Public Transportation Analysis

2023 Naan Mudhalvan - IBM Data Analytics with Cognos
Group 1 - Project 8
College : NM001 - College of Engineering Guindy
Proj_200340_Team_2

Members: Abinithi R, Abirami S V, Adithya R U, Akshaya G R,
Sai Rishi A N
Faculty Mentor : Dr. G Geetha

PHASE 2

INNOVATION

PROBLEM DEFINITION :

Analyse public transportation data to assess **service efficiency**, **on time performance**, and **passenger feedback.**
Provide insights that **support transportation improvement initiatives** and enhance the overall public transportation experience.

ANALYSIS STEPS

1. DATA PREPROCESSING

❖ Cleaning and Preprocessing the Dataset:

Handling Missing Values

Missing data can significantly affect the performance of machine learning models. There are several methods to handle missing values, including:

- ***Removing Rows:*** Rows with missing values can be removed, but this might result in losing valuable data.
- ***Filling with Mean/Median/Mode*:** Filling missing values with the mean (average), median (middle value), or mode (most frequent value) of the respective column.
- ***Advanced Imputation Techniques**:* Using advanced techniques such as K-nearest neighbors imputation or regression imputation to predict missing values based on other features.

Encoding Categorical Variables:

Machine learning algorithms generally work with numerical data. Categorical variables like RouteID and StopName need to be converted into numerical representations. Common methods include:

- ***Label Encoding:*** Assign a unique numerical label to each category. For example, for RouteID: {"RouteA": 1, "RouteB": 2, ...}.
- ***One-Hot Encoding:*** Creating binary columns for each category. For example, for RouteID, columns like RouteA, RouteB, etc., can be created with binary values indicating presence or absence.
- ***Hashing:*** Converting categories into hash values. This method is useful when dealing with a large number of categories.

❖ Creating Features:

The day of the week's information can be extracted from the 'WeekBeginning' attribute. This can be important for capturing patterns related to specific days (e.g., weekdays vs. weekends). This information can be also used to create a new feature with values like 1 for Monday, 2 for Tuesday, and so on.

❖ Splitting Data into Training and Testing Sets:

- **Training Set**: This portion of the dataset is used to train the machine learning model. It contains input features and their corresponding output labels. The model learns patterns from this data.
- **Testing Set:** This portion of the dataset is used to evaluate the performance of the trained model. It helps assess how well the model generalizes to new, unseen data. The testing set also contains input features and corresponding output labels, but the model has never seen these data points during training.

2. FEATURE ENGINEERING

Feature engineering is the process of selecting, transforming, or creating relevant features from raw data to enhance the performance of machine learning models. Well-engineered features can significantly improve a model's ability to learn patterns and make predictions. It can be performed in the following ways:

❖ Extracting Relevant Features:

Relevant features can be extracted in several ways:

- **Aggregation:** Calculating statistics like averages, sums, minimum, maximum, or standard deviations for numerical attributes. For example, the average number of boardings per day for each stop can be calculated. This aggregated information can provide valuable insights.

- **Temporal Features:** Extracting time-based features such as day of the week, month, year, or specific events like holidays can capture patterns related to specific time periods.

❖ **Feature Scaling and Normalization:**

It might be necessary to scale or normalize the features. Many machine learning algorithms are sensitive to the scale of features. Common techniques include Min-Max scaling and Z-score normalization.

❖ **Iterative Process:**

Feature engineering is often an iterative process. After creating initial features, it's essential to assess the model's performance. If the model is not performing well, revisiting and refining the features can lead to better results. Domain expertise and a good understanding of the problem domain play a crucial role in this iterative process.

3. LABEL GENERATION

Label Generation is a critical step in supervised machine learning, where the target variable (also known as the label or output) that the model will learn to predict is defined. For predicting service disruptions, a binary variable indicating whether a disruption has occurred (1) or not (0) can be created. The following steps are followed:

❖ **Defining the Disruption Criteria:**

- **Domain Knowledge:** Domain experts can be consulted to understand what constitutes a service disruption. Disruptions can vary based on the nature of the service, and experts can provide valuable insights into defining disruption criteria.
- **Identifying Patterns:** Analyzing historical data to identify patterns associated with disruptions and looking for features such as unusually low boardings, delays in schedules, incidents reported, or any other indicators that suggest service interruptions.
- **Thresholds:** Based on the analysis and expert consultation, thresholds can be set for the identified criteria. For example, a day with boardings significantly lower than the average may be considered as a disruption.

❖ **Creating the Binary Label:**

Once the disruption criteria and thresholds have been defined, create a binary label column in the dataset. Assign a value of 1 to instances (days, stops, routes, etc.) where the disruption criteria are met, and 0 to instances where there is no disruption. For instance, if a specific stop has boardings lower than a certain threshold on a given day, that day would be labelled with a 1, indicating a disruption. Days where boardings are above the threshold would be labeled as 0, indicating no disruption.

