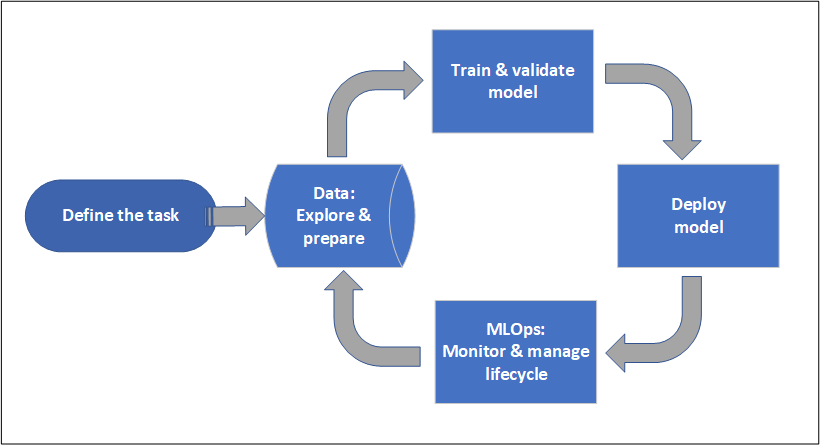
**What is Azure Machine Learning?**

Azure Machine Learning is a cloud service for accelerating and managing the machine learning project lifecycle. Machine learning professionals, data scientists, and engineers can use it in their day-to-day workflows: Train and deploy models and manage MLOps.

You can create a model in Azure Machine Learning or use a model built from an open-source platform, such as Pytorch, TensorFlow, or scikit-learn. MLOps tools help you monitor, retrain, and redeploy models.



**Who is Azure Machine Learning for?**

Azure Machine Learning is for individuals and teams implementing MLOps within their organization to bring machine learning models into production in a secure and auditable production environment.

Data scientists and ML engineers will find tools to accelerate and automate their day-to-day workflows. Application developers will find tools for integrating models into applications or services. Platform developers will find a robust set of tools, backed by durable Azure Resource Manager APIs, for building advanced ML tooling. Enterprises working in the Microsoft Azure cloud will find familiar security and role-based access control (RBAC) for infrastructure. You can set up a project to deny access to protected data and select operations.

**Task 1 : Explore the dataset**

1. Log on to the Azure Machine learning workspace from the azure portal. portal.azure.com

Graphical user interface, text

Description automatically generated

1. Upload the dataset on Azure Machine Learning Workspace

Graphical user interface, application, Word

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

1. Use the dataset\_final.csv as a dataset

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, table

Description automatically generated

Update the data type for fraud

Graphical user interface, application, table

Description automatically generated

1. Select the compute cluster provisioned as a compute. Generate the profile for the dataset

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

1. Understand the dataset through the profile.

Graphical user interface

Description automatically generated

# What is automated machine learning (AutoML)?

Automated machine learning, also referred to as automated ML or AutoML, is the process of automating the time-consuming, iterative tasks of machine learning model development. It allows data scientists, analysts, and developers to build ML models with high scale, efficiency, and productivity all while sustaining model quality. Automated ML in Azure Machine Learning is based on a breakthrough from our Microsoft Research division.

Traditional machine learning model development is resource-intensive, requiring significant domain knowledge and time to produce and compare dozens of models. With automated machine learning, you'll accelerate the time it takes to get production-ready ML models with great ease and efficiency.

**Ways to use AutoML in Azure Machine Learning**

Azure Machine Learning offers the following two experiences for working with automated ML. See the following sections to understand [feature availability in each experience](https://docs.microsoft.com/en-us/azure/machine-learning/concept-automated-ml#parity).

For code-experienced customers, [Azure Machine Learning Python SDK](https://docs.microsoft.com/en-us/python/api/overview/azure/ml/intro). Get started with [Tutorial: Use automated machine learning to predict taxi fares](https://docs.microsoft.com/en-us/azure/machine-learning/tutorial-auto-train-models).

