Machine Learning Introduction in AWS

What are the use case of Artificial Intelligence and Machine Learning services in AWS?

AI and machine learning solutions provided by AWS enable businesses to innovate, enhance customer experiences, and optimize processes.

- Innovate faster with cloud

- Solutions to business needs

- Solve real-world business problems

- Enhanced customer experiences and better decision-making

- Resources to upskill your team with educational content

Artificial intelligence and its subsets
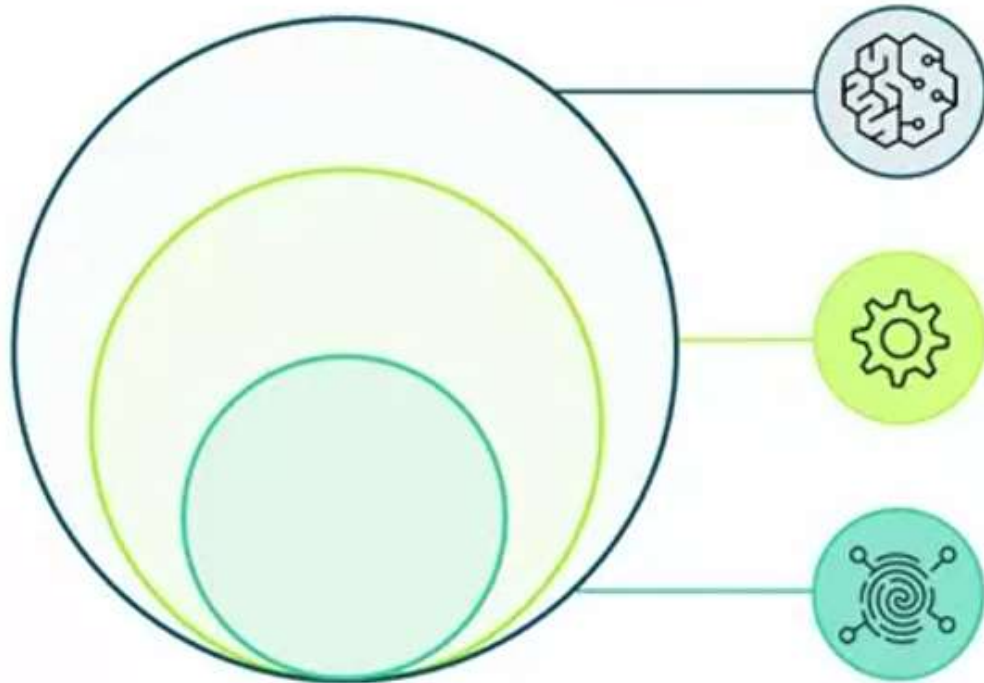
Artificial Intelligence (AI)
- A field of study that uses computers to do processes that mimic human behavior

Machine Learning (ML)
- A subset of AI
- Uses algorithms to learn and improve from training data

Deep Learning (DL)
- A subset of ML
- Uses multilayer networks to build models that are inspired by the human brain

Machine Learning Terminology

- **Training**:

  Machine learning algorithms are given labeled data to learn patterns and optimize their parameters through iterations, resulting in an improved model that can make accurate predictions.

- **Model**:

  A model is the output of training that captures patterns and relationships that are learned from training data.   It is used for formulating predictions or making judgments based on new data.

- **Inference and Hosting**:

  Deploying the trained model on a server or platform to process real-world data and generate predictions is hosting. It can act as an API endpoint which can be called for getting results.

Machine Learning Services on AWS

- **Amazon Polly**: Text-to-Speech

- **Amazon Transcribe**: Speech-to-Text

- **Amazon SageMaker**: Build and Deploy Machine Learning Models

- **Amazon Textract**: Extract Text and Data

- **Amazon Kendra**: Find Accurate Information Faster

Machine Learning Services on AWS

- **Amazon Personalize**: Personalize Online Experiences

- **Amazon Rekognition**: Automated Image and Video Analysis

- **Amazon Lex**: Conversational AI for Chatbots

- **Amazon Comprehend**: Natural-Language Processing

- **Amazon Translate**: Translation

Amazon Rekognition

Amazon Recognition is a managed computer vision service that uses deep learning technology to automate picture and video analysis. It allows for many different type analysis and tasks to be performed on images and videos.

Rekoginition is used for the following:

- Automates image and video analysis.

- Overcomes challenges like cost, time, errors, and scalability issues associated with manual analysis

- Use deep learning technology for efficient image and video analysis.

- It can perform accurate facial analysis, object detection, text detection etc.

- It can identify inappropriate content to enforce policies and regulations.

- It can perform facial analysis for user identification and verification purposes.

Amazon Rekognition available API's

What API's dose Rekoginition use and its functions?

1. **CompareFaces**: Match faces in different images.

2. **DetectFaces**: Identify faces and provide comprehensive details.

3. **DetectLabels**: Recognize objects, concepts, events with confidence scores.

4. **DetectModerationLabels**: Detect content for moderation purposes.

5. **DetectProtectiveEquipment**: Identify face, hand, and head coverings.

6. **DetectText**: Detect text within images.

7. **CelebrityRecognition**: Identify celebrity faces.

Amazon Transcribe

Amazon Transcribe offers automatic speech-to-text conversion with various use cases including customer insights, subtitling, toxicity detection, and clinical documentation.

Use Cases:

- Get insights from customer conversations.

- Create subtitles and meeting notes, enhancing accessibility and customer experience.

- Detect toxic content in audio for safer online interactions.

