# Waterfowl Travel: Team Iteration 3 Report


3/3/2025

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:

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

During this iteration, team members were assigned different machine learning models to train and test using a smaller sample of the Waterfowl data (from mid-August 2024 to mid-September 2024) exported from Movebank. The objective was to identify which model most accurately predicts Waterfowl migration. We first developed code to sort and format the data for compatibility with our models. Next, we implemented code to process the data through the models and assess the accuracy of their predictions. In addition, Jacob shifted focus during the iteration to proactively work on further website and code integration, including the implementation of Deck.gl.

- **Stories Completed**
  - As a developer, I want to implement GNN/RNN Architecture and optimize performance so that I model migration as a graph problem. [8 pts] (Revel Etheridge/Drew Burkhalter)
  - As a developer, I want the system to pull real-time weather data from APIs, so the migration predictions reflect current conditions. [5 pts] (Kenny Adams)
  - As a developer, I want to build out the LSTM architecture, train it, and optimize it, so that I can predict the next migration coordinates based on past movements. [8 pts] (Breanna Woosley)
  - As a developer, I want to implement Deck.gl and OSM with the sample set of historical data so that I can understand whether the data can be handled given its size. [8 pts] (Jacob Sullivan)
  - As a developer, I want to train the Convolutional LSTM model and to tune hyperparameters so that I can evaluate performance on predicted routes. [8 pts] (Tania Perdomo Flores)
  - As a developer, I want to preprocess the data and set up sequence generation and an input pipeline so that I can set up the infrastructure for the Stacked LSTM. [5 pts] (Breanna Woosley)
  - As a developer, I want to clean the data and perform feature engineering so that I can set up the infrastructure for the Convolution LSTM. [5 pts] (Jacob Sullivan, Tania Perdomo Flores)
  - As a user, I want an easy-to-use interface that is clear and concise so that I can navigate the app easily. [3 pts] (Jacob Sullivan)
  - As a developer, I want to build the stopover graph from historical data for the baseline for a GNN/RNN so that we can implement the GNN/RNN architecture. [5 pts] (Drew Burkhalter, Revel Etheridge)
  - As a developer, I want to preprocess the data and set up sequence generation and an input pipeline so that I can set up the infrastructure for the Random Forest Model. [5 pts] (Kenny Adams, Tania Perdomo Flores)
- **Stories Still in Progress**
  - As a developer, I want to preprocess the data and set up sequence generation and an input pipeline so that I can set up the infrastructure for the Random Forest Model. [8 pts] (Kenny Adams)
- **Stories Dropped**

- As a developer, I need the data to be integrated with my code so that it can update automatically. [8 pts]

# Plan For Next Iteration:

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

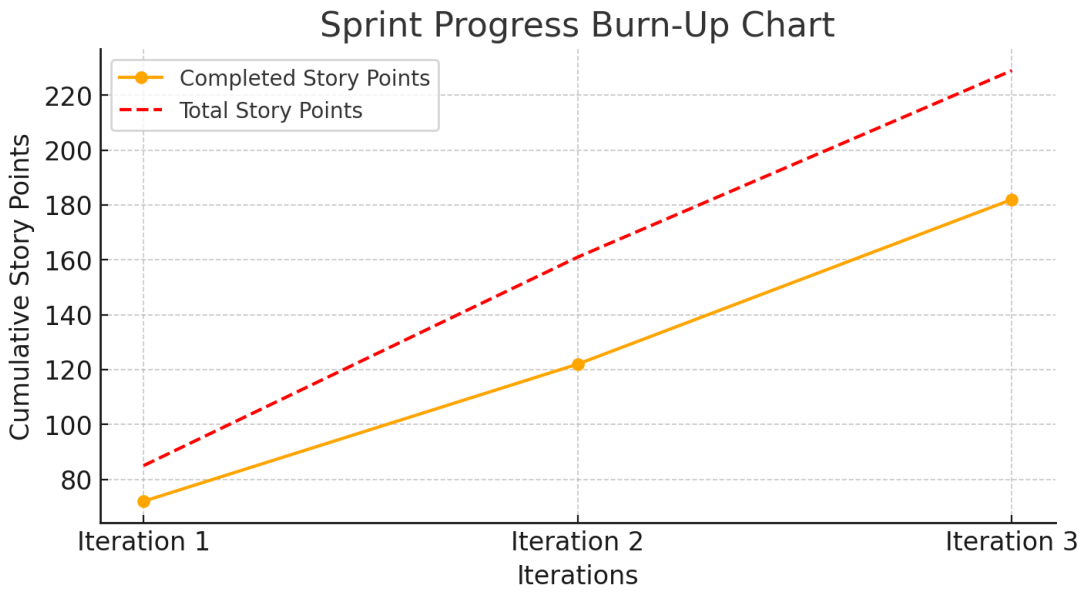
Iteration 4 will focus primarily on advancing our machine learning models. Due to the HPC being down for two weeks during this period, we were unable to thoroughly test the models we developed. Even when we had access to the HPC, we faced challenges in using it effectively. In the next iteration, we aim to gain better control over the HPC and run tests to evaluate the models' accuracy. Based on the results, we will identify areas for improvement.