❖ **Ensuring Consistency:**

Consistency has to be ensured in applying the disruption criteria across the dataset. It is crucial that the criteria and thresholds are applied uniformly to all relevant instances. Consistency in

labeling is essential for the model to learn meaningful patterns.

- **Validation and Refinement:**

The labeled data should be validated by cross-referencing it with incident reports or other reliable sources of disruption information to ensure accuracy. If discrepancies are found or the model isn't performing as expected, the disruption criteria and labels should be refined iteratively based on feedback from the validation process.

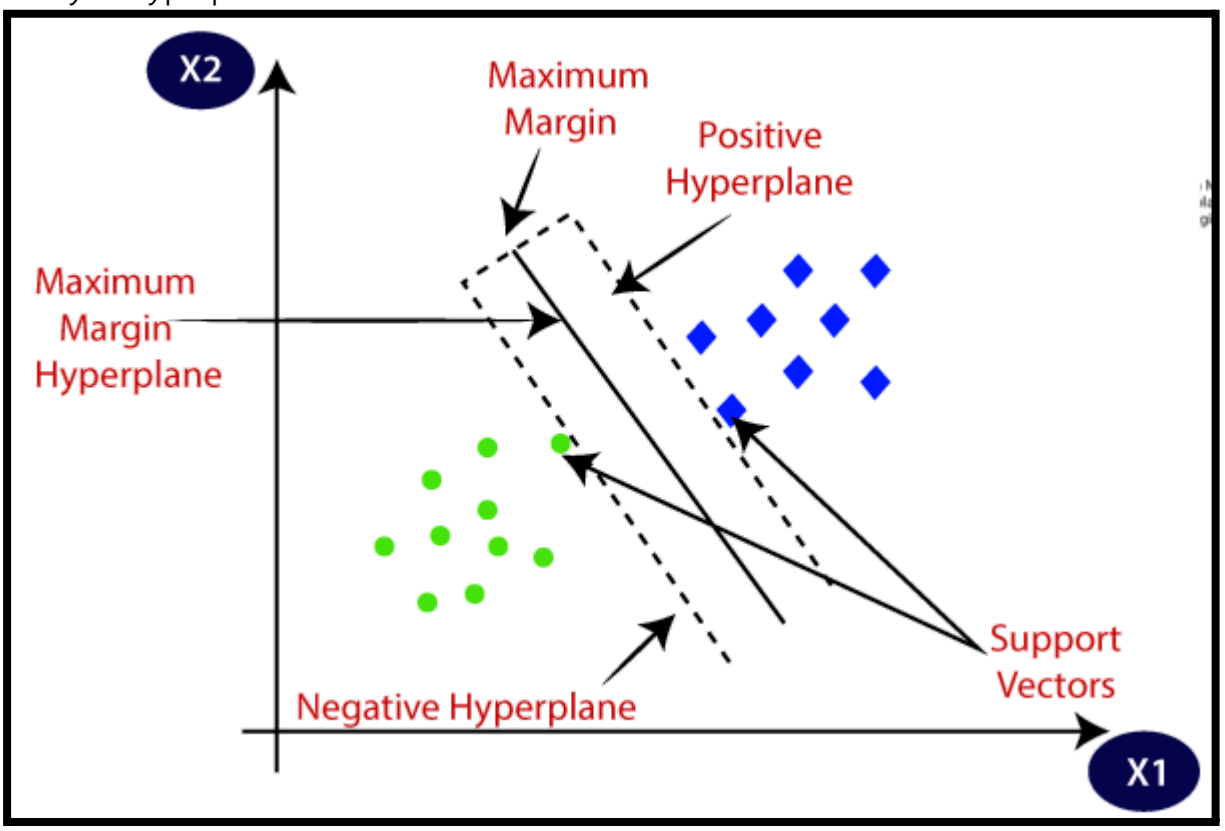
4. MODEL ARCHITECTURE

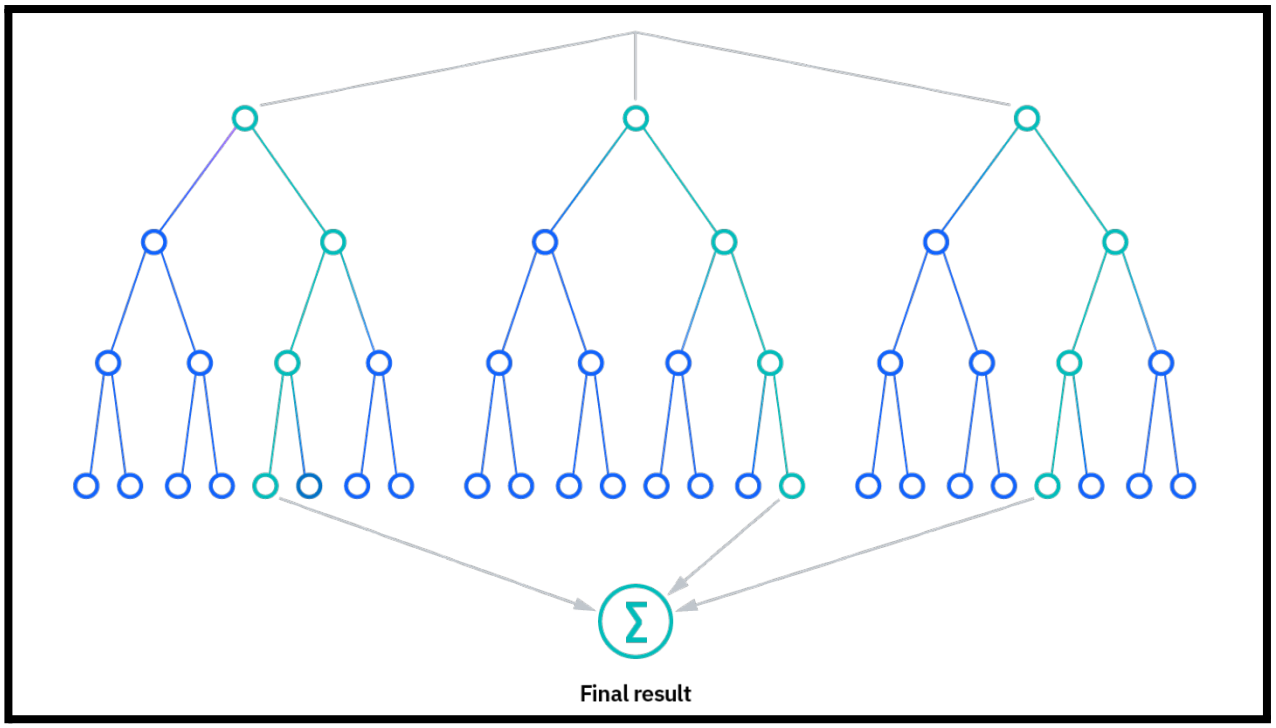
Support Vector Machine Algorithm

Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning.

The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane.

SVM chooses the extreme points/vectors that help in creating the hyperplane. These extreme cases are called as support vectors, and hence algorithm is termed as Support Vector Machine. Consider the below diagram in which there are two different categories that are classified using a decision boundary or hyperplane:



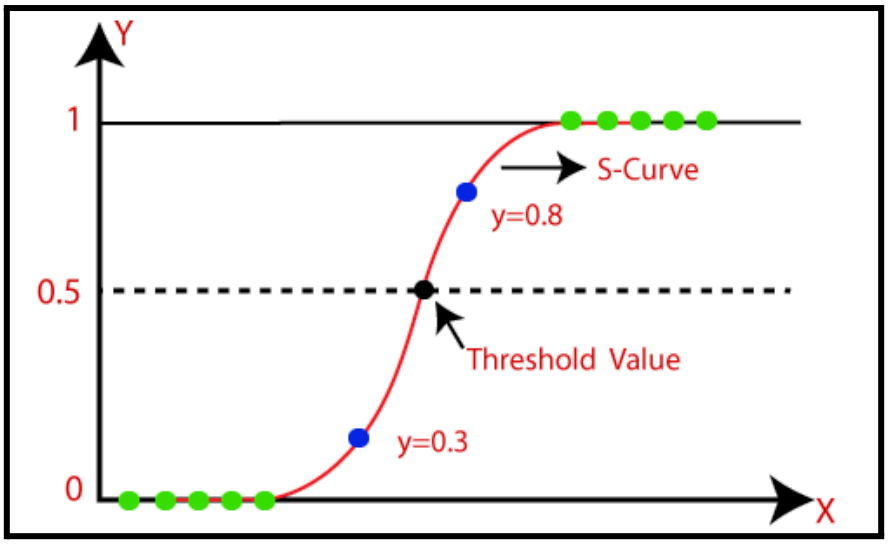


Types of SVM:

- **Linear SVM:** Linear SVM is used for linearly separable data, which means if a dataset can be classified into two classes by using a single straight line, then such data is termed as linearly separable data, and classifier is used called as Linear SVM classifier.
