RTS Preventable Accidents Identification - Mid Term Status Report

Team Members: Xiaoran Li, Weiran Lin, Meiying Chen, Weinan Hu
Sponsor: Wray Robert, Swift Adam, Dobson Christopher D.

1. Vision and Goals

1) Vision:
   Reduce future preventable accidents and improve safety.
2) Goals:
   - Identify potential causes of preventable accidents.
   - Deliver reusable models to predict future preventable accidents.

2. Milestones and Status

	Content	Anticipated Date	Actual Date(completed)
Milestone 1	Create Project Charter	10/09	09/28
Milestone 2	Data gathering and merging	10/16	10/14
Milestone 3	Data cleaning and explanatory analysis	10/30	10/25
Milestone 4	Feature engineering	11/13	11/13
Milestone 5	Develop predictive models	11/27	
Milestone 6	Conduct causal analysis	12/04	
Milestone 7	Complete final presentation and delivered result	12/09	

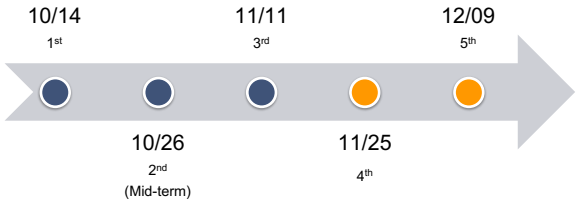
3. Progress and Completed Deliverables

1) 09/21: Project Assigned
2) 09/28: Presented our 1st version project charter to sponsor
3) 10/14: 1st presentation.
   Signed the project charter;
   Presented progress on timeline analysis, geoinformation analysis and deep learning method.
4) 10/26: Midterm Presentation(2nd Presentation).
   Presented overall progress, hotspot map of geoinformation, lag variables definition& visualization and CNN architecture.
5) 11/11: 3rd presentation. Presented progress on non-parametric tests on lag variables, feature selection and feature engineering.
6) 11/13: Delivered the feature engineering result to sponsor.

4. Upcoming Presentations and Deliverables

This section combines planning for upcoming deliverables and upcoming meetings and presentations.
Bi-Weekly meeting timeline for team 7 and sponsors:

1) 11/25: 4th Presentation. Present the predictive model results(both traditional and deep learning).
2) 11/27: Deliver models (both traditional and deep learning) to sponsor.
3) 12/04: Deliver the causal analysis results.
4) 12/09: Final (5th) Presentation. Present the overall results.



5. Risks and Opportunities

1) Imbalance dataset lead to inaccurate result.
   Possible Solutions: resampling, use the right evaluation metrics
2) The relevance of all features and accidents is low.
   Possible Solutions: try different feature selection methods like feature EDA, statistical tests, ridge/lasso regression or feature importance.
3) Small data set may lead to overfitting:
   Possible Solutions: Early stopping, Regularization

6. Baseline-to-Actual comparison

The schedule comparison is shown above, and time budget and human resource comparison is shown below

	Xiaoran Li	Weiran Lin	Melissa Chen	Weinan Hu	Sponsor
09/21/2020 -09/27/2020	3 hours (completed)	3 hours (completed)	3 hours (completed)	3 hours (completed)	0 hours
09/28/2020 – 10/04/2020	15 hours (completed)	15 hours (completed)	15 hours (completed)	15 hours (completed)	2 hours (completed)
10/05/2020 – 10/16/2020	15 hours (completed)	15 hours (completed)	15 hours (completed)	15 hours (completed)	3 hours (completed)
10/17/2020 – 10/30/2020	15 hours (completed)	15 hours (completed)	15 hours (completed)	15 hours (completed)	3 hours (completed)
10/31/2020 – 11/13/2020	15 hours (completed)	15 hours (completed)	15 hours (completed)	15 hours (completed)	3 hours (completed)
11/14/2020 – 11/27/2020	15 hours	15 hours	15 hours	15 hours	3 hours
11/15/2020 – 12/04/2020	15 hours	15 hours	15 hours	15 hours	3 hours
12/05/2020 – 12/15/2020	15 hours	15 hours	15 hours	15 hours	3 hours

7. Documentation Links

Box(Data, Dashboard, Code, Graph, Video):
https://rochester.app.box.com/folder/123094628921
Google Docs(Internal Meeting Recording, Internal Meeting Recording):
https://docs.google.com/document/d/1ZT1LQbMsVVIlK2V7FMvrxCDug6_Nfvp12YW89n6ILrA/edit#heading=h.twcuiscl0ogf

8. Outlook

Currently, we have sufficient data and resource. We have confidence to finish the project on schedule within budget. We hope we can achieve the expectations of our sponsors.