

Build an explainable AI model with AWS Sagemaker Clarify

Sarbani Maiti

AVP - Cloud Service in Magic Finserv |

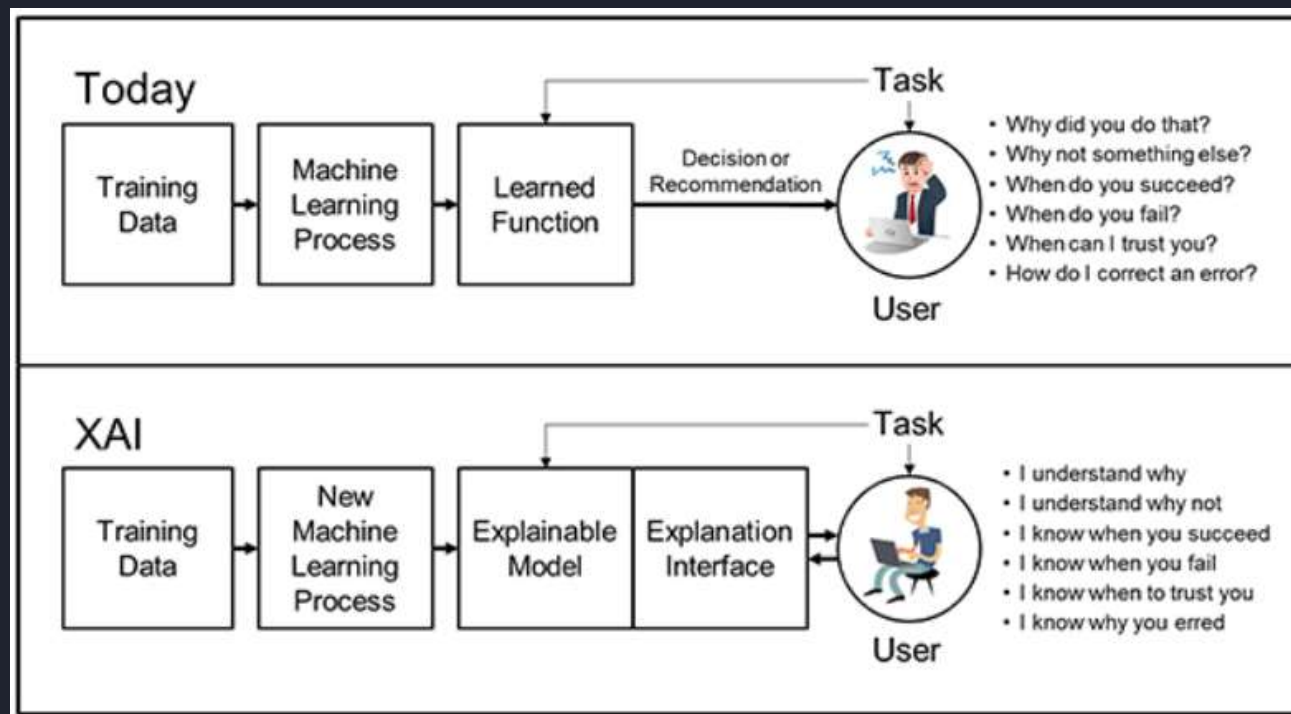
Consultant - Machine Learning | MLOps | NLP |

AWS Community Builder | Mentor

Topics

- Introduction to Explainable AI
- What is SHAPLEY
- How to build Explainable AI Model using Sagemaker Clarify
- How to build a model to explain Credit decision using Sagemaker Clarify in Sagemaker Studio.

Understand How Machines Make Decision



The diagram Source: <https://www.darpa.mil/program/explainable-artificial-intelligence>