- **Non-linear SVM:** Non-Linear SVM is used for non-linearly separated data, which means if a dataset cannot be classified by using a straight line, then such data is termed as non-linear data and classifier used is called as Non-linear SVM classifier

Logistic Regression

- Logistic regression is one of the most popular Machine Learning algorithms, which comes under the Supervised Learning technique. It is used for predicting the categorical dependent variable using a given set of independent variables.
- Logistic regression predicts the output of a categorical dependent variable. Therefore the outcome must be a categorical or discrete value. It can be either Yes or No, 0 or 1, true or False, etc. but instead of giving the exact value as 0 and 1, **it gives the probabilistic values which lie between 0 and 1**.
- Logistic Regression is much similar to the Linear Regression except that how they are used. Linear Regression is used for solving Regression problems, whereas **Logistic regression is used for solving the classification problems**.
- In Logistic regression, instead of fitting a regression line, we fit an "S" shaped logistic function, which predicts two maximum values (0 or 1).
- The curve from the logistic function indicates the likelihood of something such as whether the cells are cancerous or not, a mouse is obese or not based on its weight, etc.
- Logistic Regression is a significant machine learning algorithm because it has the ability to provide probabilities and classify new data using continuous and discrete datasets.
- Logistic Regression can be used to classify the observations using different types of data and can easily determine the most effective variables used for the classification. The below image is showing the logistic function:



Logistic Function (Sigmoid Function):

- The sigmoid function is a mathematical function used to map the predicted values to probabilities.
- It maps any real value into another value within a range of 0 and 1.
- The value of the logistic regression must be between 0 and 1, which cannot go beyond this limit, so it forms a curve like the "S" form. The S-form curve is called the Sigmoid function or the logistic function.
- In logistic regression, we use the concept of the threshold value, which defines the probability of either 0 or 1. Such as values above the threshold value tends to 1, and a value below the threshold values tends to 0.

Assumptions for Logistic Regression:

- The dependent variable must be categorical in nature.
- The independent variable should not have multi-collinearity.

Logistic Regression Equation:

The Logistic regression equation can be obtained from the Linear Regression equation. The mathematical steps to get Logistic Regression equations are given below:

- We know the equation of the straight line can be written as:

$$y= b_0 + b_1x_1 + b_2x_2 + b_3x_3 + \cdots + b_nx_n$$

- In Logistic Regression y can be between 0 and 1 only, so for this let's divide the above equation by (1-y):

$$\frac{y}{1-y}\text{; 0 for y= 0, and infinity for y=1}$$

- But we need range between -[infinity] to +[infinity], then take logarithm of the equation it will become:

$$log\left[\frac{y}{1-y}\right] = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + \cdots + b_nx_n$$

The above equation is the final equation for Logistic Regression.

Types of Logistic Regression:

On the basis of the categories, Logistic Regression can be classified into three types:

- **Binomial:** In binomial Logistic regression, there can be only two possible types of the dependent variables, such as 0 or 1, Pass or Fail, etc.
- **Multinomial:** In multinomial Logistic regression, there can be 3 or more possible unordered types of the dependent variable, such as "cat", "dogs", or "sheep"
- **Ordinal:** In ordinal Logistic regression, there can be 3 or more possible ordered types of dependent variables, such as "low", "Medium", or "High".

5. MODEL EVALUATION

❖ Performance Metrics for Classification

In a classification problem, the category or classes of data is identified based on training data. The model learns from the given dataset and then classifies the new data into classes or groups based on the training. It predicts class labels as the output, such as *Yes or No, 0 or 1, Spam or Not Spam,* etc. To evaluate the performance of a classification model, different metrics are used, and some of them are as follows:

- **Accuracy**
- **Confusion Matrix**
- **Precision**
- **Recall**
- **F-Score**
- **AUC(Area Under the Curve)-ROC**

I. Accuracy

The accuracy metric is one of the simplest Classification metrics to implement, and it can be determined as the number of correct predictions to the total number of predictions.
It can be formulated as:

$$Accuracy = \frac{Number\ of\ Correct\ Predictions}{Total\ number\ of\ predictions}$$

To implement an accuracy metric, we can compare ground truth and predicted values in a loop, or we can also use the scikit-learn module for this. Although it is simple to use and implement, it is suitable only for cases where an equal number of samples belong to each class.

When to Use Accuracy?

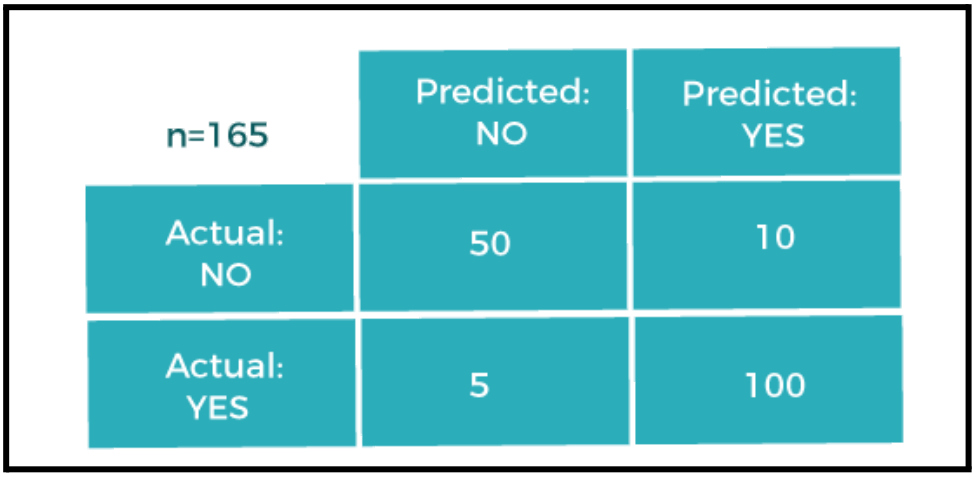
It is good to use the Accuracy metric when the target variable classes in data are approximately balanced. For example, if 60% of classes in a fruit image dataset are of Apple, 40% are Mango. In this case, if the model is asked to predict whether the image is of Apple or Mango, it will give a prediction with 97% of accuracy.

