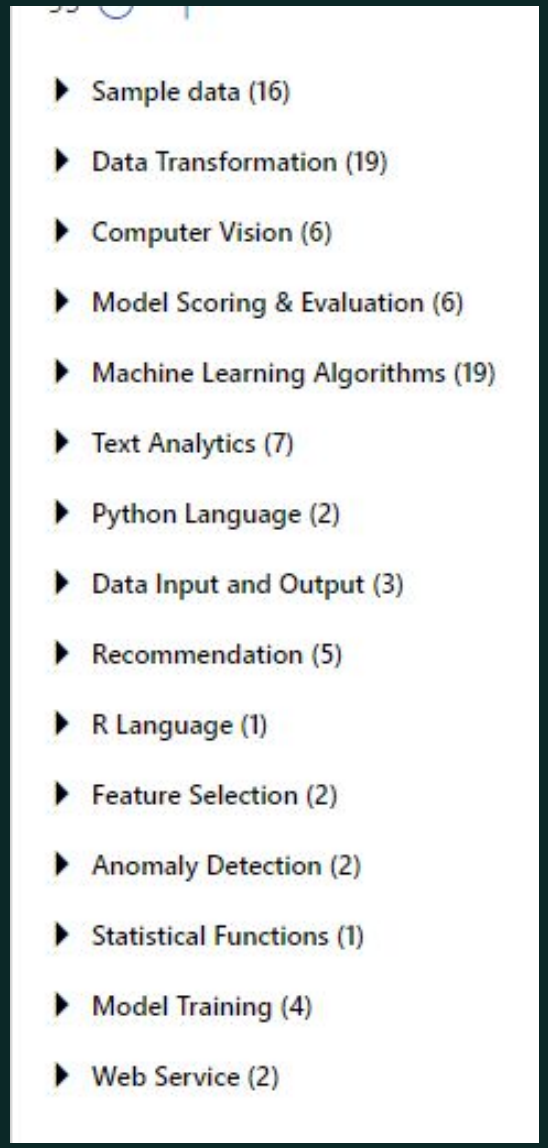
Powerful tools for social scientist in cloud computing

- Presentation by Alfred Kwadzo Torsu, MPP
  Political Science Department

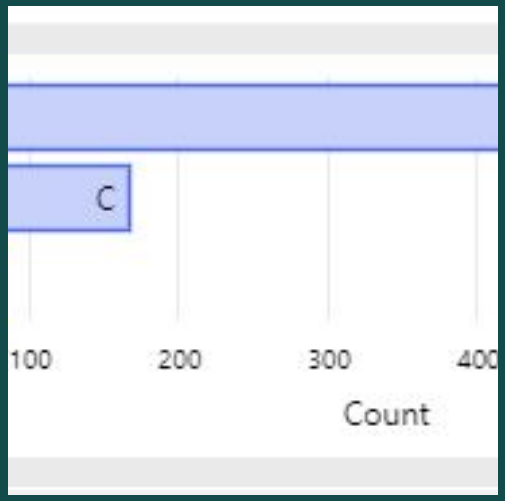
AZURE MACHINE LEARNING

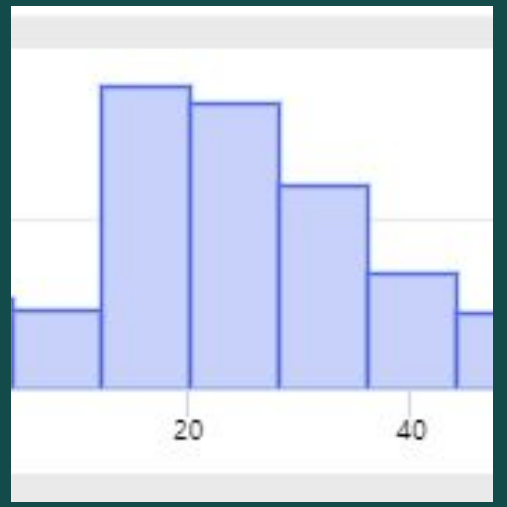
Azure Machine learning enables build machine learning pipelines that can be used to extract insights from structured and unstructured data. By using pre-built components people with low code knowledge can , you can quickly create a pipeline that performs tasks such as sentiment analysis, entity recognition, and key phrase extraction.