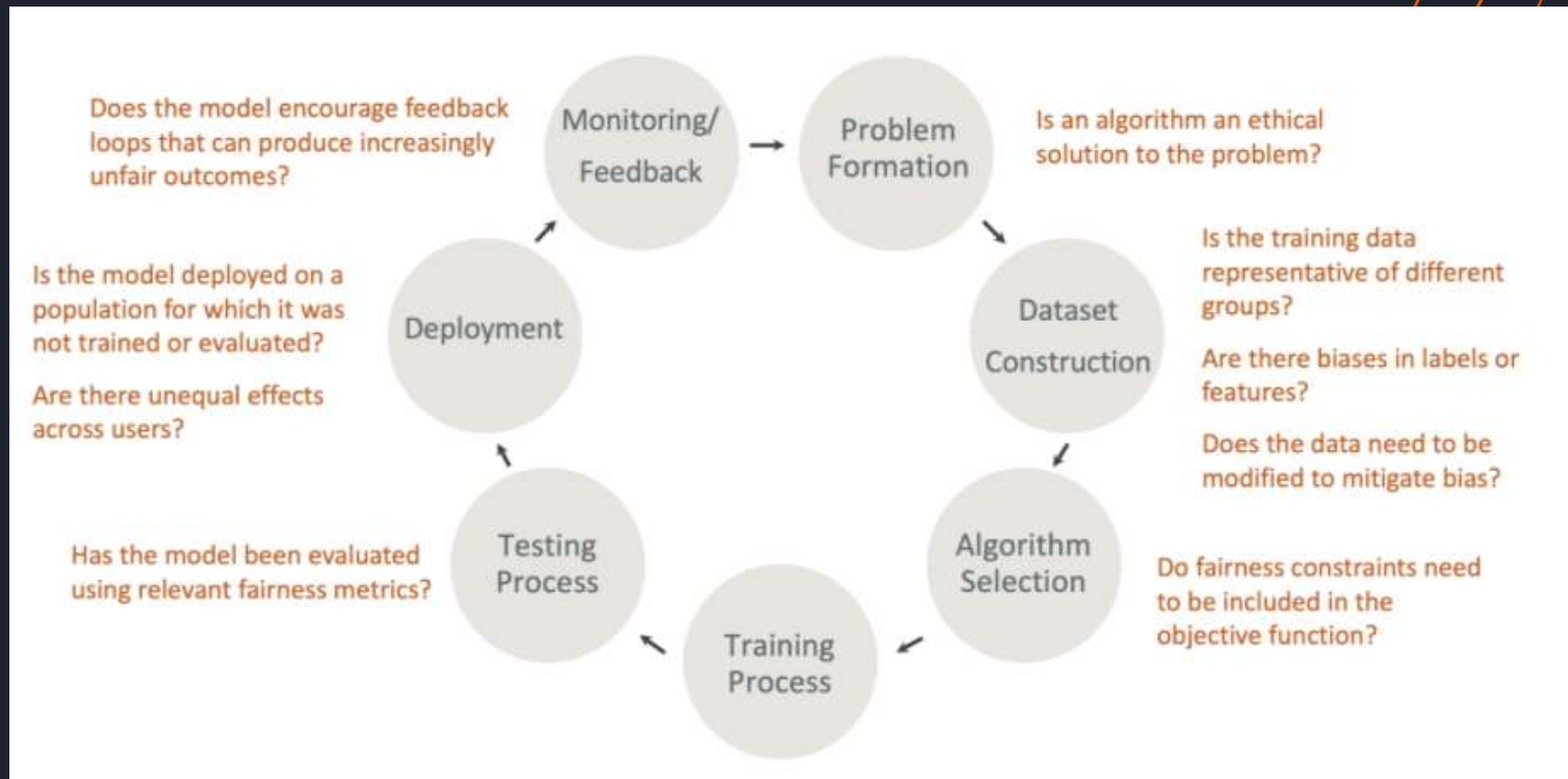
Understand How Machines Make Decision

AI-based systems are disrupting almost every industry – healthcare, finances, education, retail, media, marketing, social, economy, HR system, etc, etc. and helping us to make crucial decisions that are impacting millions of lives.

Hence it is important to understand how these decisions are made by the AI system.

Can we trust the Machine Learning, Deep Learning models, and their training algorithm?

Machine Learning Lifecycle



<https://docs.aws.amazon.com/sagemaker/latest/dg/clarify-fairness-and-explainability.html>

Explain the Model Decision

Simple Algorithms

- Bayesian classifiers,
 - Decision trees,
 - Linear Regression,
 - Logistic Regression etc ..
- possess a certain degree of interpretability, traceability, visibility, and transparency in their decision-making process at the cost of the performance

Complex Algorithms

- PCA,
 - Deep Neural Network,
 - Ensemble methods,
 - Random forests etc
- sacrifice their explainability to achieve high performance and accuracy

AWS Sagemaker clarify

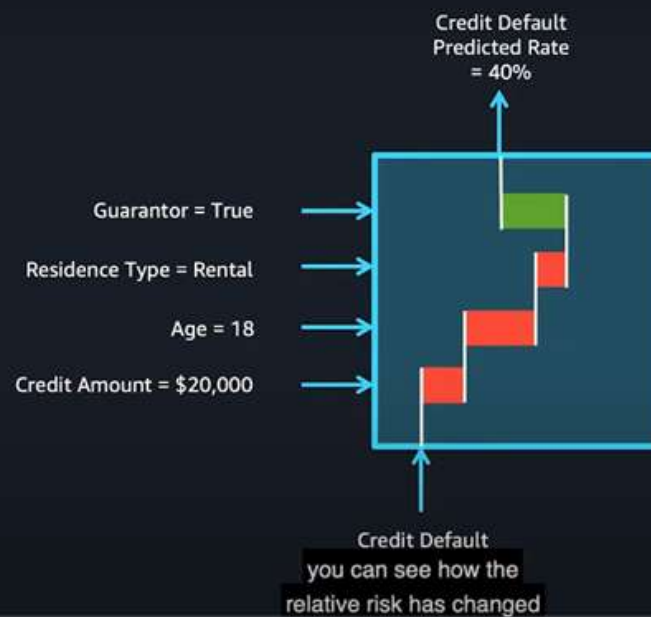
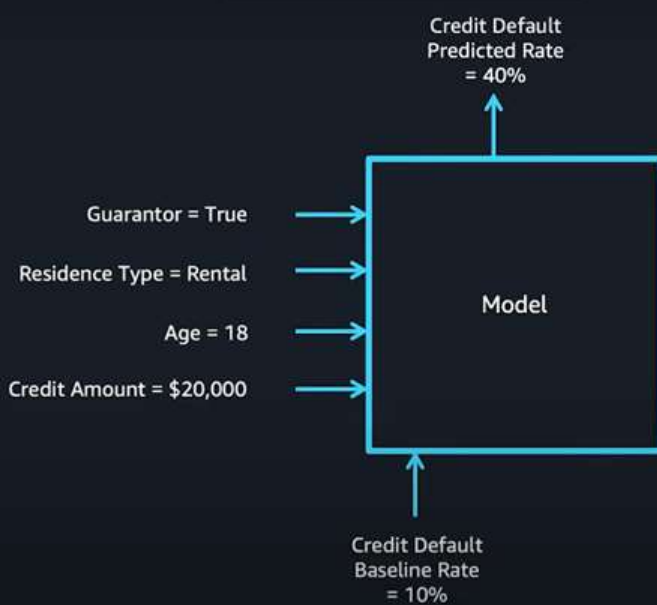
- Detecting dataset bias
 - ✓ Pre Training & Post Training Bias Detection
 - ✓ Measure bias using a various statistical metrics.
 - ✓ Help in early detection of bias in dataset
- Detecting model bias
 - ✓ Explain how feature values contribute to the predicted outcome, both for the model overall and for individual predictions.
 - ✓ Run a SageMaker Clarify analysis, which includes automatic deployment to a temporary endpoint, and computation of bias metrics using our model and dataset.
 - ✓ By computing these metrics, we can figure out if trained model has similar predictive behavior across groups.