When not to use Accuracy?

It is recommended not to use the Accuracy measure when the target variable majorly belongs to one class. For example, Suppose there is a model for a disease prediction in which, out of 100 people, only five people have a disease, and 95 people don't have one. In this case, if our model predicts every person with no disease (which means a bad prediction), the Accuracy measure will be 95%, which is not correct.

II. Confusion Matrix

A confusion matrix is a tabular representation of prediction outcomes of any binary classifier, which is used to describe the performance of the classification model on a set of test data when true values are known. A typical confusion matrix for a binary classifier looks like the below image(However, it can be extended to use for classifiers with more than two classes).



n=165	Predicted: NO	Predicted: YES
Actual: NO	50	10
Actual: YES	5	100

We can determine the following from the above matrix:

- In the matrix, columns are for the prediction values, and rows specify the Actual values. Here Actual and prediction give two possible classes, Yes or No. So, if we are predicting the presence of a disease in a patient, the Prediction column with Yes means, Patient has the disease, and for NO, the Patient doesn't have the disease.
- In this example, the total number of predictions are 165, out of which 110 time predicted yes, whereas 55 times predicted No.
- However, in reality, 60 cases in which patients don't have the disease, whereas 105 cases in which patients have the disease.

In general, the table is divided into four terminologies, which are as follows:

1. **True Positive(TP):** In this case, the prediction outcome is true, and it is true in reality, also.
2. **True Negative(TN):** in this case, the prediction outcome is false, and it is false in reality, also.
3. **False Positive(FP):** In this case, prediction outcomes are true, but they are false in actuality.
4. **False Negative(FN):** In this case, predictions are false, and they are true in actuality.

III. Precision

The precision metric is used to overcome the limitation of Accuracy. The precision determines the proportion of positive prediction that was actually correct. It can be calculated as the True Positive or predictions that are actually true to the total positive predictions (True Positive and False Positive).

$$Precision = \frac{TP}{(TP + FP)}$$

IV. Recall or Sensitivity

It is also similar to the Precision metric; however, it aims to calculate the proportion of actual positive that was identified incorrectly. It can be calculated as True Positive or predictions that are actually true to the total number of positives, either correctly predicted as positive or incorrectly predicted as negative (true Positive and false negative).

The formula for calculating Recall is given below:

$$\text{Recall} = \frac{TP}{TP+FN}$$

When to use Precision and Recall?
From the above definitions of Precision and Recall, we can say that recall determines the performance of a classifier with respect to a false negative, whereas precision gives information about the performance of a classifier with respect to a false positive.
So, if we want to minimize the false negative, then, Recall should be as near to 100%, and if we want to minimize the false positive, then precision should be close to 100% as possible.
In simple words, *if we maximize precision, it will minimize the FP errors, and if we maximize recall, it will minimize the FN error.*

V. F-Scores

F-score or F1 Score is a metric to evaluate a binary classification model on the basis of predictions that are made for the positive class. It is calculated with the help of Precision and Recall. It is a type of single score that represents both Precision and Recall. So, ***the F1 Score can be calculated as the harmonic mean of both precision and Recall, assigning equal weight to each of them.***
The formula for calculating the F1 score is given below:

$$F1 - score = 2 * \frac{precision * recall}{precision + recall}$$

When to use F-Score?
As F-score make use of both precision and recall, so it should be used if both of them are important for evaluation, but one (precision or recall) is slightly more important to consider than the other. For example, when False negatives are comparatively more important than false positives, or vice versa.

VI. AUC-ROC

Sometimes we need to visualize the performance of the classification model on charts; then, we can use the AUC-ROC curve. It is one of the popular and important metrics for evaluating the performance of the classification model.
ROC represents a graph to show the performance of a classification model at different threshold levels. The curve is plotted between two parameters, which are:

- **True Positive Rate**
- **False Positive Rate**

TPR or true Positive rate is a synonym for Recall, hence can be calculated as:

$$TPR = \frac{TP}{TP + FN}$$

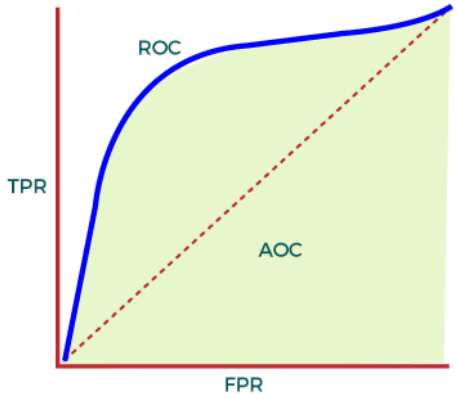
FPR or False Positive Rate can be calculated as:

$$TPR = \frac{FP}{FP + TN}$$

To calculate value at any point in a ROC curve, we can evaluate a logistic regression model multiple times with different classification thresholds, but this would not be much efficient. So, for this, one efficient method is used, which is known as AUC.