- Improve clinical documentation by quickly transcribing clinical conversations into electronic health records (EHR).

Amazon SageMaker

Amazon SageMaker offers built capabilities that enable data scientists and developers to efficiently build, train, and deploy machine learning models, while AWS would manage the infrastructure complexities.

Use Cases:

- SageMaker streamlines machine learning model development, allowing for rapid creation, training, and deployment.

- Loads data from multiple sources such as Redshift and Amazon S3 that facilitates Data preparation

- Converting and transforming raw data into meaningful features using Data Wrangling

- SageMaker Studio, an integrated development environment, provides familiar tools like visual editors, debuggers, and continuous integration/continuous deployment (CI/CD) capabilities.

Amazon Textract

Amazon Textract is a machine learning solution that extracts data from various documents, eliminating the need for manual entry or through code. Textract is suitable for all types of businesses and can significantly speed up document processing.

Use Cases:

- Extract data from documents to avoid manual data entry which is slow, expensive, and prone to error.

- Amazon Textract uses machine learning to read and process any document type and extracts text, tables, and other data from them.

- Ideal for financial services, healthcare, public sector, etc. - as it can process millions of pages in hours.

- This service from AWS can reduces document processing costs and enhances efficiency without any code or template as it automatically handles extraction.

Amazon Kendra

"Amazon Kendra is an intelligent enterprise search service that helps you search across different content repositories with built-in connectors."
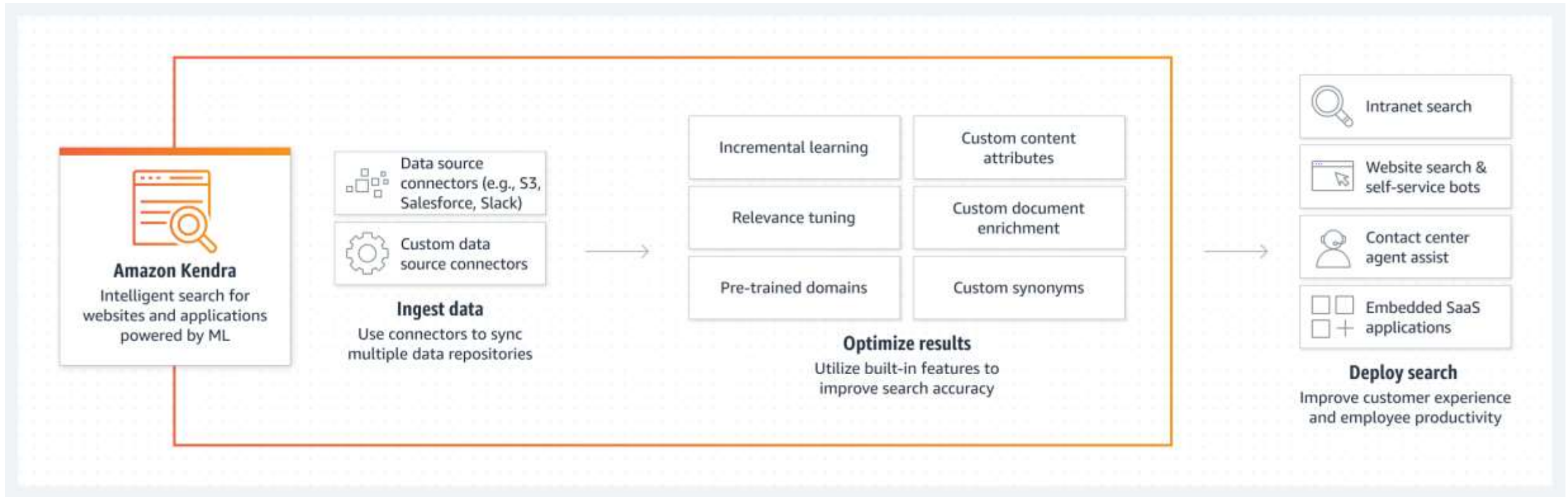
- It searches across structured and unstructured content repositories.
- It uses natural language processing (NLP) for getting precise answers.
- It can improve internal search experiences, enhancing employee productivity and data-driven decisions through a unified search interface.

Amazon Lex

"Amazon Lex is a fully managed artificial intelligence (AI) service with advanced natural language models to design, build, test, and deploy conversational interfaces in applications."

- It can be used to make virtual agents and voice assistants for simple tasks and self-service.
- Automate responses to FAQs for tech support, HR, and finance using natural language search with Amazon Kendra.
- It can be used as application bots for automating user tasks, integrating with enterprise software through AWS Lambda.

Amazon Comprehend

"Amazon Comprehend is a natural-language processing (NLP) service that uses machine learning to uncover valuable insights and connections in text."

- It can be used detect sentiment, categorize support requests, and improve products using customer survey insights.
- It is used to automate insights extraction from legal documents.
- It can classify entities, find relationships in financial documents, and handle insurance claims or mortgages.