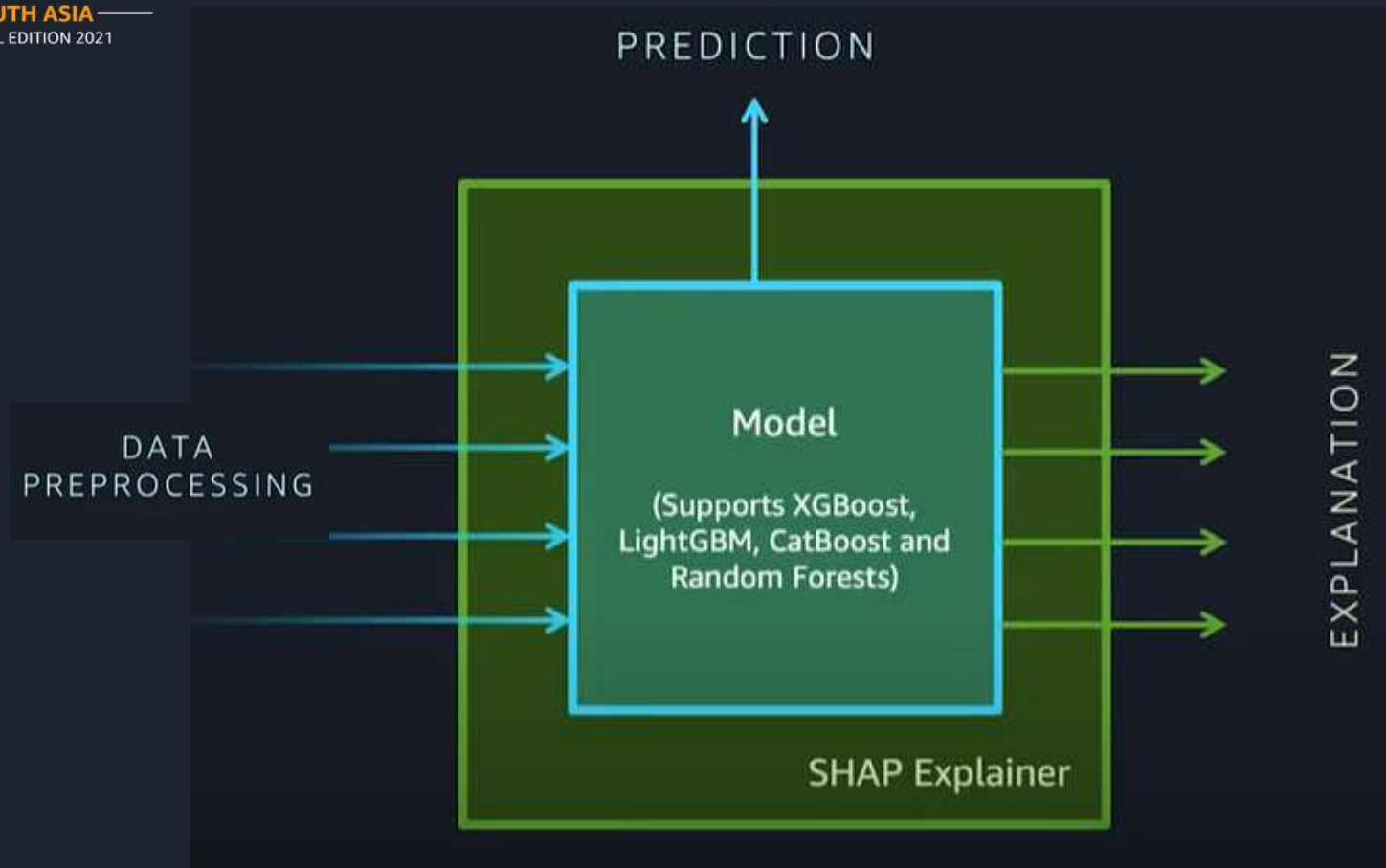
Explain Credit Decision with Sagemaker Clarify

- Classify credit applications and predict whether the credit would be payed back or not
- Bank to Reduce the risk of losing money due to unpaid credits
- Also reduce the risk of denying trustworthy customers credit which has a set of negative impacts.

Explain Credit Decision with Sagemaker Clarify

- Explain specific failure cases in model predictions
- Explain risk factor for credit default
- Achieve the stakeholder's trust in model
- Identify potential bias in model
- Comply with “right for explanation” for users





SHAPE

SHAP stands for SHapley Additive exPlanations. 'Shapley' relates to a game theoretic concept called Shapley values that is used to create the explanations.

A Shapley value describes the marginal contribution of each 'player' when considering all possible 'coalitions'.

Using this in a machine learning context, a Shapley value describes the marginal contribution of each feature when considering all possible sets of features.

'Additive' relates to the fact that these Shapley values can be summed together to give the final model prediction.

As an example, we might start off with a baseline credit default risk of 10%. Given a set of features, we can calculate the Shapley value for each feature. Summing together all the Shapley values, we might obtain a cumulative value of +30%. Given the same set of features, we therefore expect our model to return a credit default risk of 40% (i.e. $10\% + 30\%$).

Shapley Values

WITH GUARANTOR

Guarantor = True ✓
Residence Type = Rental ✓
Age = 18 ✓
Credit Amount = \$20,000 ✓
Credit Default Risk = 40%

WITHOUT GUARANTOR

Guarantor = ~~True~~ = False ✗
Residence Type = Rental ✓
Age = 18 ✓
Credit Amount = \$20,000 ✓
Credit Default Risk = 60%
Effect of Guarantor = -20%

WITH GUARANTOR

Guarantor = True ✓
Residence Type = ~~Rental~~ = Own ✗
Age = 18 ✓
Credit Amount = \$20,000 ✓
Credit Default Risk = 10%

WITHOUT GUARANTOR

Guarantor = ~~True~~ = False ✗
Residence Type = ~~Rental~~ = Own ✗
Age = 18 ✓
Credit Amount = \$20,000 ✓
Credit Default Risk = 20%
Effect of Guarantor = -10%, -20%

WITH GUARANTOR

Guarantor = True ✓
Residence Type = Rental ✓
Age = ~~18~~ = 45 ✗
Credit Amount = \$20,000 ✓
Credit Default Risk = 20%

WITHOUT GUARANTOR

Guarantor = ~~True~~ = False ✗
Residence Type = Rental ✓
Age = ~~18~~ = 45 ✗
Credit Amount = \$20,000 ✓
Credit Default Risk = 35%
Effect of Guarantor = -15%, -10%, -20%

Shapley Values

WITH GUARANTOR		WITHOUT GUARANTOR	
Guarantor = True	✓	Guarantor = True = False	✗
Residence Type = Rental = Own	✗	Residence Type = Rental = Own	✗
Age = 18	✓	Age = 18	✓
Credit Amount = \$20,000 = \$5,000	✗	Credit Amount = \$20,000 = \$5,000	✗
Credit Default Risk = 20%		Credit Default Risk = 45%	
		Effect of Guarantor = -25%, -5%, -15%, -15%, -10%, -20%	

WITH GUARANTOR		WITHOUT GUARANTOR	
Guarantor = True	✓	Guarantor = True = False	✗
Residence Type = Rental = Own	✗	Residence Type = Rental = Own	✗
Age = 18 = 45	✗	Age = 18 = 45	✗
Credit Amount = \$20,000 = \$5,000	✗	Credit Amount = \$20,000 = \$5,000	✗
Credit Default Risk = 5%		Credit Default Risk = 45%	
		Effect of Guarantor = -10%, -15%, -25%, -5%, -15%, -15%, -10%, -20%	
		Overall Effect of Guarantor = SHAP Value for Guarantor = -17%	

AWS Services

As part of the solution, the following services are used:

AWS Lambda: Used to generate a synthetic credits dataset and upload to Amazon S3.

