THE UNIVERSITY OF HONG KONG

MSc in E-Commerce and Internet Computing

ECOM7126: Machine Learning for Business and E-Commerce
(2024-25)

Assignment 1 – Medical Insurance Claims

AXE Medical is a medium size insurance company in Hong Kong wanting to use machine learning to predict health insurance claims using past claim data the company has collected over the past few years.

The dataset consists of 1,300 former successful claim cases, in .csv format and the following is the data dictionary of the dataset:

age	age of the insured
sex	gender of the insured
home	home district of the insured: Hong Kong / Kowloon / South NT / North NT
bmi	body mass index (weight in kg / height in m^2)
children	number of dependent children
smoker	yes/no
drinking	no (non-drinker) / occasional / frequent
past claims	amount in HKD of total past insurance claims

Note: NT = New Territory

Deliverable:

1. A report in PDF format.
2. The Colab notebook (Python programs with comments and notes) that you use to produce your results in .ipynb (Colab notebook format).

You should include the following in your report:

1. An account of what you have done in investigating and transforming your data and why.
2. Use **Linear Regression** to build a prediction engine. (Hint: You may not need to include all features in the dataset. Justify your decision.) DO NOT use other models for this assignment.
3. Any observation you may have in your project to the management of the company.
4. If you have used any AI tools in completing this assignment, you are required to clearly state how and what tools you have used.

Datasets provided: AXE_dataset.csv