For limited/no-code experience customers, Azure Machine Learning studio at [https://ml.azure.com](https://ml.azure.com/). Get started with these tutorials:

**Task 2 : Create your first model using AutoML**

1. Select the AutoML from the machine learning workspace.

Graphical user interface, application, Word

Description automatically generated

1. Create a new Auto Ml run

Graphical user interface, text, application, email

Description automatically generated

1. Select the dataset

Graphical user interface, text, application, email

Description automatically generated

1. Select the target column and select compute cluster as compute

Graphical user interface, text, application, email

Description automatically generated

ii. Select Accuracy as primary metric. And limit the training job time to 0.5 hours. And click save and then Finish

Graphical user interface, text, application, email

Description automatically generated

# Interpretability: model explanations in automated machine learning

To better understand your model, you can see which data features (raw or engineered) influenced the model's predictions with the model explanations dashboard.

The model explanations dashboard provides an overall analysis of the trained model along with its predictions and explanations. It also lets you drill into an individual data point and its individual feature importances. [Learn more about the explanation dashboard visualizations](https://docs.microsoft.com/en-us/azure/machine-learning/how-to-machine-learning-interpretability-aml#visualizations).

To get explanations for a particular model,

1. On the **Models** tab, select the model you want to understand.
2. Select the **Explain model** button, and provide a compute that can be used to generate the explanations.
3. Check the **Child runs** tab for the status.
4. Once complete, navigate to the **Explanations (preview)** tab which contains the explanations dashboard.

**3. Explore the ML model output**

i. When training is complete click on the experiment for model explanation.

Graphical user interface, text, application, email

Description automatically generated

ii. Explore the Data guardrails tab.

Graphical user interface, text, application, email

Description automatically generated

**Identify overfitting**

Overfitting in machine learning occurs when a model fits the training data too well, and as a result can't accurately predict on unseen test data. In other words, the model has simply memorized specific patterns and noise in the training data, but is not flexible enough to make predictions on real data.

Consider the following trained models and their corresponding train and test accuracies.

| **IDENTIFY OVERFITTING** | | |
| --- | --- | --- |
| **Model** | **Train accuracy** | **Test accuracy** |
| A | 99.9% | 95% |
| B | 87% | 87% |
| C | 99.9% | 45% |

Considering model **A**, there is a common misconception that if test accuracy on unseen data is lower than training accuracy, the model is overfitted. However, test accuracy should always be less than training accuracy, and the distinction for overfit vs. appropriately fit comes down to *how much* less accurate.

When comparing models **A** and **B**, model **A** is a better model because it has higher test accuracy, and although the test accuracy is slightly lower at 95%, it is not a significant difference that suggests overfitting is present. You wouldn't choose model **B** simply because the train and test accuracies are closer together.

Model **C** represents a clear case of overfitting; the training accuracy is very high but the test accuracy isn't anywhere near as high. This distinction is subjective, but comes from knowledge of your problem and data, and what magnitudes of error are acceptable.

**Prevent overfitting**

In the most egregious cases, an overfitted model assumes that the feature value combinations seen during training will always result in the exact same output for the target.

The best way to prevent overfitting is to follow ML best-practices including:

* Using more training data, and eliminating statistical bias
* Preventing target leakage
* Using fewer features
* **Regularization and hyperparameter optimization**
* **Model complexity limitations**
* **Cross-validation**

iii. Further explore model metrics and performances.

