11/1/21

- Extended Literature Survey performed
- Dealt with existing inconsistencies in data

18/1/21

- Prepared presentation for review 1
- Decided upon deliverables and tasks for phase 2 of capstone project

26/1/21

- **Review 1 capstone**
- Panel discussion about probable models to be used to predict the pandemic spread

1/2/21

- Discussed about several models which are fit to be used to predict future cases
- Implementation of first model, Linear Regression is begun
- Data is obtained from districts.csv to train the model

8/2/21

- Linear regression is implemented initially. On concluding that the data does not fit one least squares line, decided to implement Piecewise Linear Regression
- Piecewise Linear regression provided high accuracy on testing and comparing with recorded confirmed cases
- Challenge faced during the process was selection of optimal number of segments

15/2/21

- Preparation for review 2 is begun
- Further efforts to calculate the optimal number of segments
- Due to computational limitations and enormous size of dataset which can accommodate a large number of optimal segments, optimal number of segments could not be computed
- Decided to set the optimal number of segments to 3 on viewing the distribution of confirmed cases in the training data and owing to the high accuracy obtained

25/2/21

- **Review 2 capstone**
- Model implementation is presented to the panel
- Discussion regarding the optimal number of segments

7/3/21

- Other models to predict the cases are discussed
- Decided upon choosing 3 other models: Hierarchical Bayesian Model, Simple Exponential Smoothing and XGBoost

15/3/21

- Training and testing of data on the new models performed
- Encountered different issues in all three models such as selection of optimal values of parameters to be passed to the models
- Obtained the guidance of mentor with respect to the issues faced while training the models
- Presentation for review 3 is prepared

30/3/21

- **Review 3 capstone**
- Presented the implementation and results of the 3 new models
- Panel pointed out issues with the models and provided time to improve the models before the final ESA

4/4/21

- Simple exponential smoothing is modified so as to predict cases for more than one day
- Hierarchical Bayesian model presents output as a range of values in lognormal form, hence the mean of the anti-log of the values is used
- Optimal number of n_estimators for XGBoost is found out to be 1000 from gridSearchCV

11/4/21

- Report work is begun
- Other issues with the models are rectified and hence accuracy of all 3 models is improved

18/4/21

- Preparation for final meet with mentor scheduled on 28th April
- Report is completed
- Paper in IEEE format is begun
- All models are finalized and are working up to expectations

28/4/21

- Final meet with Prof. Nitin V Pujari, mentor where all the models are presented
- Accuracies and implementation of each model is discussed
- After the meet, preparation for final ESA scheduled on 4th May is begun
- Completed IEEE paper, video, poster, website for visualization of prediction using each model and presentation for final ESA

4/5/21

- **Final ESA for capstone project phase 2**
- All the four models are presented
- Previous issues which were pointed out by the panel are rectified except optimal number of segments for Piecewise Linear Regression which forms a part of future work