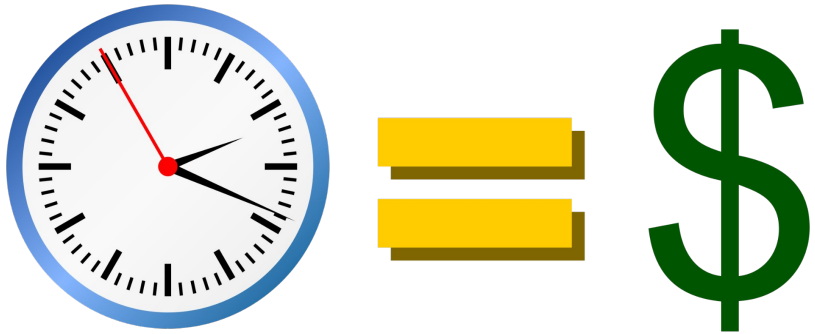
MileApp Online Assessment: Task Completion Time Prediction

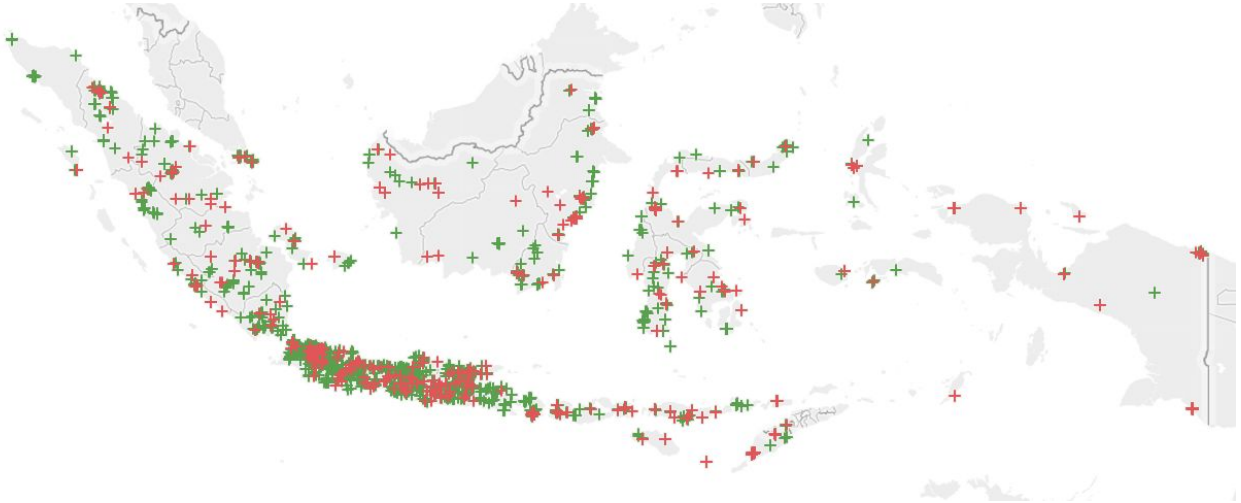
Abdullah Hadi

Idea



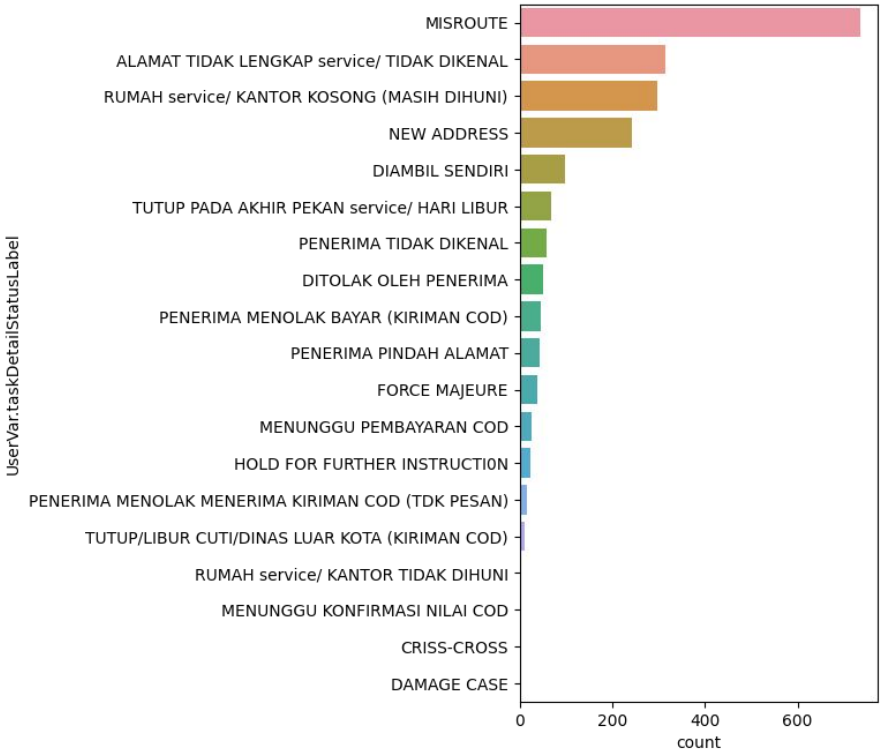
Time is money, and finding out an estimate on when a certain delivery time can be finish can open possibility for our business to use that time more efficiently, like routing more delivery. So with that idea and the data that has been given i purpose and implemented a machine learning model that predict the time needed for a delivery task to be completed. Before we jump on how the models perform, in the next couple of slide i will show some insight and understanding of the data that i found important, others finding can be seen at the notebooks

