Simulation Storyboard
Create and run AI models

1. Now that you've set up your data set in cloud storage, it is time to build the AI models. Select **New asset** to start creating your AI model.

    AI models
    Creating artificial intelligence models used to be an incredibly complex and difficult task. But today you can show off to your colleagues how IBM Watson Studio does this for you—automatically!

2. On the left side, there is a **Tool type** list. Select the **Automated builders** tool. **AutoAI** is the only Automated builder displayed. With AutoAI, you will be able to quickly set up and run AI models using your data to train and test the models. Select **AutoAI**.

3. Creating and testing AI models is called experimenting. Now it's time to create your AutoAI experiment. Type "Loan Risk" in the **Name** field to name your model and press **Enter**.

4. In the **Define configuration** section, you'll see a machine learning service isn't associated with your project yet. Let's take care of that now. Select the **Associate a Machine Learning service instance** link.

5. On the **Associate service** page, you'll see all the services you've created that can be associated with your experiment. Right now, there's only one service displayed: **Machine Learning-Risk_Fraud**. Select the checkbox next to the service and then select **Associate**.

6. Now that you've associated your machine learning service, select **Reload** to refresh the page.

7. You can see Machine Learning-Risk_Fraud displayed in the **Watson Machine Learning Service Instance** field now. You're ready to create your AutoAI experiment. Select **Create**.

8. It's time to add your data source. Select **Select from project** to find the data source you added previously.

    Your experiment
    You promised the bank you'd run several different machine learning algorithms competitively. You'd train them all on the same partial set of **historic data** about people who took out loans and then paid them back or defaulted on them. Then you'd test how well they performed, based on what they'd learned, by feeding them identical sets of **new data**, without telling them who defaulted. Each algorithm would make predictions about everyone in the new data set, classifying people as **Risk** or **No Risk**.

9. On the left, you'll see **Categories** listed. Select **Data asset** to reveal your data assets.

10. You can see your data asset. Select the check box next to: **german_credit_data_training.csv**.

11. On the right, information about the asset is displayed including: name, asset type, size, and when it was modified and created. Select **Select asset**.

12. Now it's time to configure the details for your experiment. The first question that displays asks if you want to create a time series forecast. For this experiment, you don't want to do that, so select **No**.

13. The next question that displays asks what you want to predict. This is asking what column from your data set you want the AI model to predict. Select **Select prediction columns** to open the drop-down list.

14. Next to **Risk** are the letters "STR". This indicates that the data type in the **Risk** column is a String. Strings are a sequence of letters, digits, punctuation, and so on. In this case, it's text data. Since you're trying to *predict which individuals are good risks for loans*, select **Risk.**

15. The **PREDICTION TYPE** that Watson AutoAI selected is Binary Classification. You're trying to determine whether an individual presents a **Risk** or **No Risk**, which is a classification with only two options, so leave the prediction type as **Binary Classification**. Select the **Next arrow** to continue.
16. By default, **Accuracy & run time** is selected in the **OPTIMIZED FOR** field. This means when Watson is evaluating which algorithm is best it will optimize by balancing accuracy and speed. Select **Experiment settings** to make further changes.
17. This is the **General** tab. It shows what you selected on the previous page. **Binary classification** is the prediction type, **No Risk** is the positive class, and **Accuracy** is the optimize metric. Select the **Next arrow** to continue.
    ### Competitive algorithms
    Some folks watching you had expected you to create mathematical algorithms that could analyze data and predict credit risk. It's time to astonish them. How? By using Watson Studios to line them up quickly.
18. Scroll down to view the **Algorithms to include** section, notice **Gradient Boosting Classifier** is not selected. Select the check box next to **Gradient Boosting Classifier**.
19. Scroll down to view the **Algorithms to use** section. Watson AutoAI lets you test up to four algorithms. Select the **Next arrow** to continue.
20. By default, Watson will select the best two algorithms to test on the data. Select **Data source** from the left menu.
21. Scroll down to view the **Training and holdout method** section.
    The **Training data split** slider lets you choose how much of the data set to use for training and how much to use for testing. By default, 90% of the data set is used for training and 10% of the data set is used to test how the AI model is performing. Hover over the underlined words to see more about each field in this section.
22. Below the Training data split is the **Select features to include** section. Here, you can select which columns (or features) in the data set to use for training and testing. There are three pages of data. Select the **forward arrow** to page through the columns in the data set.
23. All the features in the data set are available except **Risk** because that's what is being tested. By default, all the features are selected. Select the check box next to **Telephone** to deselect it and not use this feature in the experiment.
24. Select **Save settings** to confirm all your choices.
25. Everything is set up! Select **Run experiment** to start the experiment.
    ### Launch time!
    You're about to launch the experiment. Your models will ingest the historic training data, then run the new data and make their predictions. Like people watching a horse race, everyone leans in to see what happens.
26. The Relationship map shows by default. Select the **Swap view** link under **Progress map** to display the Progress map.
27. Here, the Progress map shows the two models that are being tested. Take a moment to review. Select the **Next arrow** to continue.
28. Watch as the AutoAI experiment runs. This will only take 45 seconds, but would take roughly 5 minutes when performed in the live environment.
29. Notice that each model has four pipelines (or algorithms) for a total of eight pipelines. The pipeline that's producing the most accurate predictions and is the fastest is shown with a star. Select the pipeline with the star.

30. This displays the **Pipeline details**. By default, the ROC curve is shown. Select **Confusion matrix** from the **Model viewer** on the left.
31. This displays the Confusion matrix. You will learn more about the confusion matrix and what it means next.

Conclusion: You successfully ran an AutoAI experiment in IBM Watson Studio. You determined which of the eight pipelines makes the most accurate predictions based on the data set and generated a confusion matrix.