Amazon Translate

"Amazon Translate is a neural machine translation service that delivers fast, high-quality, affordable, and customizable language translation."
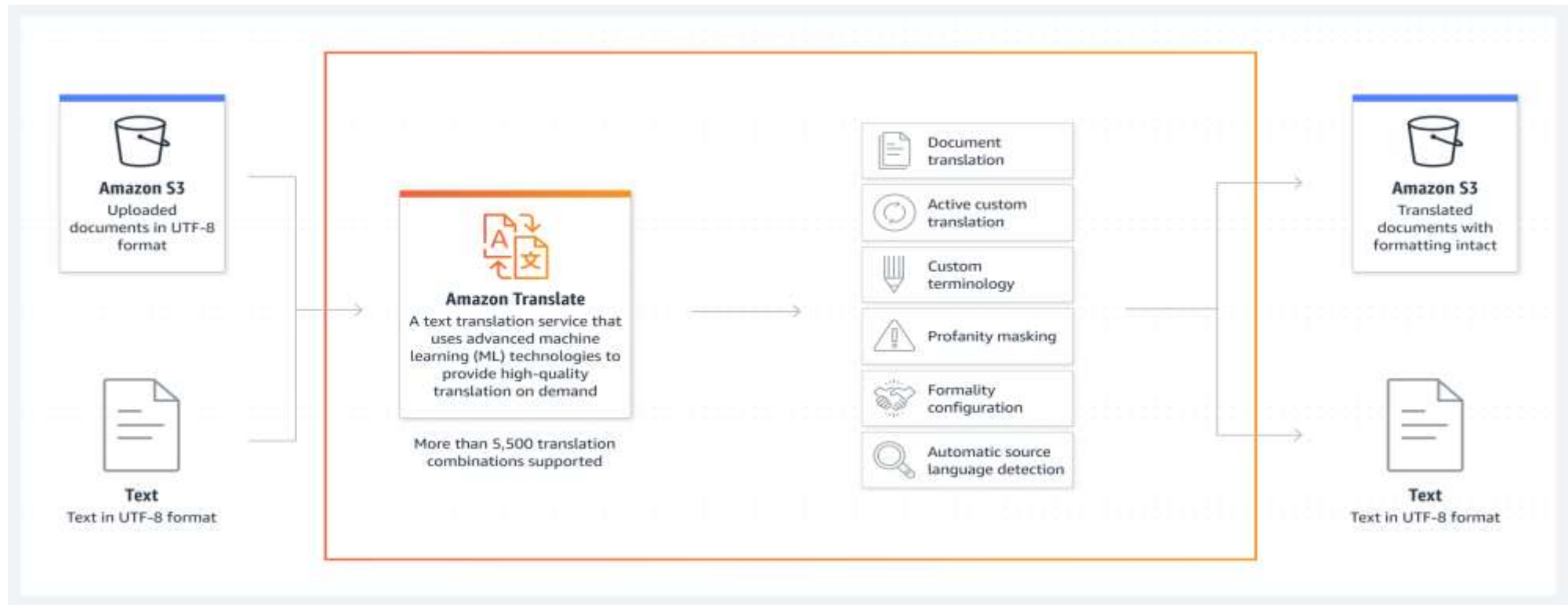
- Translate is designed to bridge language gaps for effective communication and growth in organizations.
- This service translates languages for content localization, subtitles, and communication between speakers of different languages.
- Translate can Integrate into applications and process the content in various formats.

Solving a business problem with AWS ML services

- Understanding the problem.

- Select a related service that can solve the problem.

- Select the right ML algorithm for the problem.

- Read about the security the service provides.

- Consider the Latency of the service that you use.

- Consider the accuracy of AWS ML services as it varies based on the specific use case.

Machine Learning Regression Models with AWS

Regression is a type of supervised learning where the goal is to predict a continuous numeric value based on input features, such as the price of a house, or how many people will use a city's bike rental service etc.

How to build a regression Model with AWS?

1. Amazon SageMaker Autopilot

2. Amazon Redshift ML

3. Amazon Forecast

4. Custom Solutions with EC2 and Custom Code

What is a Regression Model

ML models for regression problems predict a numeric value. For training regression models, Amazon ML uses the industry-standard learning algorithm known as linear regression.

Examples of Regression Problems

- "What will the temperature be in Seattle tomorrow?"

- "For this product, how many units will sell?"

- "What price will this house sell for?"

Regression Analysis Working

How does regression analysis work?

- Identify the question
- Collect historical data
- Train the regression analysis model
- Make predictions for unknown values

$$f(x) = \frac{1}{1 + e^{-x}}$$

If you plot this logistic regression equation, you will get an S-curve as shown below.
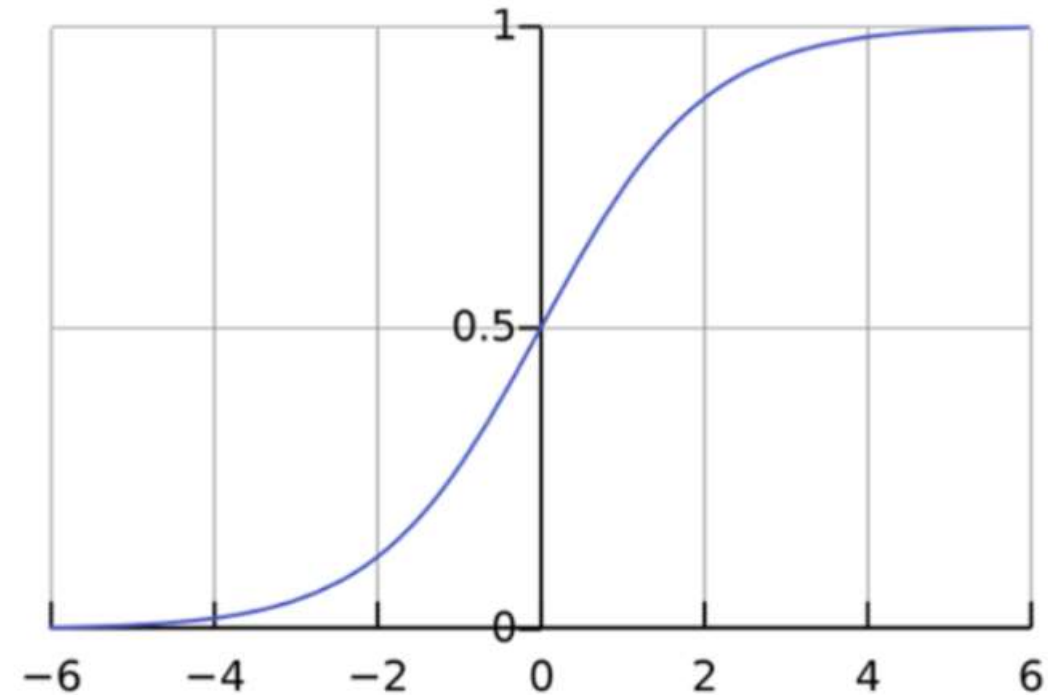
How does the logistic regression model work?