AWS Glue: Used to crawl datasets, and transform the credits dataset using Apache Spark.

Amazon S3: Used to store datasets and the outputs of the AWS Glue Job.

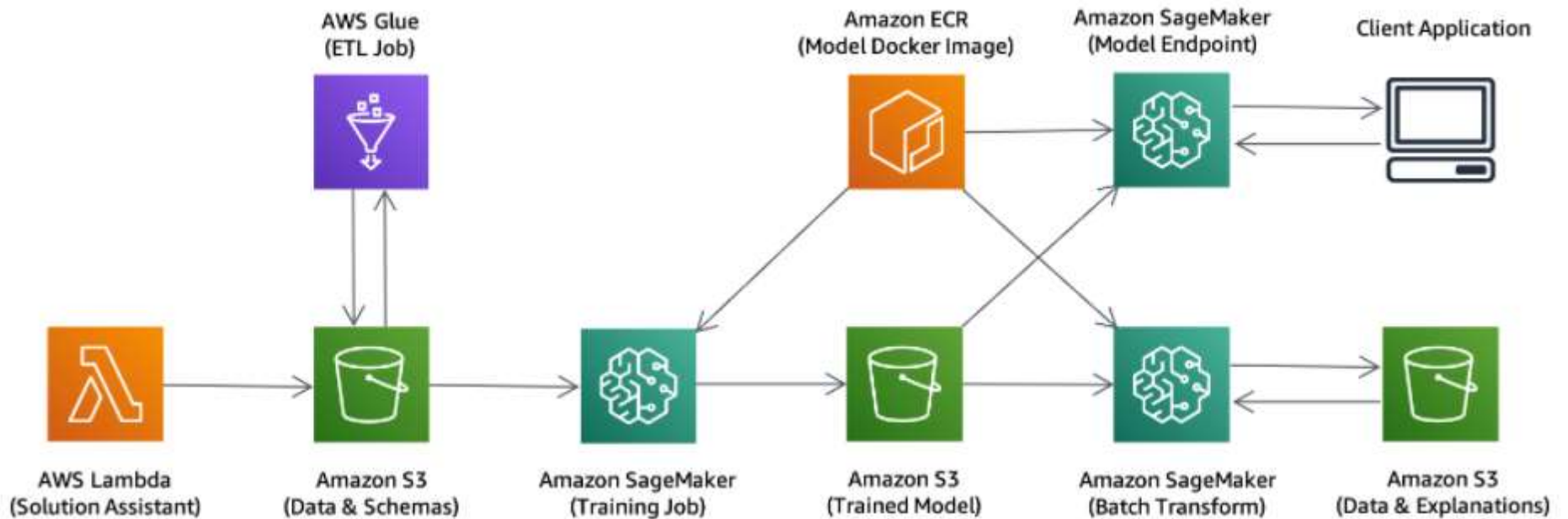
Amazon SageMaker Notebook: Used to train the LightGBM model.

Amazon ECR: Used to store the custom docker image with pre built Scikit-learn + LightGBM training environment. Used to train the model inside.

Amazon SageMaker Endpoint: Used to deploy the trained model and SHAP explainer.

Amazon SageMaker Batch Transform: Used to compute explanations in batch.

Architecture



Demo

Amazon SageMaker Studio

File Edit View Run Kernel Git Tabs Settings Help

OPEN TABS

- Launcher

Launcher

Get started

Explore one-click solutions, models, and tutorials

SageMaker JumpStart

[Solution: Detect malicious users and transactions](#)

[Solution: Demand forecasting](#)

[Go to SageMaker JumpStart](#)

Build models automatically

SageMaker Autopilot

[Video: Get started with Autopilot](#)

[Blog: Getting started with Autopilot](#)

[New autopilot experiment](#)

ML tasks and components

New compilation job

Create a new compilation job. [View compilation jobs](#)

New feature group

Create a new feature group in the feature store to logically group and manage features. [View feature store](#)

SAGEMAKER JUMPSTART LAUNCHED ASSETS

View launched solutions, deployed model endpoints and training jobs created with JumpStart.

[Browse JumpStart](#)

Solutions (5)

3 minutes ago

Your launched solutions.

[Explain Credit Decisions](#)

Ready - 23 days ago

[Fraud Detection in Financial Transactions](#)

Ready - 26 days ago

[Detect Malicious Users and Transactions](#)

Ready - 22 days ago

[Document Understanding](#)

Ready - 24 days ago

[Explain Credit Decisions](#)

Launcher

SageMaker JumpStart

about 1 hour ago

SageMaker JumpStart

Discover and deploy one-click solutions, pretrained models, and example notebooks

Search

Featured content



Corporate Credit Rating Prediction

Financial Services

Use multimodal data to predict corporate credit rating.



RoBERTa-SEC-WIKI-Large

Community Model - Text

Task:	Text Embedding
Pre-training Dataset:	S&P 500 10-K/10-Q (2010-2019) and Wikipedia
Fine-tunable:	No
Source:	Glue NLP



SEC Filings Retrieval w/ Summarizer and Scoring

Financial Services

This notebook shows how Amazon SageMaker JumpStart Industry helps deploy a simple solution to retrieve SEC filings and construct a dataframe of mixed tabular-and-text (TabText) data.

SAGEMAKER JUMPSTART LAUNCHED ASSETS

View launched solutions, deployed model endpoints and training jobs created with JumpStart.