Once each model is functioning properly, we plan to compare their performance and determine which one is the most effective. Additionally, we are exploring more efficient methods for retrieving weather data from NOAA. Our goal is to reduce calculation time by minimizing API requests and limiting the volume of data we pull by using specific weather data based on the area of the duck's location instead of the entire weather data for the U.S.

- As a developer, I want to define features and train them on past location data with labeled routes so that I can predict likely stopover locations. [2 pts] (Jacob Sullivan)
- 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 refined map to discuss with the client/stakeholder. [8 pts] (Jacob Sullivan)
- As a developer, I want to refine my machine learning model and predictions so that I can finalize migration prediction results for Stacked LSTM. [13 pts] (Breanna Woosley)
- As a developer, I want to refine my machine learning model and predictions so that I can finalize migration prediction results for the Random Forest Model. [13 pts] (Kenny Adams)
- As a developer, I want to refine my machine learning model and predictions so that I can finalize migration prediction results for the GNN. [13 pts] (Revel Etheridge)
- As a developer, I want to refine my machine learning model and predictions so that I can finalize migration prediction results for the RNN. [13 pts] (Drew Burkhalter)
- As a developer, I want to refine my machine learning model and predictions so that I can finalize migration prediction results for the Convolution LSTM. [13 pts] (Tania Perdomo Flores)
- As a developer, I want to be able to extract weather information efficiently from NOAA so that I can incorporate it with the ML model. [13 pts] (Assigned to all team members)
- 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. [5 pts] (Assigned to all team members)

# Burnup Chart:

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




As shown in the burnup chart, we started with 68 story points, of which 60 have been completed. This iteration faced delays due to the HPC being down for two weeks, followed by an additional week of learning how to interact with it. However, we made up for lost time by implementing project changes, including the new map and switching to a more efficient coding language for algorithm development. Overall, the iteration was successful, with all stories completed except for one—the finalization of the random forest algorithm. The other three stories were completed and will be refined in iteration 4. With an extra three weeks allocated for iteration 4, we should be able to get back on track with our projected velocity.

# Retrospective Summary:

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

**What Has Gone Well?**

- **Machine Learning Model Training and Validation**
  The team has made progress in training and validating machine learning models, ensuring it aligns with the project's objectives. Initial test runs indicate promising results, with refinements improving prediction accuracy.
- **HPC Meeting and Understanding**
  Discussions around high-performance computing (HPC) have helped clarify its role in handling

large-scale computations. The team has a better understanding of how to leverage HPC resources for model training.

- **Clarifying Questions Addressed**
  Several key uncertainties regarding model implementation, data integration, and computational resource allocation have been resolved. This has streamlined decision-making and improved workflow efficiency.

**What Could Be Improved?**

- **Efficiency with HPC**
  While the team has gained knowledge on HPC usage, there is still room for improvement in optimizing job submissions, managing computational resources, and debugging issues related to parallel processing.
- **GitHub Usage and Code Management**
  There have been some challenges in consistently pushing updates to GitHub and maintaining an organized repository. Implementing a more structured version control process will help improve collaboration and reduce redundancies.

**What Questions Do You Have About the Project or Process?**

- **Model Accuracy Optimization**
  - What specific techniques or adjustments can be made to further improve model accuracy while maintaining computational efficiency?
- **Efficient API Usage**
  - What are the best practices for structuring NOAA API calls to minimize request times while retrieving the necessary data? Are there bulk download options that could reduce request overhead?

**What Action Items Need to Be Addressed?**

- **Optimize NOAA API Calls**
  - Research NOAA's API documentation for batch request methods or data bundling options.
  - Develop a strategy for reducing redundant API calls and caching frequently used data.
- **Standardize Data Preprocessing**
  - Implement a preprocessing pipeline to clean, normalize, and format weather data before feeding it into the model.
  - Identify which weather parameters are most relevant to migration patterns and filter out unnecessary data.
- **Feature Engineering for Model Integration**
  - Determine the best approach for encoding weather data into the machine learning model.
  - Experiment with aggregating weather variables over different time windows to identify the most predictive features.

# 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 third iteration of this project, our spider diagram reflects strong performance across *Team Communication, Team Collaboration, Workload,* and *Stakeholder Engagement*. Open communication, clear responsibilities, and strong stakeholder relationships have been key to our success. Internally, structured communication has kept everyone aligned with project goals. The client's active engagement through weekly meetings has also helped refine features and ensure user stories stay on track.

However, *Project Progress* scored lower due to delays in accessing the high-performance computer (HPC). First, we waited for initial access, then hardware issues in the department caused further setbacks, and we continued waiting for availability after the incident. Without the HPC, we could not run our prediction models, as our machines lacked the necessary processing power. We eventually gained access in the final week of this iteration, but this left little time to fully utilize it.

Now that we have access to the HPC, we expect greater flexibility in running our models without the resource limitations that previously slowed progress. With this capability, we can run models more efficiently, implement additional features, and work with a larger sample of historical data provided by the client. This should eliminate major resource roadblocks and allow us to make more significant strides in the next iteration. Additionally, as we've improved our workload management and team collaboration, we anticipate an overall boost in team temperature, fostering a more productive and cohesive working environment.

# 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/BIErlnBzQJY

# 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)