization of duck migration patterns. In summary, Iteration 4 was marked by trial and error in weather integration, refining and selecting key ML models, and minor adjustments to the visualization map.

Retrospective Summary:

Author: Breanna Woosley

Contributors: Revel Etheridge, Drew Burkhalter, Jacob Sullivan, Kenny Adams, Breanna Woosley, Tania Perdomo Flores

What Has Gone Well?

- Team Communication
 - The team maintained strong and consistent communication throughout the iteration, which supported clear task assignments and collaboration.
- Efficiency with GitHub and HPC
 - There was noticeable improvement in managing code via GitHub and leveraging HPC resources more effectively. Version control was smoother, and job submissions were more organized.
- NOAA Weather Data Integration
 - The process of extracting and understanding NOAA weather data has become more streamlined. The team now has clarity on the structure and use of the weather data relevant to the project's predictive goals.
- Scrum Master Organization
 - Our Scrum Master has done a great job keeping tasks organized and helping maintain focus on iteration goals.

What Could Be Improved?

- Project Progress
 - The project is not behind schedule, but we experienced some initial delays due to limited access to high-performance computing resources, which impacted early code troubleshooting. Now that access has been restored, development is moving forward and aligned with upcoming milestones.
- Seeking Technical Support
 - There were missed opportunities to reach out for help with challenging aspects of the project. Proactively consulting experts or mentors could help avoid unnecessary delays.

What Questions Do You Have About the Project or Process?

- Data Vectorization Techniques
 - What are the most efficient and scalable ways to vectorize temporal and spatial weather data for machine learning models?
- KD-Tree Validation
 - How can we best confirm the accuracy and performance of our KD-tree implementation, especially in matching spatial data points?
- Haversine Error Calculations
 - What are acceptable error thresholds for haversine-based distance calculations in this context, and how do they impact model precision?

What Action Items Need to Be Addressed?

- Model Result Integration
 - Develop a method to combine results from the GNN and LSTM models. Explore ensemble strategies or decision logic for integrating predictions.
- Debug Weather Data Pipeline
 - Continue troubleshooting the weather integration pipeline. Pay special attention to edge cases, data mismatches, and API inconsistencies.

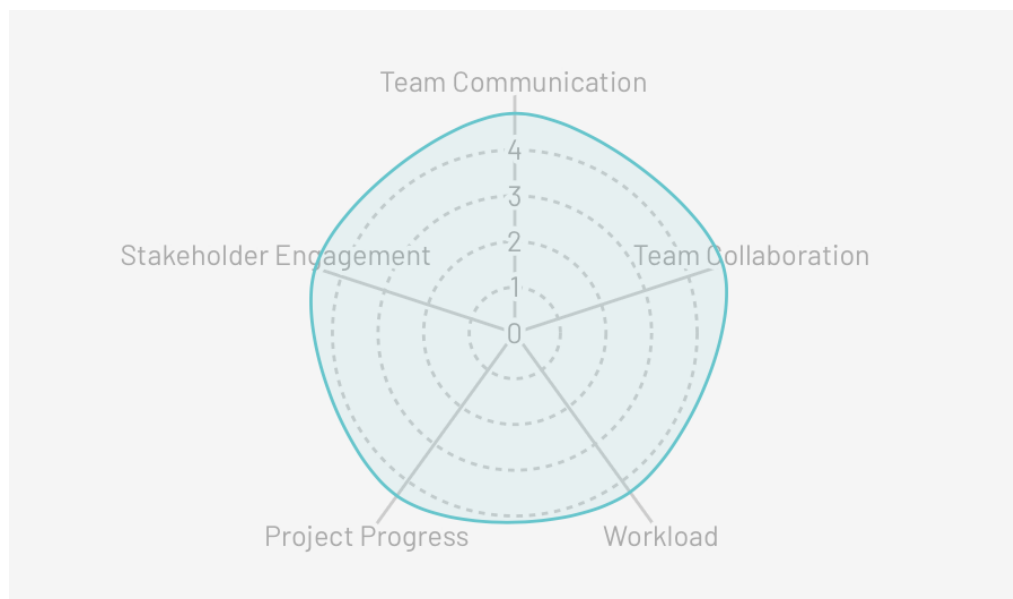
- Coordinate Feature Engineering
 - Align feature engineering efforts across both models to ensure consistency in input data and maximize predictive synergy.

Team Temperature:

Author: Tania Perdomo Flores

Contributors: Revel Etheridge, Drew Burkhalter, Jacob Sullivan, Kenny Adams, Breanna Woosley, Tania Perdomo Flores

Spider chart along with a description and summary of our team's temperature.



Summary:

After the fourth iteration of this project, our spider diagram reflects a well-balanced, yet slightly irregular, distribution of team performance across key areas. Team communication and collaboration remain strong, indicating that internal discussions and teamwork dynamics are effective. However, stakeholder engagement is slightly lower than previous iterations, primarily due to missed communication with our client and the client missing a meeting. This gap is not significant to our progress, however, as we have had consistent stakeholder engagement.

Project progress and workload appear moderate, mainly due to delays caused by waiting for the HPC and figuring out the weather API integration for our models. While these technical hurdles slowed momentum, we've made significant strides toward the end of the iteration. Though we feel slightly behind, these challenges have been addressed, positioning us to hit the ground running in the next iteration. With these roadblocks cleared, we can expect smoother execution and a stronger trajectory moving forward.

Showcase:

Author: Tania Perdomo Flores

Contributors: Revel Etheridge, Drew Burkhalter, Jacob Sullivan, Kenny Adams, Breanna Woosley, Tania Perdomo Flores

At the end of this iteration, our team prepared a showcase of the effort completed. The showcase demonstrates the specific user stories completed during the iteration.

<https://youtu.be/-iSHUvNT1Ao>

Appendix:

Author: Jacob Sullivan

Contributors: Revel Etheridge, Drew Burkhalter, Jacob Sullivan, Kenny Adams, Breanna Woosley, Tania Perdomo Flores

- Kanban Board and GitHub
 - <https://github.com/users/JakeSul1023/projects/1/views/1> (Kanban Board GitHub)
 - https://github.com/JakeSul1023/Team-7_Waterfowl (Repository GitHub)
 - <https://github.com/JakeSul1023/DuckData-WebApp> (Web-Page Repo)