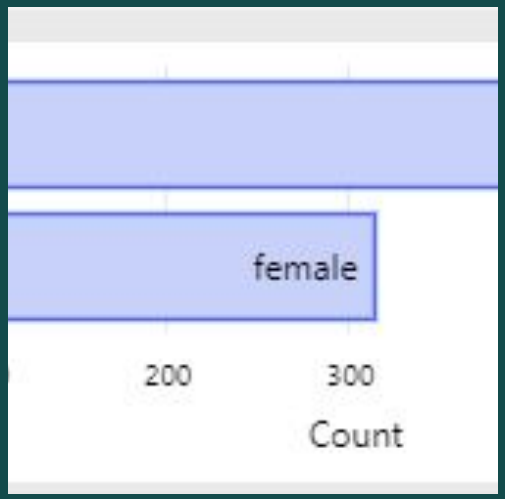


Azure Machine Learning Data Preview Function

- With Azure data preview function, users can instantly visualize inserted data and derive valuable insights without any additional actions, enabling them to make informed decisions even before utilizing machine learning models.





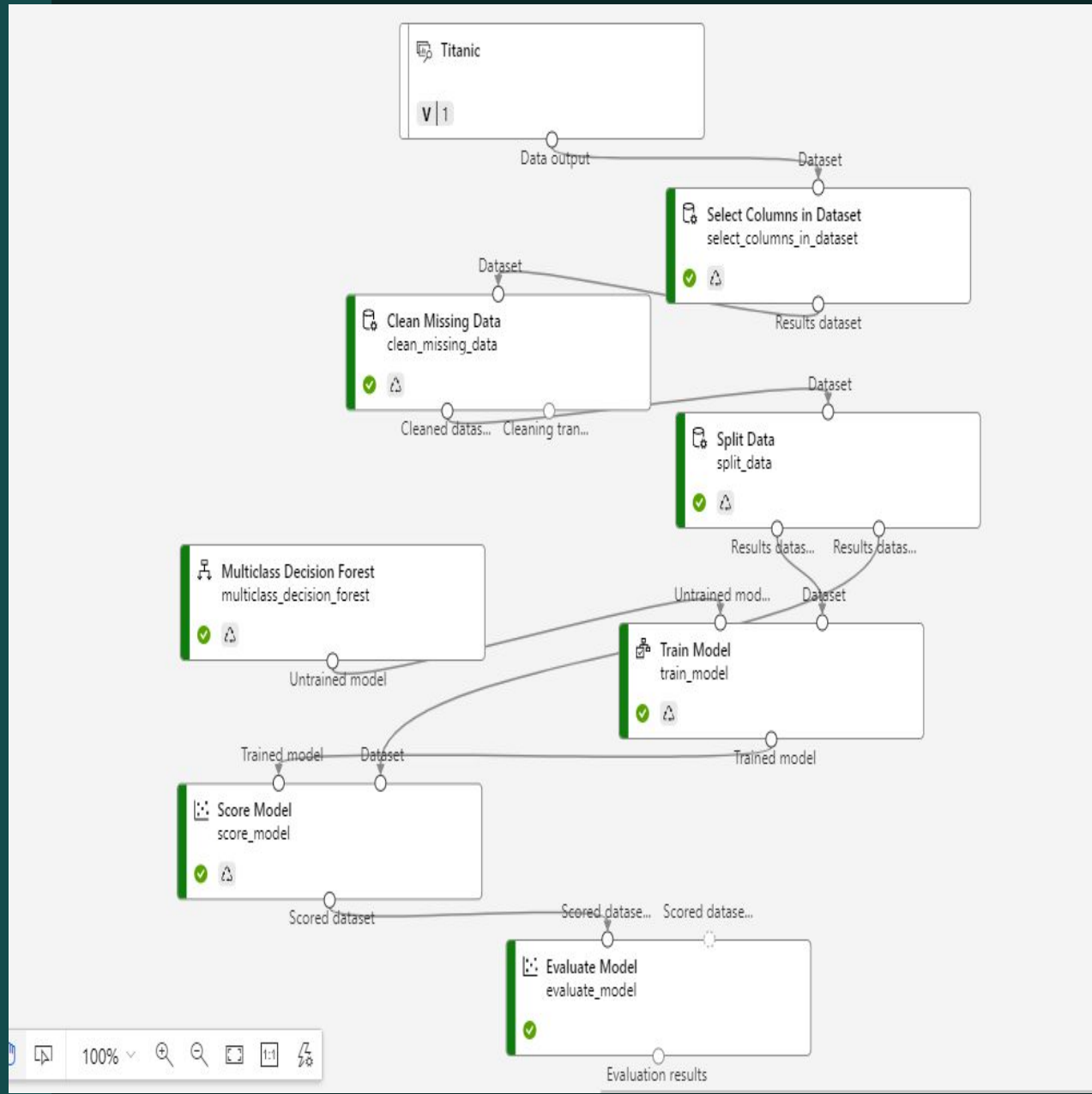


- This is what a typical Machine pipeline looks like.
- Titanic data set consist of information about people on board of the titanic. Gender, Fare, Age, Survive etc
- Select Columns in Dataset: allows you to select which columns in your data set your want to use in your model.
- Clean missing data: allows you to clean data set and replace missing data.
- Split data : this function splits data into various parts depending on the user configuration.