- Equations - $y = 2*x$
- Variables
- Logistic regression function >

Example - Regression Models with SageMaker Autopilot

Example: Creating a Regression model using SageMaker Autopilot to predict house price

Data Source – Kaggle House Pricing Dataset

Step 1 : Create a bucket in Amazon S3 and upload the training dataset.
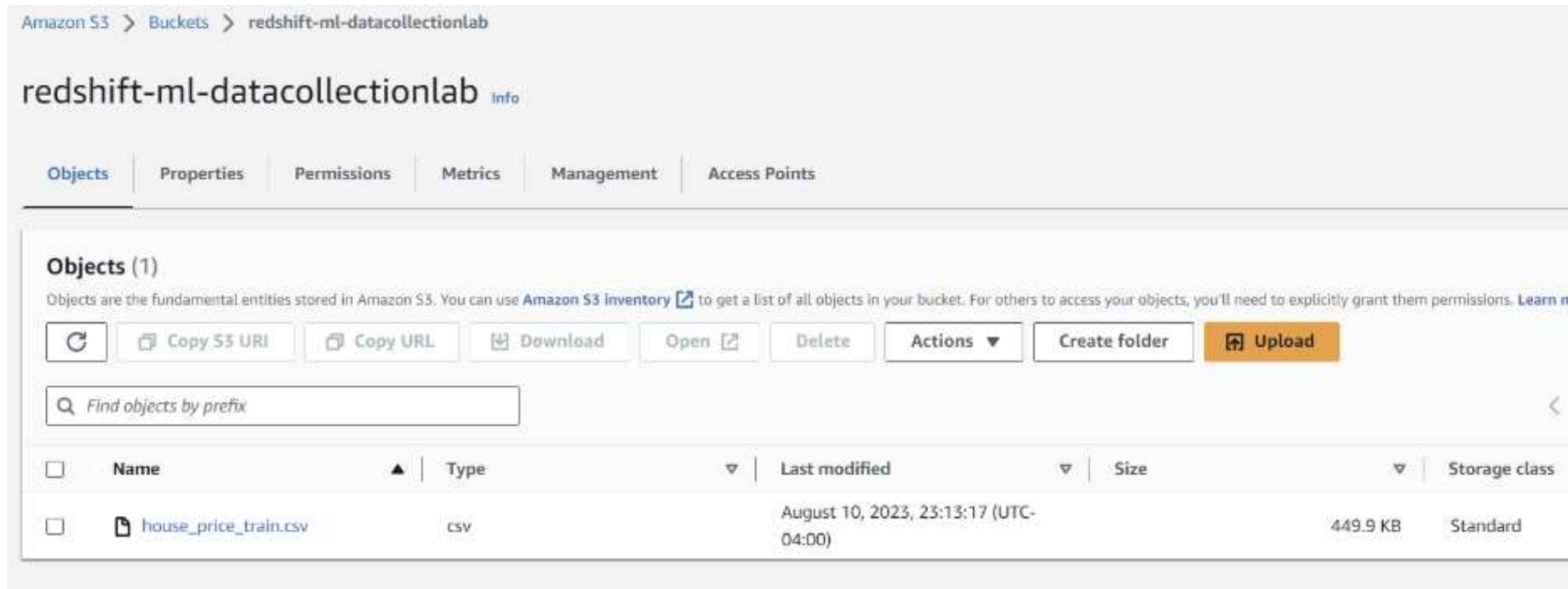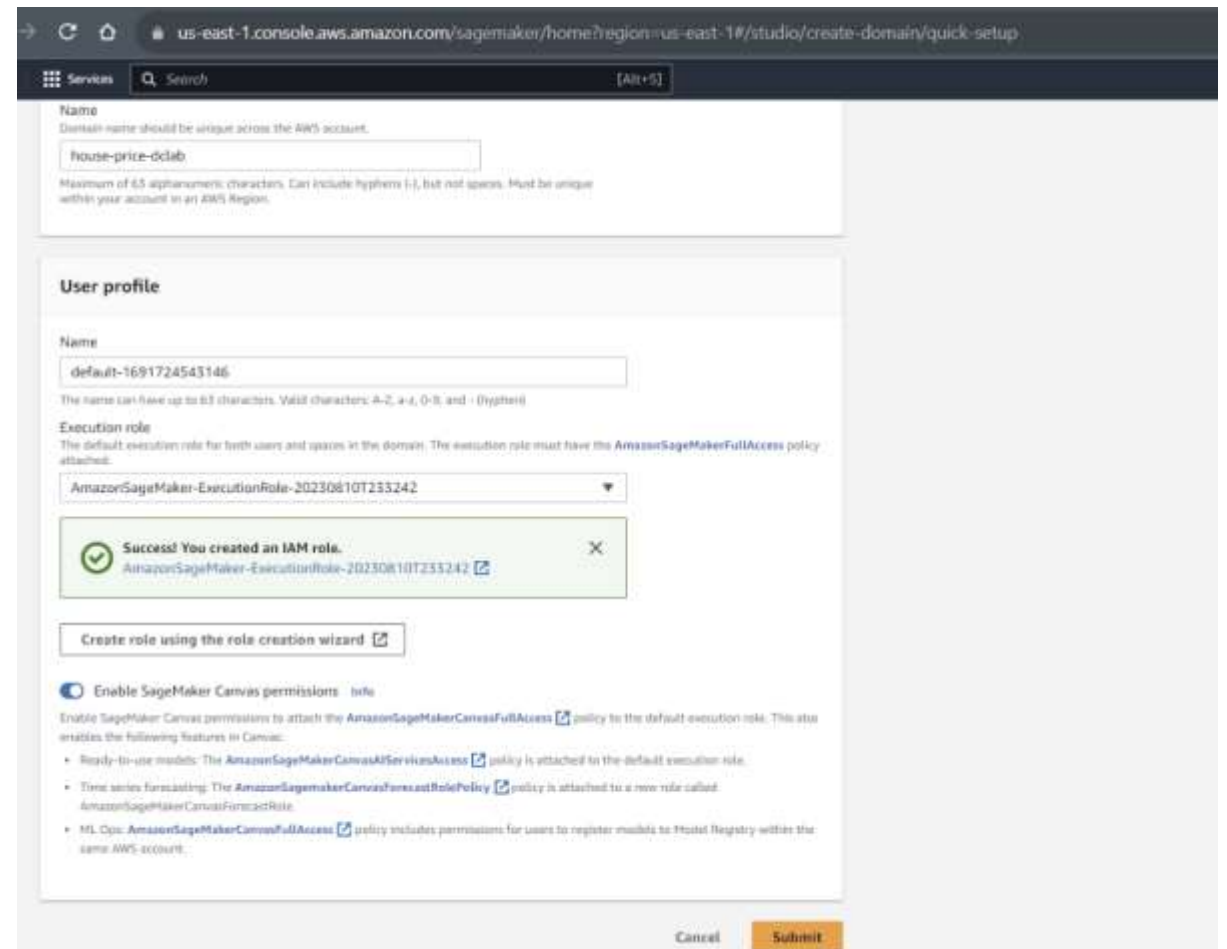
Example - Regression Models with SageMaker Autopilot