🔍 Browse JumpStart

Solutions (5)

less than 20 seconds ago

Your launched solutions.

Explain Credit Decisions

Ready · 23 days ago

Fraud Detection in Financial Transactions

Ready - 26 days ago


Detect Malicious Users and Transactions

Ready - 22 days ago

Document Understanding

Ready · 24 days ago

Explain Credit Decisions

 SageMaker JumpStart

 Explain Credit Decisions

 Explain Credit Decisions X

0_introduction.ipynb

 1_datasets.ipynb

SOLUTION

Explain Credit Decisions

 Browse JumpStart

Launch Solution

Launch an end-to-end solution that will create supporting AWS infrastructure and tools to allow you to run specific ML workflows for different use cases.

Launch

Solution

Given the increasing complexity of machine learning models, the need for model explainability has been growing lately. Some governments have also introduced stricter regulations that mandate a right to explanation from machine learning models. In this solution, we take a look at how [Amazon SageMaker](#) can be used to explain individual predictions from machine learning models.

As an example application, we classify credit applications and predict whether the credit would be paid back or not (often called a credit default).

AWS Sagemaker Studio

The screenshot displays the AWS SageMaker Studio web interface. The top navigation bar includes 'Amazon SageMaker Studio' and a menu with 'File', 'Edit', 'View', 'Run', 'Kernel', 'Git', 'Tabs', 'Settings', and 'Help'. On the left, a file explorer shows a directory structure with 'S3Downloads' and 'solutions_local_metadata.j...'. The main content area has three tabs: 'Launcher', 'SageMaker JumpStart', and 'Corporate Credit Rating Predi'. The 'Corporate Credit Rating Predi' tab is active, showing a 'SOLUTION' section titled 'Corporate Credit Rating Prediction' with a 'Browse JumpStart' button. Below this is a 'Launch Solution' section with a description: 'Launch an end-to-end solution that will create supporting AWS infrastructure and tools to allow you to run specific ML workflows for different use cases.' and a 'Launch' button. At the bottom, a 'Corporate Credit Rating' section provides a detailed description of the solution, mentioning it is a 'text-enhanced' credit rating model based on numeric features (Altman's 5 financial ratios) and SEC filings, aiming to improve credit rating predictions.

Amazon SageMaker Studio

File Edit View Run Kernel Git Tabs Settings Help

Launcher SageMaker JumpStart Corporate Credit Rating Predi

SOLUTION

Corporate Credit Rating Prediction

Browse JumpStart

Launch Solution

Launch an end-to-end solution that will create supporting AWS infrastructure and tools to allow you to run specific ML workflows for different use cases.

Launch

Corporate Credit Rating

This SageMaker JumpStart Industry solution provides a template for a *text-enhanced* credit rating model. It shows how to take a model based on numeric features (in this case, Altman's famous 5 financial ratios) combined with texts from SEC filings so as to achieve an improvement in the prediction of credit ratings. You are not restricted to the 5 Altman ratios; you can add more variables as needed or completely change the variables. The main objective of this solution notebook is to show how SageMaker JumpStart Industry can help process NLP scoring of SEC filings text and use the *Altman's Z-score* to compute the Altman's 5 financial ratios to enhance features.

SAGEMAKER JUMPSTART LAUNCHED ASSETS

View launched solutions, deployed model endpoints and training jobs created with JumpStart.

[Browse JumpStart](#)

Solutions (4)

2 minutes ago

Your launched solutions.

[Explain Credit Decisions](#)

Ready · 23 days ago

[Fraud Detection in Financial Transactions](#)

Ready · 27 days ago

[Document Understanding](#)

Ready · 24 days ago

[Detect Malicious Users and Transactions](#)

Ready · 22 days ago

SageMaker JumpStart

Corporate Credit Rating Predi

Corporate Credit Rating Predi

Deployed less than 5 seconds ago

Base Solution: Corporate Credit Rating Prediction

Deploy Status

Launching: 15% Complete less than 20 seconds ago

The resources for this solution are being created. Please allow a few minutes for all the resources to be created before proceeding.

Generated Artifacts

View Artifacts Generated by the solution. less than 20 seconds ago

Name	Type
sagemaker-soln-ccr-js-1wd00o-demo-endpoint	Endpoint
sagemaker-soln-ccr-js-1wd00o-demo-model	Model

SAGEMAKER JUMPSTART LAUNCHED ASSETS

View launched solutions, deployed model endpoints and training jobs created with JumpStart.

[Browse JumpStart](#)

Solutions (4)

9 minutes ago

Your launched solutions.

[Explain Credit Decisions](#)

Ready · 23 days ago

[Fraud Detection in Financial Transactions](#)

Ready · 27 days ago

[Document Understanding](#)

Ready · 24 days ago

[Detect Malicious Users and Transactions](#)

Ready · 22 days ago

SageMaker JumpStart

Corporate Credit Rating Predi

Corporate Credit Rating Predi

Deployed less than 5 seconds ago

Base Solution: Corporate Credit Rating Prediction

Deploy Status

Launching: 99% Complete less than 20 seconds ago

The resources for this solution are being created. Please allow a few minutes for all the resources to be created before proceeding.

Generated Artifacts

View Artifacts Generated by the solution. less than 20 seconds ago

Name	Type	Status
sagemaker-soln-ccr-js-1wd00o-demo-endpoint	Endpoint	Creating
sagemaker-soln-ccr-js-1wd00o-demo-model	Model	Complete

AWS Sagemaker Studio

The screenshot displays the AWS Sagemaker Studio console. On the left, the 'SageMaker resources' sidebar shows a 'Projects' dropdown menu, a search bar, and a 'Create project' button. Below this, a list of projects is visible, with 'demosagemakerstudio' selected. The main panel shows the 'Explain Credit Decisions' solution, which was deployed 11 days ago. It includes a 'Browse JumpStart' button, a section titled 'Open solution in Studio' with an 'Open Notebook' button, and a 'Delete solution' section with a 'Delete all resources' button.

SageMaker resources
Select the resource to view.

Projects

Search

Create project

Name

demosagemakerstudio

End of the list

SageMaker JumpStart X Explain Credit Decisions X

Explain Credit Decisions
Deployed 11 days ago

Browse JumpStart

Base Solution: Explain Credit Decisions

Open solution in Studio

Open a notebook to walk through the solution using the data provided or customize it to use your own.