AUC: Area Under the ROC curve

AUC is known for **Area Under the ROC curve**. As its name suggests, AUC calculates the two-dimensional area under the entire ROC curve, as shown below image:



AUC calculates the performance across all the thresholds and provides an aggregate measure. The value of AUC ranges from 0 to 1. It means a model with 100% wrong prediction will have an AUC of 0.0, whereas models with 100% correct predictions will have an AUC of 1.0.

When to Use AUC

AUC should be used to measure how well the predictions are ranked rather than their absolute values. Moreover, it measures the quality of predictions of the model without considering the classification threshold.

When not to use AUC

As AUC is scale-invariant, which is not always desirable, and we need calibrating probability outputs, then AUC is not preferable.

Further, AUC is not a useful metric when there are wide disparities in the cost of false negatives vs. false positives, and it is difficult to minimize one type of classification error.

❖ Performance Metrics for Regression

Regression is a supervised learning technique that aims to find the relationships between the dependent and independent variables. A predictive regression model predicts a numeric or discrete value. The metrics used for regression are different from the classification metrics. It means we cannot use the Accuracy metric (explained above) to evaluate a regression model; instead, the performance of a Regression model is reported as errors in the prediction. Following are the popular metrics that are used to evaluate the performance of Regression models.

- **Mean Absolute Error**
- **Mean Squared Error**
- **R2 Score**
- **Adjusted R2**

I. Mean Absolute Error (MAE)

Mean Absolute Error or MAE is one of the simplest metrics, which measures the absolute difference between actual and predicted values, where absolute means taking a number as Positive.

To understand MAE, let's take an example of Linear Regression, where the model draws a best fit line between dependent and independent variables. To measure the MAE or error in prediction, we need to calculate the difference between actual values and predicted values. But in order to find the absolute error for the complete dataset, we need to find the mean absolute of the complete dataset.

The below formula is used to calculate MAE:

$$MAE = 1/N \sum |Y - Y'|$$

Here,
Y is the Actual outcome, Y' is the predicted outcome, and N is the total number of data points.
MAE is much more robust for the outliers. One of the limitations of MAE is that it is not differentiable, so for this, we need to apply different optimizers such as Gradient Descent. However, to overcome this limitation, another metric can be used, which is Mean Squared Error or MSE.

II. Mean Squared Error

Mean Squared error or MSE is one of the most suitable metrics for Regression evaluation. It measures the average of the Squared difference between predicted values and the actual value given by the model.
Since in MSE, errors are squared, therefore it only assumes non-negative values, and it is usually positive and non-zero.
Moreover, due to squared differences, it penalizes small errors also, and hence it leads to over-estimation of how bad the model is.
MSE is a much-preferred metric compared to other regression metrics as it is differentiable and hence optimized better.
The formula for calculating MSE is given below:

$$MSE = 1/N \sum (Y - Y')^2$$

Here,
Y is the Actual outcome, Y' is the predicted outcome, and N is the total number of data points.

III. R Squared Score

R squared error is also known as Coefficient of Determination, which is another popular metric used for Regression model evaluation. The R-squared metric enables us to compare our model with a constant baseline to determine the performance of the model. To select the constant baseline, we need to take the mean of the data and draw the line at the mean.
The R squared score will always be less than or equal to 1 without concerning if the values are too large or small.

$$R^2 = 1 - \frac{MSE(Model)}{MSE(Baseline)}$$

IV. Adjusted R Squared

Adjusted R squared, as the name suggests, is the improved version of R squared error. R square has a limitation of improvement of a score on increasing the terms, even though the model is not improving, and it may mislead the data scientists.
To overcome the issue of R square, adjusted R squared is used, which will always show a lower value than R^2. It is because it adjusts the values of increasing predictors and only shows improvement if there is a real improvement.
We can calculate the adjusted R squared as follows:

$$R_a^2 = 1-\left[\left(\frac{n-1}{n-k-1}\right) \times \left(1 - R^2\right)\right]$$

Here,
n is the number of observations
k denotes the number of independent variables
and R_a^2 denotes the adjusted R^2

6. TUNING AND OPTIMIZATION

The configuration and hyperparameter tuning can profoundly influence a model's performance. A slight tweak can be the difference between a mediocre outcome and stellar results. For instance, the Adam optimizer, a popular optimization method in deep learning, has specific hyperparameters that, when fine-tuned, can lead to faster and more stable convergence during training.