- Train model: This will train the model according to targets variables such as sentiment commands.
- Multiclass Decision forest allows users a model that can categorize data into multiple groups by using a collection of decision trees.

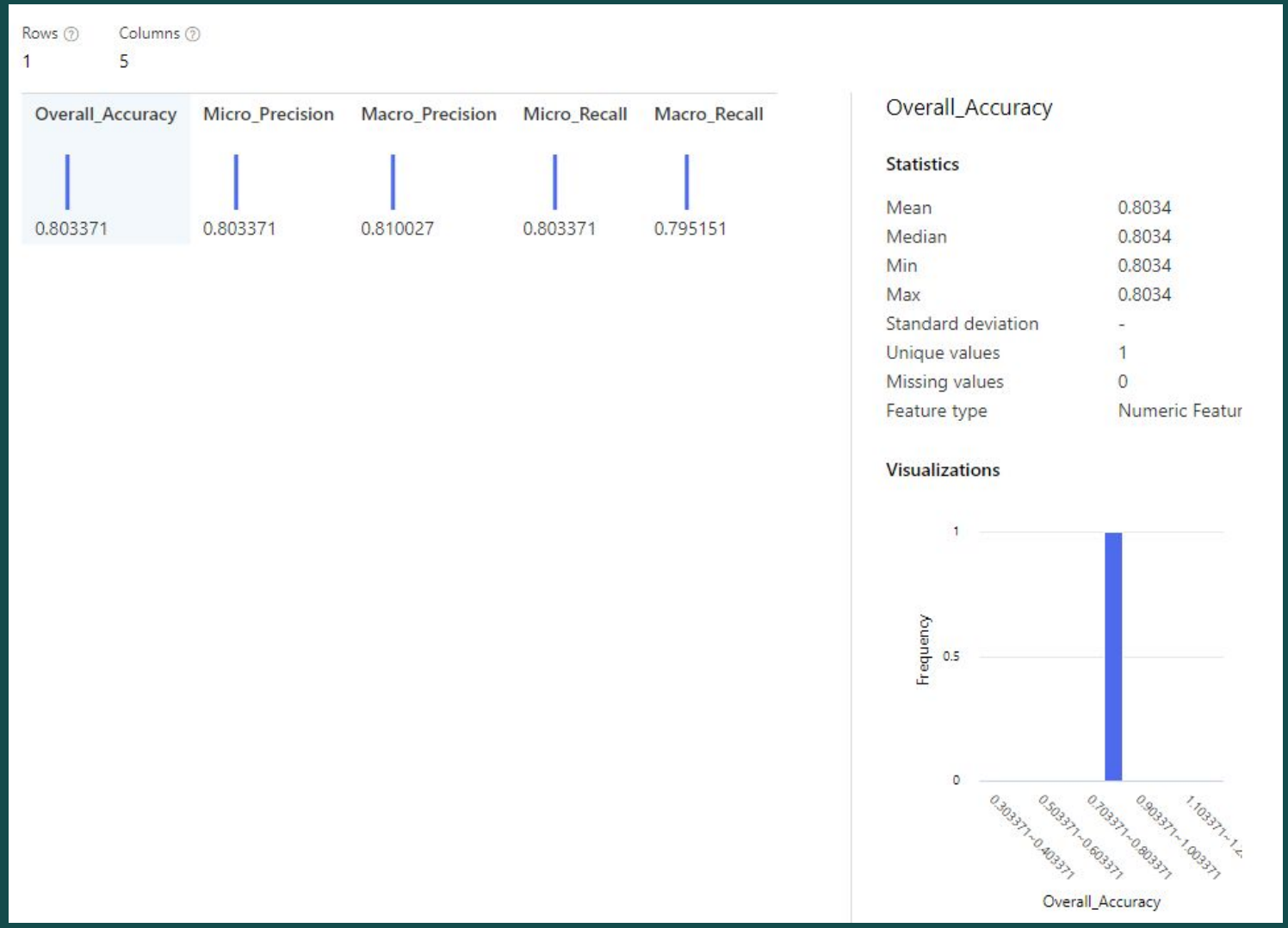
- Score module will apply a trained machine learning model to a dataset to predict the target variable without any further model training or evaluation.

