The current state of your project: what works, and what doesn't. This needs to match the code and data committed in your team's repo on GitHub.

Both the random forests model and the multiple regression model have been refined to yield a smaller loss. The team has created a testing document to systematically test hyperparameters. The ensemble is now complete and runs without any errors and the model has been incorporated into the website. Students have independently verified that the website can be run using the steps outlined in the README. We have begun to work on the RNN as soon as we received feedback from milestone 2, but we are not done yet.

Any feature changes to the proposal. As we discussed in class, no reason is necessary for M1, but any feature changes in M2 et seq will need a technical justification. Obviously, feature changes need to keep in spirit with the project, so you can't for example remove all the hard parts :)

We will reintroduce the RNN model to our ensemble and thus change the number of models in our ensemble to 3.

On the website, a square or hexagonal outline will surround the location of the wildfire as opposed to a circle. This is due to the limitations of drawing complex polygons with the gmplot library.

The current challenges / bottlenecks you are facing, both technical and otherwise, and what you are thinking of doing about it.

The team is still hoping to get access to the fire fuel data. As an alternative, we are looking into getting data from 'British Columbia Fuel Hazard Assessment and Abatement Fire Risk' which is publicly available. Challenges for the alternative data include: obtaining data from an access-only web app.

We foresee a data related bottleneck as the RNN uses different data from our other two models. RNN requires data following a set time interval from the same location. We do not have this. The team will meet on the tenth to discuss work arounds.

For each team member, what tasks were done and which tasks are underway.

Student 1 (Anna Wang): Tested and optimized the Random Forest model with student 2. refer to the training document below: https://docs.google.com/spreadsheets/d/1iEXAQaKGtLsYO9Hde9jv-uWRv64l4uHNLjKWk1qWP2E/edit?usp=sharing. Created the ensemble (notebook, Part 6). Made changes to part 5 so both models' input features and labels used the same dataframe. Wrote project description for the final report. Students 1 and 3 are making progress on the RNN model.

Student 2 (James Ardian): Tested and optimized the Random Forest model with student 1. Student 2 also made some headway, and is currently attempting, to integrate the full ML model into the website.

Student 3 (Michele): split the dataset where 80% of the data is used for training and the rest is used for testing. Originally, all of the data was used to train and the same data was used to test for accuracy. Student 3 also did pair programming with Student 4 to normalize the data by using z scores. This minimized the loss compared to what we had in Milestone 3. Student 1 also tried applying the log to the dataset. This approach minimized the loss significantly and we will use this method. Students 3 and 1 began work on the RNN model.

Student 4 (Alex): Worked with student 3 on paired programming tasks. Normalized data using z-scores. Experimented using t-distributions to remove outliers (ie. remove data point if a threshold number of features were above 3 SD from the mean). Improved loss slightly.