In real-world applications, hyperparameter search and fine-tuning become even more evident. Consider a scenario where a pre-trained neural network, initially designed for generic image recognition, is repurposed for a specialized task like medical image analysis. Its accuracy and reliability can be significantly enhanced by searching for optimal hyperparameters and fine-tuning them for this dataset. This could mean distinguishing between accurately detecting a medical anomaly and missing it altogether.

Furthermore, as machine learning evolves, our datasets and challenges become more complex. In such a landscape, the ability to fine-tune models and optimize hyperparameters using various optimization methods is not just beneficial; it's essential. It ensures that our models are accurate, efficient, adaptable, and ready to tackle the challenges of tomorrow.

HYPERPARAMETER OPTIMIZATION

Hyperparameter optimization focuses on finding the optimal set of hyperparameters for a given model. Unlike model parameters, these hyperparameters are not learned during training but are set before the training begins. Their correct setting can significantly influence the model's performance.

Grid Search

Grid Search involves exhaustively trying out every possible combination of hyperparameters in a predefined search space. For instance, if you're fine-tuning a model and considering two hyperparameters, learning rate and batch size, a grid search would test all combinations of the values you specify for these hyperparameters.

For an SVM applied to this problem, two critical hyperparameters are:

- The type and parameters of the kernel: For instance, if using the Radial Basis Function (RBF) kernel, we need to determine the **gamma value**.
- The **regularization parameter (C)** determines the trade-off between maximizing the margin and minimizing classification error.

By training the SVM with each combination and validating its performance on a separate dataset, grid search allows us to pinpoint the combination that yields the best classification accuracy.

Random Search

Random Search, as the name suggests, involves randomly selecting and evaluating combinations of hyperparameters. Unlike Grid Search, which exhaustively tries every possible combination, Random Search samples a predefined number of combinations from a specified distribution for each hyperparameter.

Consider a scenario where a financial institution develops a machine learning model to predict loan defaults. The dataset is vast, with numerous features ranging from a person's credit history to current financial status.

The model has several hyperparameters like learning rate, batch size, and the number of layers. Given the high dimensionality of the hyperparameter space, using Grid Search might be

computationally expensive and time-consuming. By randomly sampling hyperparameter combinations, the institution can efficiently narrow down the best settings with the highest prediction accuracy, saving time and computational resources.

Bayesian Optimization

Bayesian Optimization is a probabilistic model-based optimization technique particularly suited for optimizing expensive-to-evaluate and noisy functions. Unlike random or grid search, Bayesian Optimization builds a probabilistic model of the objective function. It uses it to select the most promising hyperparameters to evaluate the true objective function.

Bayesian Optimization shines in scenarios where the objective function is expensive to evaluate. For instance, training a model with a particular set of hyperparameters in deep learning can be time-consuming. Using grid search or random search in such scenarios can be computationally prohibitive. By building a model of the objective function, Bayesian Optimization can more intelligently sample the hyperparameter space to find the optimal set-in fewer evaluations.

Bayesian Optimization is more directed than grid search, which exhaustively tries every combination of hyperparameters, or random search, which samples them randomly. It uses past evaluation results to choose the next set of hyperparameters to evaluate. This makes it particularly useful when evaluating the objective function is time-consuming or expensive.

7. DEPLOYMENT

Once an ML model is ready, the next step is to make it available for users to make predictions. Placing an ML model in an environment for users to interact with and use for decision making refers to model deployment.

The Process of Deploying Machine Learning Models

ML model deployment involves the following steps:

1. **Develop, create, and test the model in a training environment:** This step requires rigorous training, testing, and optimization of the model to ensure high performance in production. The model training step determines how models perform in production. ML teams must collaborate to optimize, clean, test, and retest model code.
2. **Movement of models to deployment environment:** After rigorous testing and optimizing model code, the top-performing models undergo preparation for deployment. Models need a deployment environment that contains all the hardware resources and data required to make the model perform optimally. Different deployment environments include:
    - **Containers:** Most teams use a container deploying environment because containers are reproducible, predictable, and easy to modify and update, making collaboration among engineers easy. Containers encompass all the hardware, configurations and dependencies necessary to deploy the model, improving consistency among ML teams.
    - **Notebooks:** Jupyter and AWS Sagemaker are common notebooks used by data scientists for experimentation in the ML lifecycle. However, notebooks present difficulties like reproducibility and testing for teams. To efficiently use notebooks in the production workflow, teams should consider code organization, reusability, and dependencies, among other factors.
    - **In-App environments:** This environment works when certain limitations or constraints exist around using data outside the application.

