✓ **Congratulations! You passed!**
TO PASS 80% or higher

Keep Learning

GRADE
100%

Data Transforms and Feature Engineering: End of Module Quiz

LATEST SUBMISSION GRADE

100%

1. The *.fit_transform* method corresponds to which scikit-learn interface(s)? Choose one answer.

   1 / 1 point

   ○ Transformer, Estimator, Predictor

   ⦿ Transformer, Estimator

   ○ Estimator, Predictor

   ○ Transformer

   ○ Transformer, Predictor

   ✓ **Correct**
   Correct!

2. True/False. A principal reason for emphasizing the use of pipelines in the AI workflow is to have a consistent platform that enables comparison of many variants of the workflow.

   1 / 1 point

   ⦿ True

   ○ False

   ✓ **Correct**
   Correct!

3. Which of the following statements describes the **best strategy** to address class imbalance?

   1 / 1 point

   ○ If there is a lot of data just use under-sampling otherwise use outlier detection algorithms.

   ○ Determine the best variant of SMOTE by comparisons and use it.

   ○ Continue to collect data until you have balanced classes.

   ○ Use an outlier detection algorithm or SVM instead of a re-sampling technique.

   ⦿ Compare re-sampling approaches to a baseline and to detection algorithms.

   ✓ **Correct**
   Correct!

4. Which of the following statements is **not** a feature of the package imbalanced-learn imbalance?

   1 / 1 point

   ○ Has a suite of over and under-sampling methods implemented

   ○ Works with TensorFlow

   ○ Has a number of tutorials to work from

   ⦿ Has outlier detection algorithm packaged as part of library

   ○ Works with scikit-learn pipelines

   ✓ **Correct**
   Correct!

5. Which of the following statements does **not** describe valid use case for dimensionality reduction?

   1 / 1 point

   ○ Principal components analysis to process images used in classification.

   ○ Non-negative matrix factorization to resolve topics from a corpus of words.

   ○ t-distributed stochastic neighbor embedding to visualize the results of a clustering algorithm.

   ○ Using an ANOVA to select a subset of features

   ⦿ Down-sampling of the majority class

6. True/False. tSNE is a reasonable alternative to PCA because it describes a wider variety of structures. However, it is still recommended to use another dimensionality reduction method, like PCA if the number of features is very high. `1 / 1 point`

- ◉ True
- ○ False

7. Which statement best describes why visualization of topics can have an impact on the business opportunity? `1 / 1 point`

- ◉ Because sharing with domain experts might enable topic-specific feature engineering
- ○ Because visual inspection can help choose the number of topics
- ○ Because we are able to see the top words with each topic
- ○ Because we are able to see the relative importance of each topic across the corpus
- ○ Because domain experts can visually inspect the validity of the topics

True

False