Example: Creating a Regression model using SageMaker Autopilot to predict house price

Step 2 : Setting up a SageMaker domain.

AWS Service : SageMaker Studio

- Select Domain

- Create a SageMaker domain

- Quick setup

- Create IAM Role : AmazonSageMakerFullAccess

- Specify S3 Bucket

- Save – will create a new Domain

- Launch SageMaker Studio

Example - Regression Models with SageMaker Autopilot

Example: Creating a Regression model using SageMaker Autopilot to predict house price

Step 3 : Import the dataset on SageMaker and process the data

Inside SageMaker Studio

- Data Wrangler

- Import data

- Data sources, choose Amazon S3

- Select the bucket

- Select the dataset and Import

- Transform - Add step

- Select Manage Columns

- Drop Column – Id

- Export and Train

Example - Regression Models with SageMaker Autopilot

Example: Creating a Regression model using SageMaker Autopilot to predict house price

Step 4 : Create an Autopilot experiment

Inside SageMaker Studio

- Naming the experiment.

- Setting the Target Value:

   SalePrice

- Select Training method

- Select the ML problem type:

   Regression
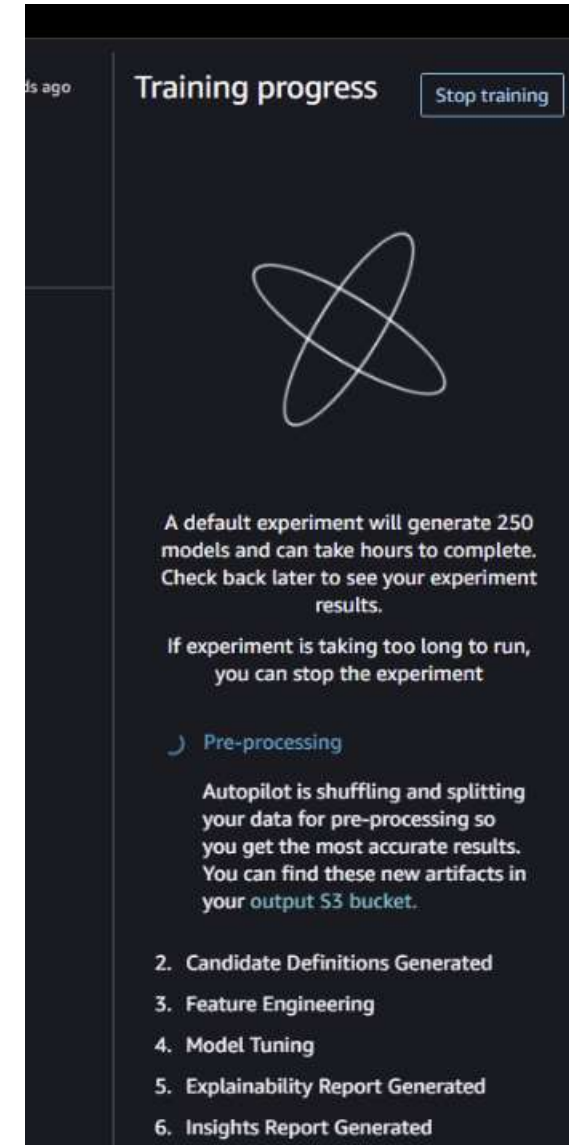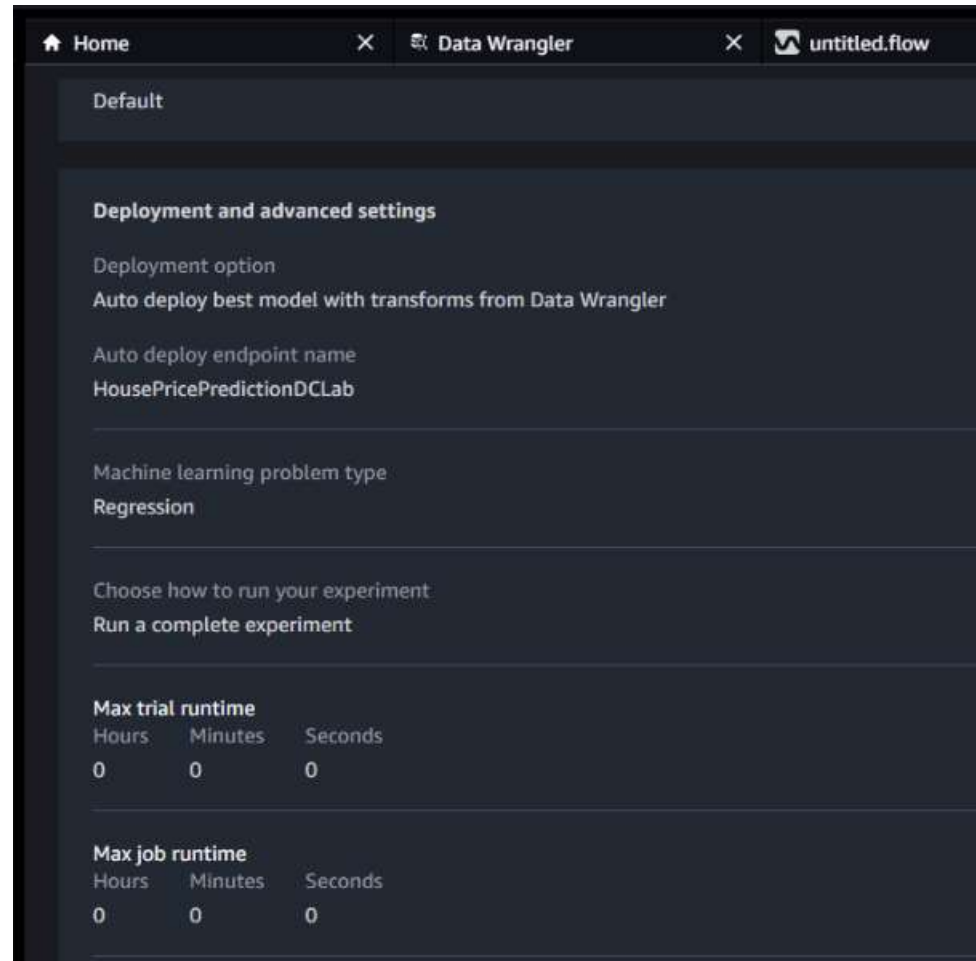
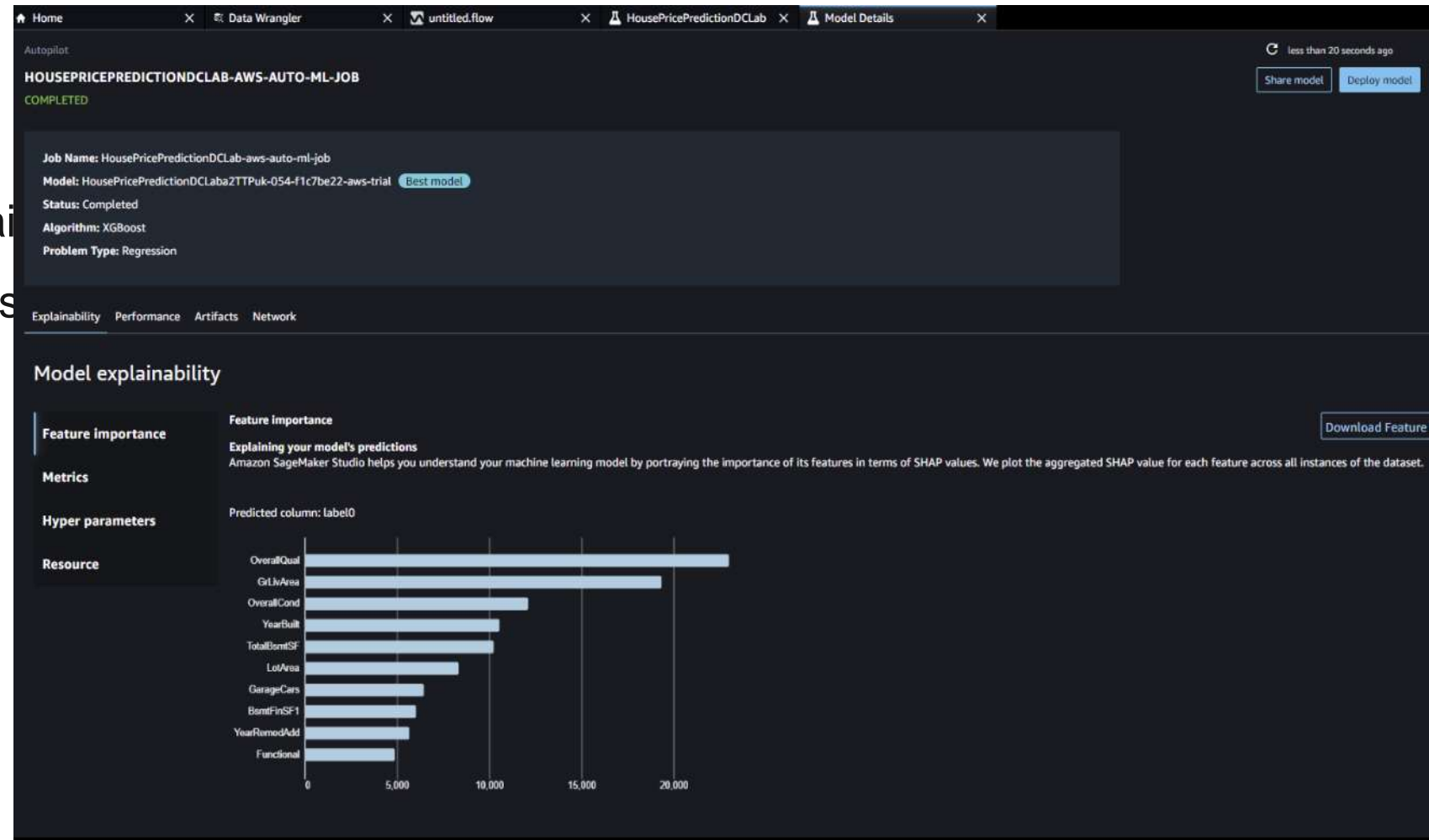- Create Experiment

