WEEKLY STATUS REPORT

REPORT DATE	TRACK NAME	PREPARED BY
Oct 10th, 2021	IntegrityM (Data wrangling and Visualization)	SeungHeon Han, Yihang Zhao, Yuwen Luo

STATUS SUMMARY

After the past few weeks of cleaning and manipulating dataset Medicare Part B, we are currently in the step of modeling and visualization. However, the computing power of our laptops keeps interfering with our modeling processes, so we ask our clients for AWS services. Besides that, our clients will provide us with a new "IELE Plus" dataset, which could give us more fraud labels on fraud labeling processes.

PROJECT OVERVIEW

MAIN TASK	SUB TASK	% DONE	DUE DATE	ASSIGNED TO	NOTES
Visualize outliers in datasets	Part B	100%	Oct 10th, 2021	SeungHeon Han	
	Part D	100%	Oct 10th, 2021	Yihang Zhao	
	DMEPOS	100%	Oct 10th, 2021	Yuwen Luo	
Transformation of outliers in the dataset	Log Transformation	100%	Oct 10th, 2021	Yihang Zhao	Lots of NaN values are created
	Square Root Transformation	100%	Oct 10th, 2021	Yihang Zhao	Lots of NaN values are created
Model Fitting	Logistic Regression	50%	Oct 10th, 2021	Yihang Zhao	

	KNN	10%	Oct 10th, 2021	Yihang Zhao	
	EBM	30%	Oct 10th, 2021	SeungHeon Han	GridSearch Failure
	XGBoost	50%	Oct 10th, 2021	Yuwen Luo	GridSearch Failure
	Random Forest	50%	Oct 10th, 2021	Yuwen Luo	GridSearch Failure
Over-Sampling & Under-Sampling Part B using SMOTE skills	Creating more fraud labels in dataset	100%	Oct 10th, 2021	SeungHeon Han	

RISK AND ISSUE HISTORY

ISSUE	ASSIGNED TO	DATE
We have not yet received additional data (LEIE Plus) from our clients - the clients are supposed to give us the new dataset because the data is not publicly available.	Yihang Zhao, Yuwen Luo, SeungHeon Han	Oct 10th, 2021
Due to the large size of data, we have a memory usage problem when training models including a grid search and model selection (KNN, RF, XGB, EBM).	Yihang Zhao, Yuwen Luo, SeungHeon Han	Oct 10th, 2021
AWS credits have not been provided from the clients. We have been still waiting for their response.	Yihang Zhao, Yuwen Luo, SeungHeon Han	Oct 10th, 2021

CONCLUSIONS & OUTLOOK

We will fit more machine learning models after clients provide us some cloud computing service. Furthermore, our model could improve after clients provide us with 'IELE PLUS' dataset.