Follow-along instructions: Build and cross-validate a random forest model with Python

Accessing and utilizing resources in this section

While watching the video that follows this reading, you may find it helpful to track the instructor's progress by following along in your own Jupyter notebook. To do so, open the annotated follow-along guide for the video. The content in this notebook is identical to the content shown in this lesson's instructional video. In addition to that content, you'll find additional information throughout the notebook. That information is provided to explain the purpose of each concept covered, why the code is written in a certain way, and tips for running the code.

Steps to complete before each video:

1. Read this page of instructions.
2. Open the Annotated follow-along guide: Build and cross-validate a random forest model with Python, which contains a version of the same notebook the instructor will use in the video.
3. Follow along with the instructor as they go over the code in the notebook.
4. Learn from the instructor and practice running the code in your notebook.

Data dictionary

In this lesson's video, your notebook will include the **bank_customer_churn_data.csv** dataset. This dataset represents bank churn data from a bank in Europe. "Churn" is a term that refers to customers leaving and taking their business elsewhere.

The dataset contains:

10,000 rows – each row is a unique customer of the bank

14 columns:

Column name	Type	Description
RowNumber	int	Row numbers from 1 to 10,000
CustomerId	int	Customer's unique ID assigned by bank
Surname	str	Customer's last name
CreditScore	int	Customer's credit score. This number can range from 300 to 850.
Geography	str	Customer's country of residence
Gender*	str	Categorical indicator
Age	int	Customer's age (years)
Tenure	int	Number of years customer has been with bank
Balance	float	Customer's bank balance (Euros)
NumOfProducts	int	Number of products the customer has with the bank
HasCrCard	bool	Indicates whether the customer has a credit card with the bank
IsActiveMember	bool	Indicates whether the customer is considered active
EstimatedSalary	float	Customer's estimated annual salary (Euros)
Exited	bool	Indicates whether the customer churned (left the bank)

*Many societies now understand that gender is not binary and there is fluidity in the category across people and time, so creating categories that allow for that leads to more accurate data collection processes.