- Evaluate: evaluates the performance of the model by comparing the predicted values with the actual values in the dataset.



Results

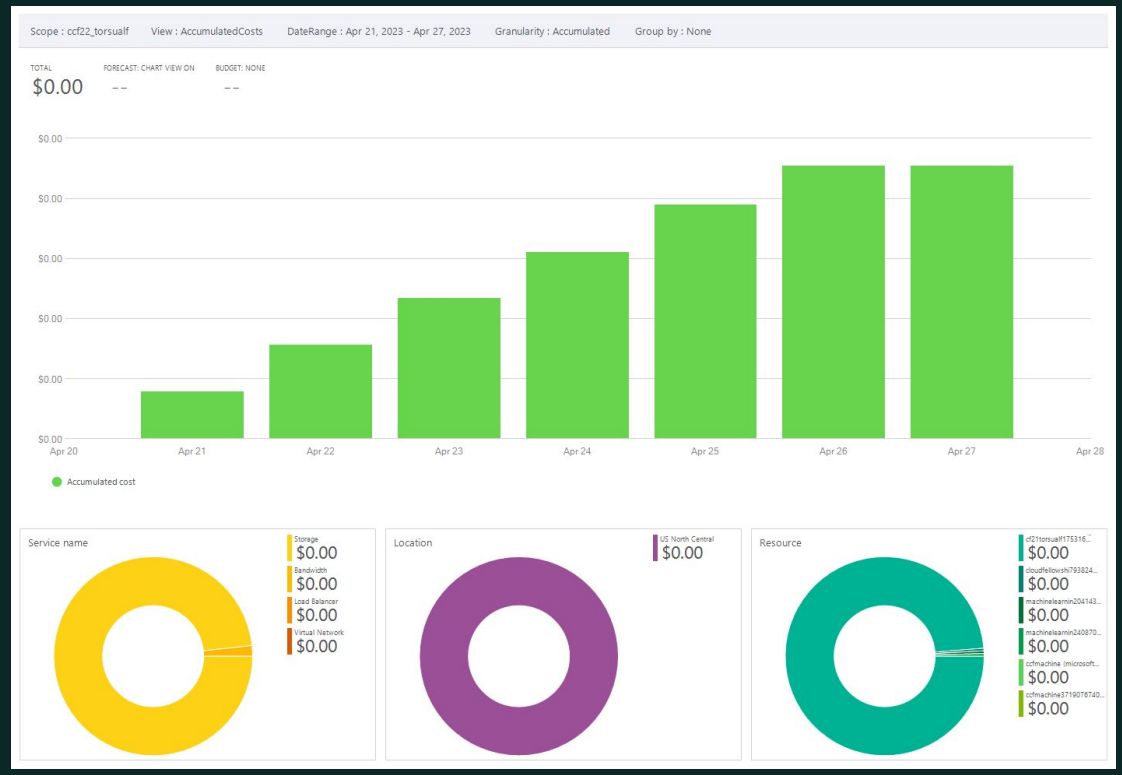
Overall Accuracy of the model is 80%. It is able to predict those who survived 80% of the time..



Cost component

- Azure Machine learning automatically creates a Storage and a VM. However, an additional computing instance or cluster is need to train and process data. These come at various cost seen in the table below.

Instance	vCPU(s)	RAM	Linux VM Price	Machine Learning Service Surcharge	Pay As You Go Total Price	1 year reserved total price	3 year reserved total price
D2 v3	2	8 GiB	$0.096/hour	$0/hour	$0.096/hour	$0.058/hour ~40% savings	$0.037/hour ~62% savings
D4 v3	4	16 GiB	$0.192/hour	$0/hour	$0.192/hour	$0.115/hour ~40% savings	$0.074/hour ~62% savings
D8 v3	8	32 GiB	$0.384/hour	$0/hour	$0.384/hour	$0.229/hour ~40% savings	$0.148/hour ~62% savings
D16 v3	16	64 GiB	$0.768/hour	$0/hour	$0.768/hour	$0.458/hour ~40% savings	$0.295/hour ~62% savings
D32 v3	32	128 GiB	$1.536/hour	$0/hour	$1.536/hour	$0.916/hour ~40% savings	$0.590/hour ~62% savings
D64 v3	64	256 GiB	$3.072/hour	$0/hour	$3.072/hour	$1.832/hour ~40% savings	$1.180/hour ~62% savings



Challenges

- Azure being a constantly evolving cloud platform, frequently introduces changes that can render even three-month tutorial obsolete, making it necessary to stay updated regularly, thus increasing the difficulty of the learning curve.