ML VAPING EFFECT PROJECT WEEK 8 REPORT

Prepared By: Ng Jun Kiat (u7338876)

INTRODUCTION

1. This report describes the changed made to the model in Week 7, and the possible follow-up actions.

CHANGES TO MODEL

2. Multi-variate Random Forest (RF) to Separate Single-variate RFs

   The model now predicts Quality Adjusted Life Years (QALYs) and Health System Costs in two separate Random Forests. The results for predicting Health System Costs appear to be much better using this approach (elaborated further in the **Results** section).

3. Up-sampling instead of Duplicating

   Gaussian noise is added to variables in addition to duplicating the dataset. [To clarify] Would we still use Gamma distribution for Health System Cost if the mean for this variable is negative?

4. Using Average Age instead of Age Group

   This change is to more accurately reflect the different ages of people. In addition, Gaussian noise is also applied on *'average_age'* variable. [To clarify] Do we use normal distribution for *'average_age'* or log normal distribution?

5. Change Metric from MSE to Mean Absolute Percentage Error (MAPE).

MAPE is used to better reflect the proportion of error to the size of the variable. The formula for MAPE is as follows:

$$M = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right|$$

M = mean absolute percentage error
n = number of times the summation iteration happens
A_t = actual value
F_t = forecast value

RESULTS

6. Compared to the previous model, Health System Cost prediction appears to be slightly better. **Figure 1** shows the results for QALY and Health System Cost predictions.

Figure 1: Results Table

	n_trees	max_depth	MAPE/%
QALY	50	15	20.1%
Health System Cost	50	20	153%

TO-DO

7. Considering my presentation is in Week 12, one more week could be spent on improving the model (Week 9), and I will aim to produce some results with data from the vaping research paper by Week 10. Possible ways to improve the model are:

- Normalise variables
- Research more on up-sampling methods
- Try different model