Open Notebook

Delete solution

If you no longer want this solution you can easily delete the resources it created. Selecting the button below will automatically delete all standard resources that were created when launching the solution, including the S3 bucket.

Delete all resources

AWS Sagemaker Studio

The screenshot displays the AWS SageMaker Studio web interface. On the left, a file explorer shows a directory structure with notebooks: 0_introduction.ipynb, 1_datasets.ipynb, 2_training.ipynb, 3_endpoint.ipynb, 4_batch_transform.ipynb, and 5_conclusion.ipynb, all modified 20 minutes ago. The main panel shows the 'SageMaker JumpStart' page, which includes a search bar and a 'Featured content' section. Two featured items are visible: 'Corporate Credit Rating Prediction' (Financial Services) and 'RoBERTa-SEC-WIKI-Large' (Community Model - Text). The RoBERTa model details include: Task: Text Embedding, Pre-training Dataset: S&P 500 10-K/10-Q (2010-2019) and WikiPedia, and Fine-tunable: No.

Amazon SageMaker Studio

File Edit View Run Kernel Git Tabs Settings Help

/ ... / jumpstart-prod-ecd_183g1s / notebooks /

Name	Last Modified
0_introduction.ipynb	20 minutes ago
1_datasets.ipynb	20 minutes ago
2_training.ipynb	20 minutes ago
3_endpoint.ipynb	20 minutes ago
4_batch_transform.ipynb	20 minutes ago
5_conclusion.ipynb	20 minutes ago

Launcher

about 1 hour ago

SageMaker JumpStart

Discover and deploy one-click solutions, pretrained models, and example notebooks

Search

Featured content

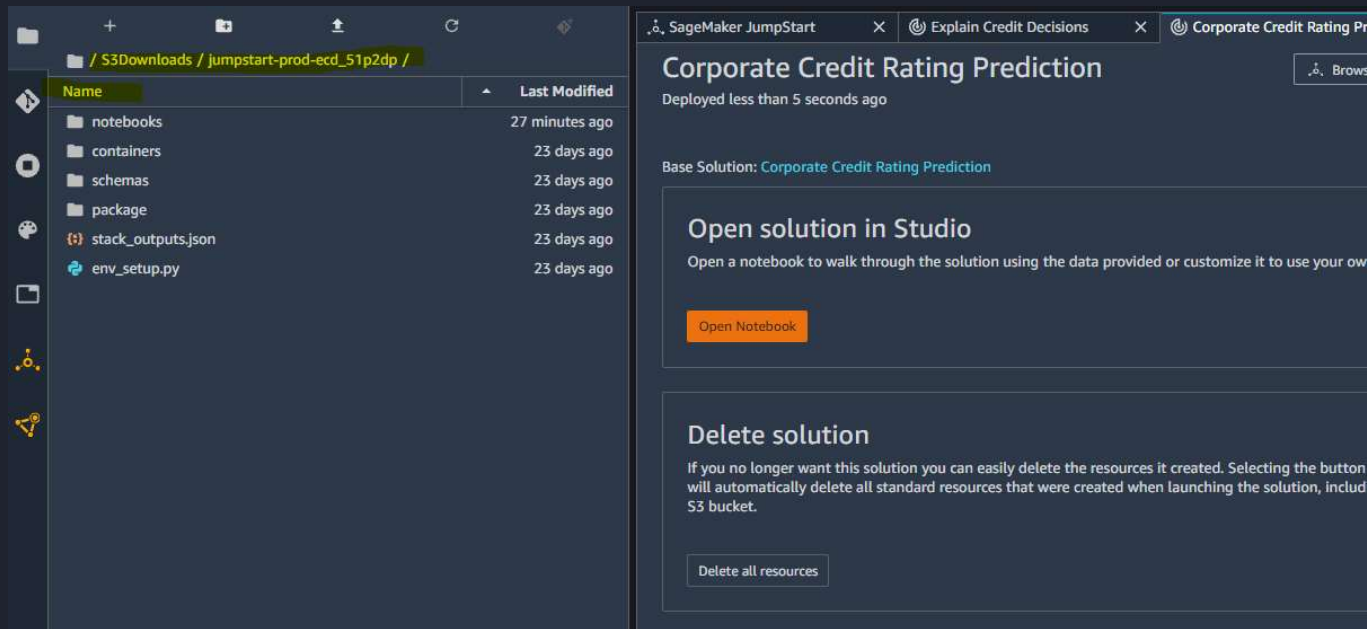
Corporate Credit Rating Prediction
Financial Services

Use multimodal data to predict corporate credit rating.

RoBERTa-SEC-WIKI-Large
Community Model - Text

Task:	Text Embedding
Pre-training Dataset:	S&P 500 10-K/10-Q (2010-2019) and WikiPedia
Fine-tunable:	No

AWS Sagemaker Studio



Once the solution is launched the notebooks are run in order notebook_1 to notebook_5 and Code directory structure is created in Sagemaker Studio as shown above. Source codes are available in container folder. These codes are required when we will customize the solution for our own data or model.

Stages

- Our solution is split into the following stages, and each stage has its own notebook:
- Introduction: We take a high-level look at the solution components.
- Datasets: We prepare a dataset for machine learning using AWS Glue.
- Training: We train a LightGBM model using Amazon SageMaker, so we have an example trained model to explain.
- Endpoint: We deploy the model explainer to a HTTP endpoint using Amazon SageMaker and visualize the explanations.
- Batch Transform: We use Amazon SageMaker Batch Transform to obtain explanations for our complete dataset.
- Dashboard: We develop a dashboard for explanations using Amazon SageMaker and Streamlit.
- Conclusion: We wrap things up and discuss how to clean up the solution.

Model Deployment

- Once the SageMaker training job is completed, a number of trained model artifacts are stored in S3 bucket
- Retrieve the model data (i.e. model.tar.gz) from the most recent trained model
- Define the model using SKLearnModel() to deploy which includes the explainer logic.
- Calling model.deploy() will start a container to host the model.
- Entities are used ->

```
entities = [  
    'data',  
    'features',  
    'descriptions',  
    'prediction',  
    'explanation_shap_values',  
    'explanation_shap_interaction_values'  
]
```

Model Explanation

- Call `explainer.predict` with features (for a credit application) to obtain a prediction and explanation.
- Visualize Explanations with bokeh & waterfall chart.
- A waterfall chart can be used to show the cumulative effect of each feature.
- Green arrows indicate that the feature *decreased* the predicted credit default risk for the individual credit application.
- While red arrows indicate that the feature *increased* the predicted credit default risk for the individual credit application.
- After all features have been considered, we reach the final predicted credit default risk (at the top of the chart).