Graphical user interface, application

Description automatically generated

Application

Description automatically generated

Chart

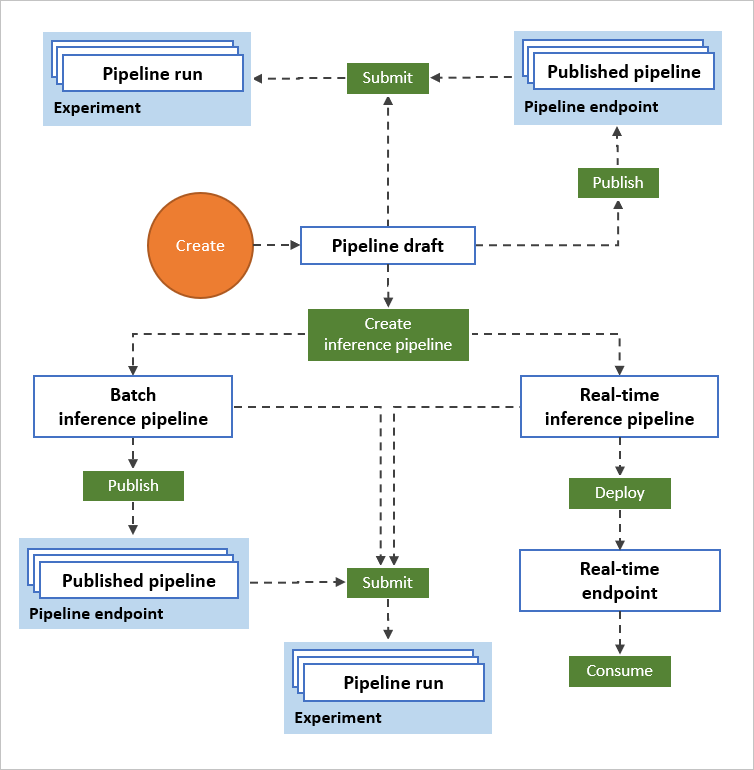
Description automatically generated with medium confidence

**Since the model clearly has a class imbalance issue we need to find a remediation and build another model. We will use Azure Machine Learning Designer for this.**

# What is Azure Machine Learning designer?

Azure Machine Learning designer is a drag-and-drop interface used to train and deploy models in Azure Machine Learning. With ML designer. Use a visual canvas to build an end-to-end machine learning workflow. Train, test, and deploy models all in the designer:

* Drag-and-drop [datasets](https://docs.microsoft.com/en-us/azure/machine-learning/concept-designer#datasets) and [modules](https://docs.microsoft.com/en-us/azure/machine-learning/concept-designer#module) onto the canvas.
* Connect the modules to create a [pipeline draft](https://docs.microsoft.com/en-us/azure/machine-learning/concept-designer#pipeline-draft).
* Submit a [pipeline run](https://docs.microsoft.com/en-us/azure/machine-learning/concept-designer#pipeline-run) using the compute resources in your Azure Machine Learning workspace.
* Convert your **training pipelines** to **inference pipelines**.
* [Publish](https://docs.microsoft.com/en-us/azure/machine-learning/concept-designer#publish) your pipelines to a REST **pipeline endpoint** to submit a new pipeline that runs with different parameters and datasets.
* Publish a **training pipeline** to reuse a single pipeline to train multiple models while changing parameters and datasets.
* Publish a **batch inference pipeline** to make predictions on new data by using a previously trained model.
* [Deploy](https://docs.microsoft.com/en-us/azure/machine-learning/concept-designer#deploy) a **real-time inference pipeline** to a real-time endpoint to make predictions on new data in real time.



# Task 4 Design a new model using ML designer

1. Select ML designer from the top menu and select new pipeline.

Graphical user interface, application, email

Description automatically generated

1. Select compute cluster as the default target

**Always connect the modules to the previous output before adding settings.**

Graphical user interface, text, application, email

Description automatically generated

1. Select the dataset

Graphical user interface, application

Description automatically generated

# SMOTE

Synthetic Minority Oversampling Technique (SMOTE) is a statistical technique for increasing the number of cases in your dataset in a balanced way. The module works by generating new instances from existing minority cases that you supply as input. This implementation of SMOTE does not change the number of majority cases.

The new instances are not just copies of existing minority cases. Instead, the algorithm takes samples of the feature space for each target class and its nearest neighbors. The algorithm then generates new examples that combine features of the target case with features of its neighbors. This approach increases the features available to each class and makes the samples more general.

SMOTE takes the entire dataset as an input, but it increases the percentage of only the minority cases. For example, suppose you have an imbalanced dataset where just 1 percent of the cases have the target value A (the minority class), and 99 percent of the cases have the value B. To increase the percentage of minority cases to twice the previous percentage, you would enter 200 for SMOTE percentage in the module's properties.

