Okay guys,

Since Kayode only has 30min I am going to do the breakdown of the presentation, I am going to pose the breakdown of the presentation and the report on discord.

Here is the breakdown of the work for presentation:

Kayode (3 slides):

|  |
| --- |
| How was the new data labeled/collected? |
| How does the newly added data compare with the original data? |
| Your part of the discussion |

Bhavyai (3 slides):

|  |
| --- |
| How was the data preprocessed? |
| How are the misclassifications of the best performing model distributed? (Go through one example) |
| Your part of the discussion |

Mike (3 slides + intro and conclusion)

|  |
| --- |
| How do the models perform on the original data vs the new + original data? |
| How does the performance of the models change based on the choice of hyper parameters? |
| My part of the discussion |

Please give me your portion of the presentation no later than **3pm Monday Dec 6**!

I accept slides plus 1 audio file per slide (do not included within powerpoint, send it to me in separate files) or video in mp4 (don’t send me all in one video, separate the recordings in the sections I listed out for you).

Now for the report. As I mentioned in a previous message, below is the breakdown for the report based on what we already did (you can find it on github <https://github.com/meng-ucalgary/ensf-612-project/blob/main/612%20report%20work%20distribution.xlsx> ):

|  |  |
| --- | --- |
| Section | Resource |
| Page 1 contents | Mike |
| Abstract/Intro/conclusion | Bhavyai |
| How was the new data labeled/collected? | Kayode |
| How does the newly added data compare with the original data? | Kayode |
| How was the data preprocessed? | Bhavyai |
| How do the models perform on the original data vs the new + original data? | Mike |
| How does the performance of the models change based on the choice of hyper parameters? | Mike |
| How are the misclassifications of the best performing model distributed? | Kayode/Bhavyai |
| Discussion | All |

Note on the last section: “How are the misclassifications of the best performing model distributed?”, you (meaning both Kayode and Bhavyai) would each need to manually go over 200 entries and label why it is misclassified as per the project template. You can find it here on github for the manual annotation entries: <https://github.com/meng-ucalgary/ensf-612-project/blob/main/project_data/manual_classification/model%20vs%20manual%20work%20distribution.xlsx>

I initially wanted to assign 50 entries to myself but I am bogged down trying to figure out porting our code from sklearn into sparkml and so I am a bit tied up. I will try to contribute to the 50 entries if I have time, but I can’t guarantee so I put two columns right now listing out who is doing which entry and another column is which I would take on if I have time after the sparkml stuff.

For the report sections plus the manual annotations, please submit this no later than December 12 afternoon! We will setup a meeting on December 11 to finalize everything for the report.