Solution Code & Explanation

Please refer the reference slide for the link of:

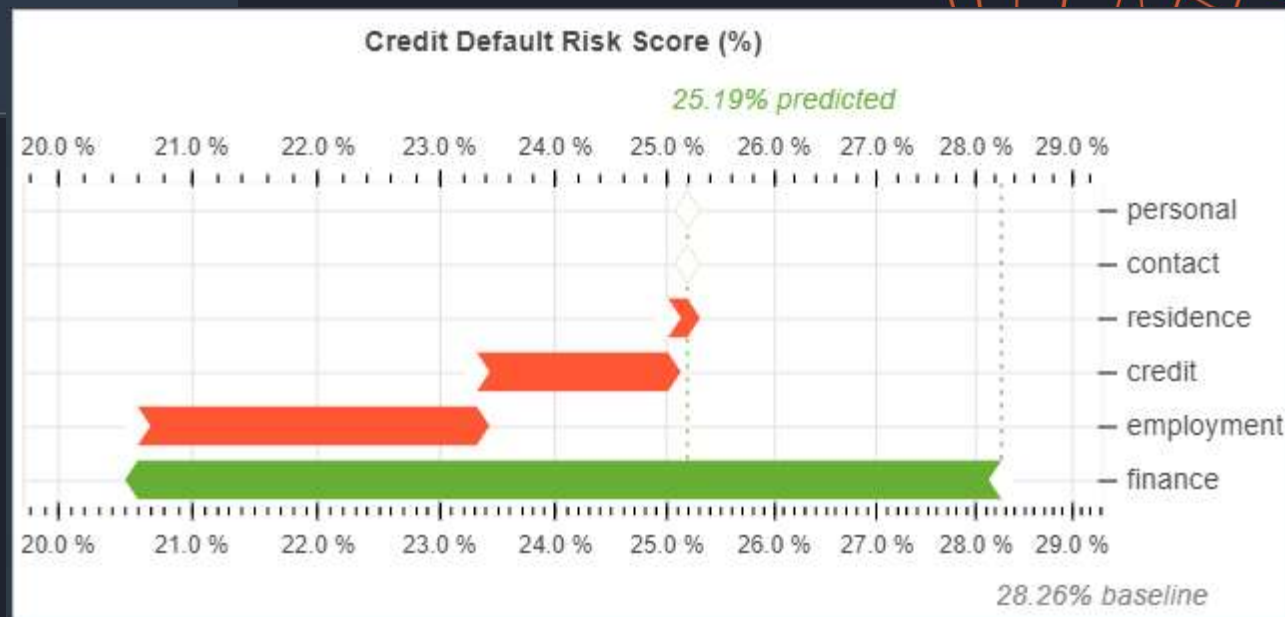
- Youtube : Explaining Credit Decisions with Amazon SageMaker – Webinar by Solution Developer
- AWS Lab github links

Model Explanation

```
x_axis_label = 'Credit Default Risk Score (%)'  
summary_waterfall = visuals.WaterfallChart(  
    baseline=explanation_summary['expected_value'],  
    shap_values=explanation_summary['shap_values'],  
    names=explanation_summary['feature_names'],  
    descriptions=explanation_summary['feature_descriptions'],  
    max_features=10,  
    x_axis_label=x_axis_label,  
)  
summary_waterfall.show()
```

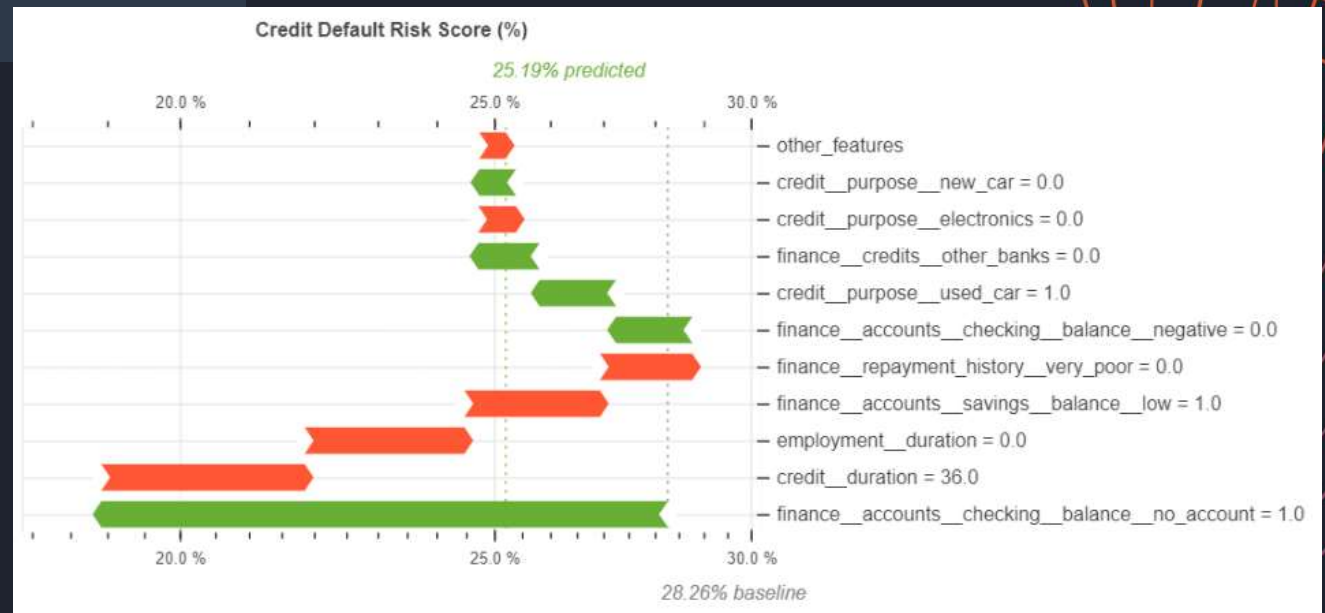
We can see from the summary waterfall chart that features related to finance have the largest combined effect on the credit default risk.