1. Search and select SMOTE to help with class imbalance

Graphical user interface, application

Description automatically generated

Diagram

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, application, Word

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

iii. Select only required columns

Graphical user interface, application

Description automatically generated

Diagram

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

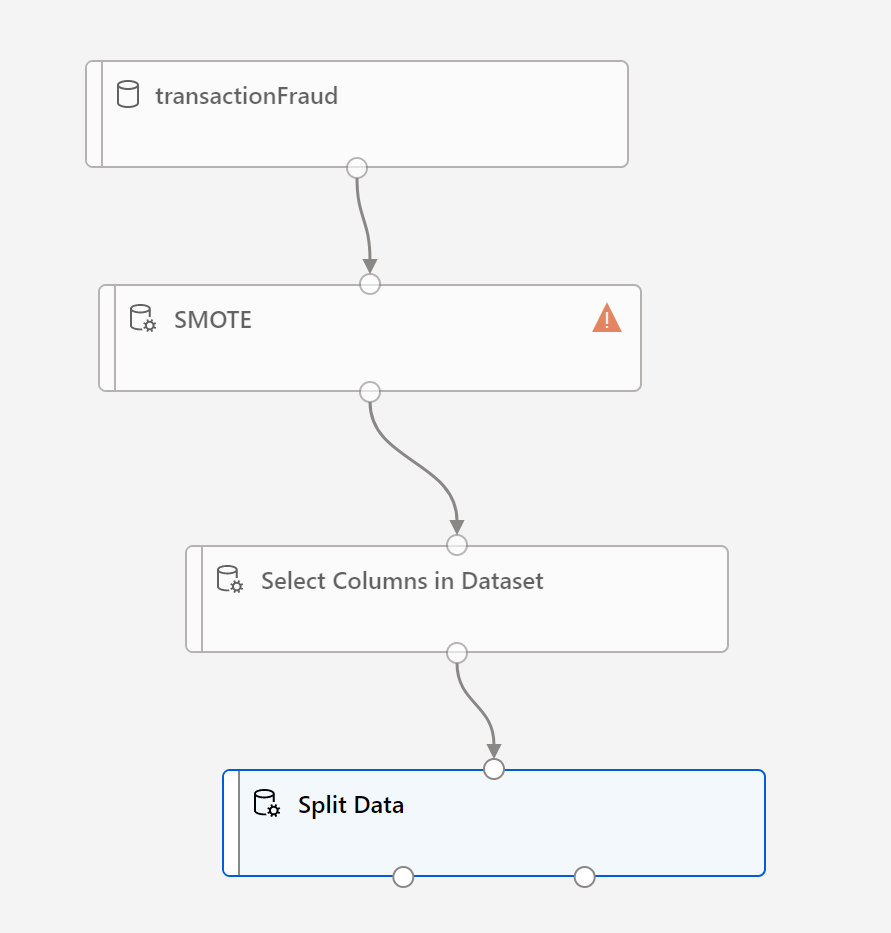
Graphical user interface, application

Description automatically generated

1. Split the data as test and train datasets

Graphical user interface, text, application, email

Description automatically generated



Graphical user interface, text, application, email

Description automatically generated

1. Select the two class boosted decision tree as the algorithm

Graphical user interface, text, application, email

Description automatically generated

Diagram

Description automatically generated

1. Add the train model module

Graphical user interface, text, application

Description automatically generated

Graphical user interface, text, application

Description automatically generated

Diagram

Description automatically generated

Graphical user interface, text, application

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

1. Add the score model to test against the test data

Graphical user interface, text, application, email

Description automatically generated

Diagram

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

1. Add the evaluate model module and submit the run

Graphical user interface, application

Description automatically generated

Diagram

Description automatically generated

Graphical user interface

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

x. Observe model training and output

Graphical user interface, application, email, website

Description automatically generated

Graphical user interface

Description automatically generated

Graphical user interface

Description automatically generated

Graphical user interface

Description automatically generated

Chart

Description automatically generated

**Summary**

A similar model is deployed by the coach for the future use, and they will share the details after participants finish building models.