Insight and Understanding 1: Point task status label distribution



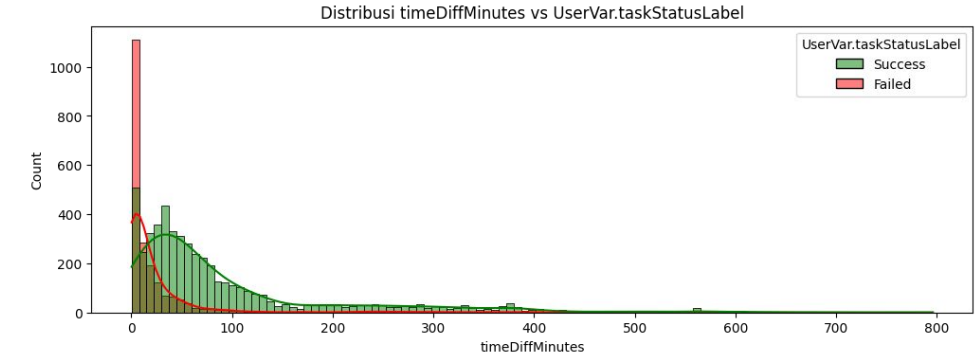
This map scatterplot was created using Tableau, from here we can see that Java island has the most dense delivery task than other island

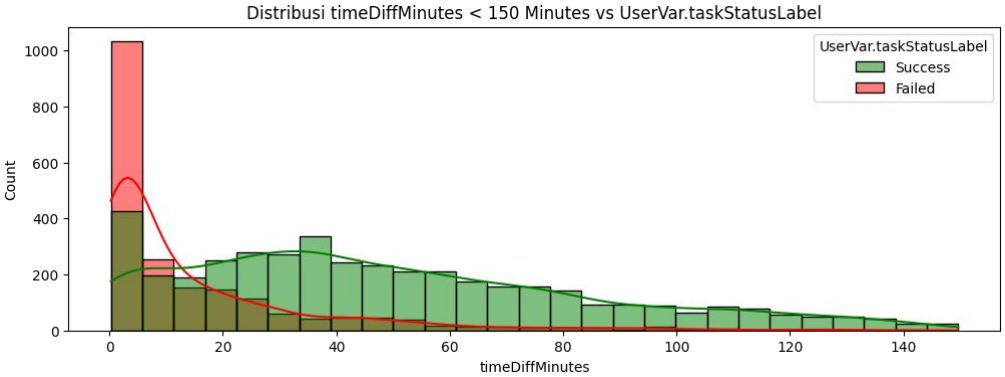
Insight and Understanding 2: Top problem for failed task



misroute is the highest problem for a task to be failed, damage case and criss-cross only occurred in one task, its not showing cause missroute and other problems has a greater number count

Insight and Understanding 3: does time to complete task impact the task to fail?





from the chart we can say that a lower time task completion has a higher chance for a task to fail

Machine Learning Part

When we build our model we use the following column/feature

- 'taskCreatedTime'
- 'cod.amount',
- 'UserVar.branch_dest',
- 'UserVar.receiver_city',
- 'UserVar.weight',
- 'UserVar.branch_origin',
- 'isCOD',

It will predict the feature 'timeDiffMinutes'

For the model type we use treebase ensemble model and also a ANN. the metric that we will use is r squared, MSE, and RMSE.

We also use a AutoML library called FALM

Machine Learning Results

Model	R Squared	MSE	RMSE
Random Forest	0.83	1752.98	41.86
FLAML : xgb_limitdepth	0.81	1910.69	43.71
ANN	0.54	4792.44	69.22

The best model to chose is random forest

Visualizing Model Prediction