Although features related to finance reduce the credit default risk, the features related to employment bring the risk back up again to a certain degree.



Explain Prediction

```
detailed_waterfall = visuals.WaterfallChart(  
    baseline=explanation['expected_value'],  
    shap_values=explanation['shap_values'],  
    names=explanation['feature_names'],  
    feature_values=explanation['feature_values'],  
    descriptions=explanation['feature_descriptions'],  
    max_features=10,  
    x_axis_label=x_axis_label  
)  
detailed_waterfall.show()
```

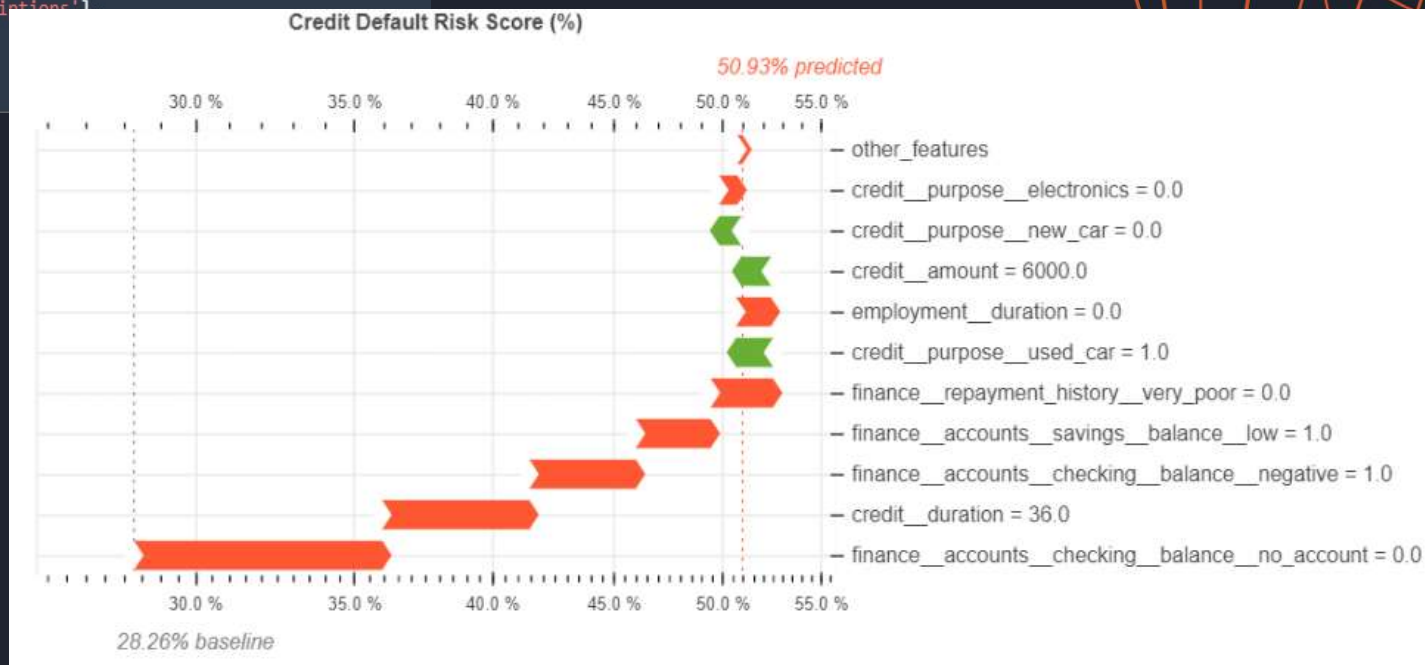


Explain Prediction

- Detailed waterfall chart shows that
 - Not having a checking account with the same bank indicates a lower credit default risk.
 - The credit to purchase a used car is associated with a lower credit default risk
 - After this we see a number of features that increase the credit default risk:
 - A credit amount of 6000 EUR, a lack of employment and a credit duration of 36 months.
 - Another potential area for investigation, would be related to the repayment history feature.
 - Not having a very poor repayment history is associated with a higher credit default risk score.

Counterfactual Example

```
counter_sample = dict(sample)
counter_sample['finance__accounts__checking__balance'] = 'negative' # from 'no_account'
counter_output = explainer.predict(counter_sample)
counter_explanation = visuals.detailed_explanation(counter_output)
visuals.WaterfallChart(
    baseline=counter_explanation['expected_value'],
    shap_values=counter_explanation['shap_values'],
    names=counter_explanation['feature_names'],
    feature_values=counter_explanation['feature_values'],
    descriptions=counter_explanation['feature_descriptions'],
    max_features=10,
    x_axis_label=x_axis_label,
).show()
```



Endpoint Example

Amazon SageMaker Studio

File Edit View Run Kernel Git Tabs Settings Help

SageMaker resources
Select the resource to view.

Endpoints

1 row selected 0/20 filters

Search column name to start

Name

sagemaker-soln-ecd-js-183g1s-explainer

End of the list

less than 10 seconds ago

MODEL MONITORING

Endpoint: sagemaker-soln-ecd-js-183g1s-explainer

Test inference Data quality Model quality Model explainability Model bias Monitoring job history AWS settings

Test inference

Test and validate your model by sending a request to the SageMaker hosted endpoint and receiving a response from your model.

JSON editor

```
1 {  
2     
3 }
```

Send Request

less than 5 seconds ago

Counterfactual Example

- Now let's switch the value of the checking account balance of the applicant from no account to negative.
- We can then see how the overall prediction of the model changes, and also see the updated contribution of this feature.
- Clearly, this application has become substantially more risky.

References

Youtube : Explaining Credit Decisions with Amazon SageMaker – Webinar
<https://www.youtube.com/watch?v=Nlwz4cU68T8&t=1204s>
by AWS Solution Developer

AWSLABS : <https://github.com/aws-labs/sagemaker-explaining-credit-decisions>

My Github link : <https://bit.ly/3BzAsQT>

My Linkedin : <https://www.linkedin.com/in/sarbani-maiti-35b89111/>

Thank You