Task 6.2

Michael Christopher – s224830467

1a. missing values

* Categorical features -> imputed with mode
* Continuous features -> imputed with median

1b. Split

* Stratified train/test split with ratio 80/20 to preserve class proportions

1c. Feature types

* Continuous features: identified by numeric dtype not originating from encoded categorical
* Categorical features: original object/string columns

1d. Encoding

* Label encoding categorical features to continuous and map target to 0, 1, and 2
* Standardization applied to features

1e. Label distribution

* A screenshot of a computer

  AI-generated content may be incorrect.Imbalance

A white paper with black text

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

A white paper with black text

AI-generated content may be incorrect.A screenshot of a document

AI-generated content may be incorrect.A screenshot of a document

AI-generated content may be incorrect.