- Training the Model

Example - Regression Models with SageMaker Autopilot

Example: Creating a Regression model using SageMaker Autopilot to predict house price

Step 4 : Create an Autopilot experiment (continued)

Inside SageMaker Studio

- Experiment made 100 trials

- Check Job profile tab for detai

- Best Trail - View model details

Example - Regression Models with SageMaker Autopilot

Example: Creating a Regression model using SageMaker Autopilot to predict house price

Step 5 : Deploy ML model

Inside SageMaker Studio

- Best Trail - Deploy model

- Configure the following details

- Wait for the endpoint

deploy

Example - Regression Models with SageMaker Autopilot

Example: Creating a Regression model using SageMaker Autopilot to predict house price

Step 6 : Invoke deployed ML model

Inside SageMaker Studio

- Create IAM User with permission

  boundary as AmazonSageMakerFullAccess

- For this user create Access Key

- Inside Postman add generated keys and do

  the configuration to call the endpoint

- Set Content-Type to text/csv.

- Paste any record from the test dataset

- Review the Predicted Price of the

  House
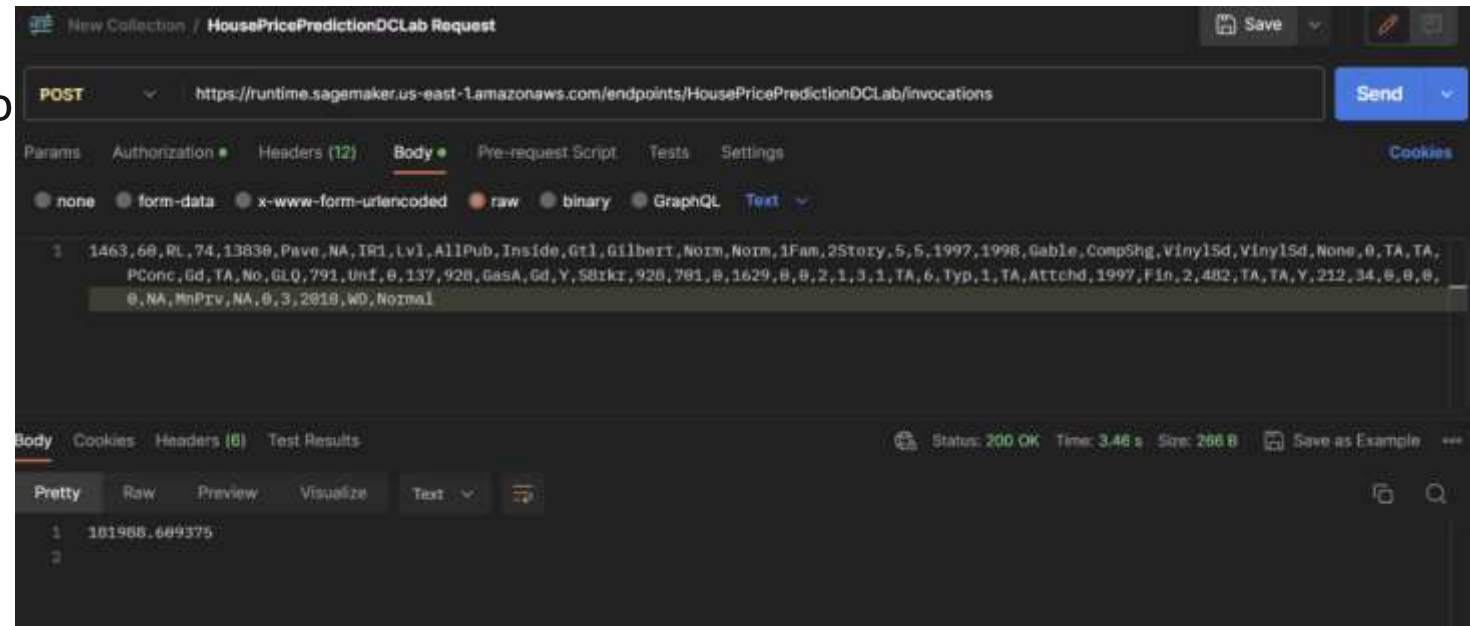
Improving Accuracy of the Regression Model

