Assessment Question 1

Code structure and workflow

* modelconfig file has redundant range of parameters, it is again later defined in the workflow files
* functions on the modelconfig files hard coded various parameters
* addition of unit test could be beneficial
* currently data source based off csv file, can be extended to connecting live database and perform updated training when new data refreshed

Technique Selection

* To start I would suggest to use simpler methods such as logistic regression
* Data dropna default is to drop any records with any empty cells, this might be too aggressive – can consider filling in missing values

Modelling approach

* Feature selection and engineering is missing. Features can be correlated and thus should processed before putting into model.
* Normalization for features
* In both approaches, metrics were based from cross validation – we can consider splitting into train/test prior, and use the test set as the final evaluation

Evaluation Approach

* Code is only evaluating based on accuracy of model. But we want to consider both precision and recall. If client is keen to reduce insult to a value to 0, then we want to weight the recall metrics heavier in order to capture all frauds even at a cost of false positive. The tradeoff decision should be further discussed with client