1. Collect data: Increase the number of training examples.

2. Feature processing: Add more variables and better feature processing.

3. Model parameter Tuning: Consider alternate values for the training parameters used by your learning algorithm

Refining a machine learning model that fits your needs often requires an iterative approach. You can achieve this by collecting more diverse data, improving feature processing, and carefully tuning model parameters.

Each iteration contributes to better understanding and fine-tuning, ultimately leading to a more predictive and effective machine learning model.

Regression and Accuracy Metrics

- Regression is a technique used to model the relationship between dependent and independent variables.

- Accuracy metrics in regression tasks include Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE).

- RMSE measures the distance between predicted and actual values, evaluating predictive accuracy.

- Amazon ML uses RMSE to assess the performance of regression models.

Analyzing Residuals in Regression Model

- Residuals are differences between true and predicted target values in evaluation data.

- Residuals reveal what portion of the target a model can't predict accurately.

- Positive residuals indicate underestimation, negative ones indicate overestimation.

- A bell-shaped histogram of residuals centered around zero signifies random prediction errors.

- Lack of a zero-centered bell shape suggests a pattern in the model's prediction errors.

- Structured residuals might indicate a need to include more variables to capture unaccounted patterns.
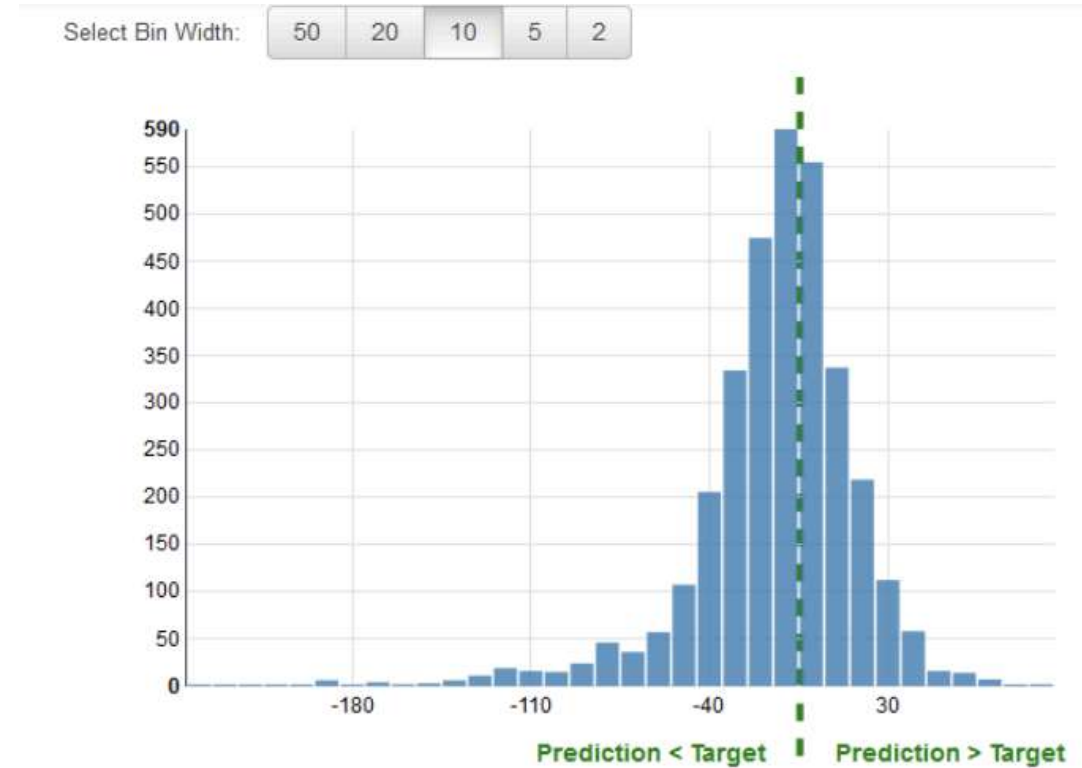

Figure 3: Distribution of residuals for a Regression model

Measuring ML Model Accuracy

Root Mean Square Error (RMSE) in Amazon ML for Regression:

- Amazon ML uses RMSE as the industry standard metric for regression tasks.

- RMSE measures the distance between predicted and actual numeric target values.

- Smaller RMSE values indicate better predictive accuracy.

- A perfect model's RMSE would be 0, signifying correct predictions.

$$RMSE = \sqrt{1/N \sum_{i=1}^{N} (actual\ target - predicted\ target)^2}$$

This example shows evaluation data that contains N records

Applications of Regression Model

Regression has several real-world applications in many different industries.

- **Manufacturing** – Probability of part failure.

- **Healthcare** – Predicting the likelihood of disease in patients

- **Finance** – Analyze financial transactions for fraud and assess loan applications and insurance applications for risk.

- **Marketing** – Analyze user responses to different words and images and create high-performing advertisements