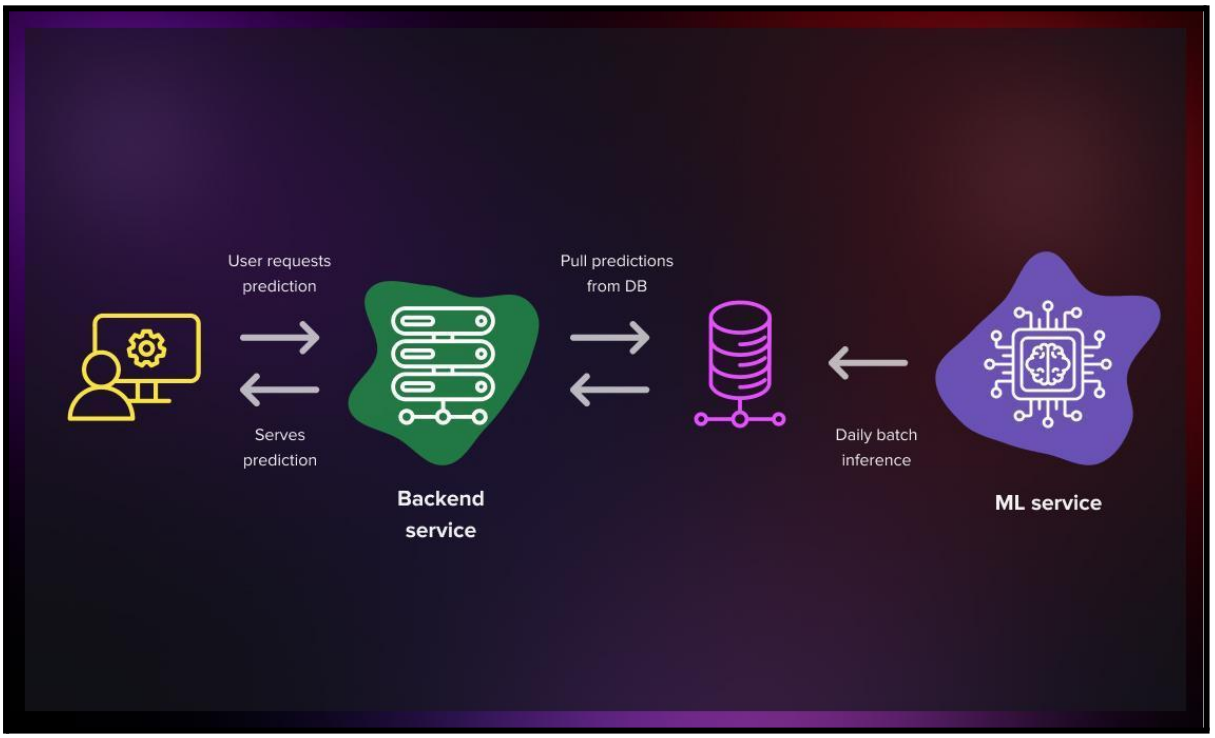
3. **Making models available for end users:** ML teams must choose how to make models available for their users. Most can be available on demand or deployed to edge devices.

4. **Monitoring:** The ML lifecycle continues after deployment. Deployed models must undergo constant monitoring to evaluate the performance and accuracy of models over time. Because data is in a continual state of motion and change, model degradation may occur. In this case, automating the ML workflow to monitor and retrain models constantly helps ensure the longevity of models.

Which ML model deployment methods to use?

You choose deployment methods depending on the specific task at hand and the business problem you need to solve. Below, we look at the main methods.

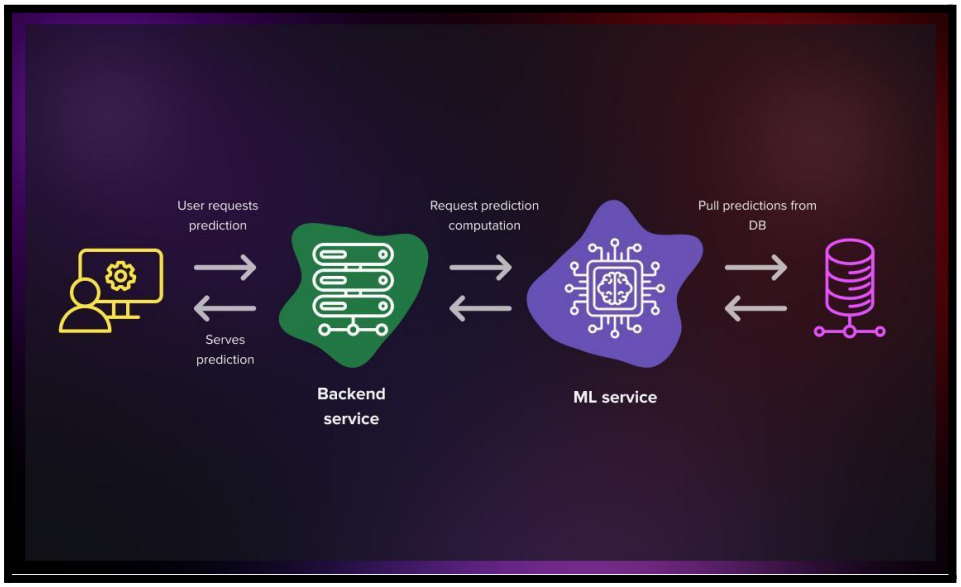
Batch deployment



This approach is suitable for scenarios where data is collected over a period of time and processed offline in larger batches.

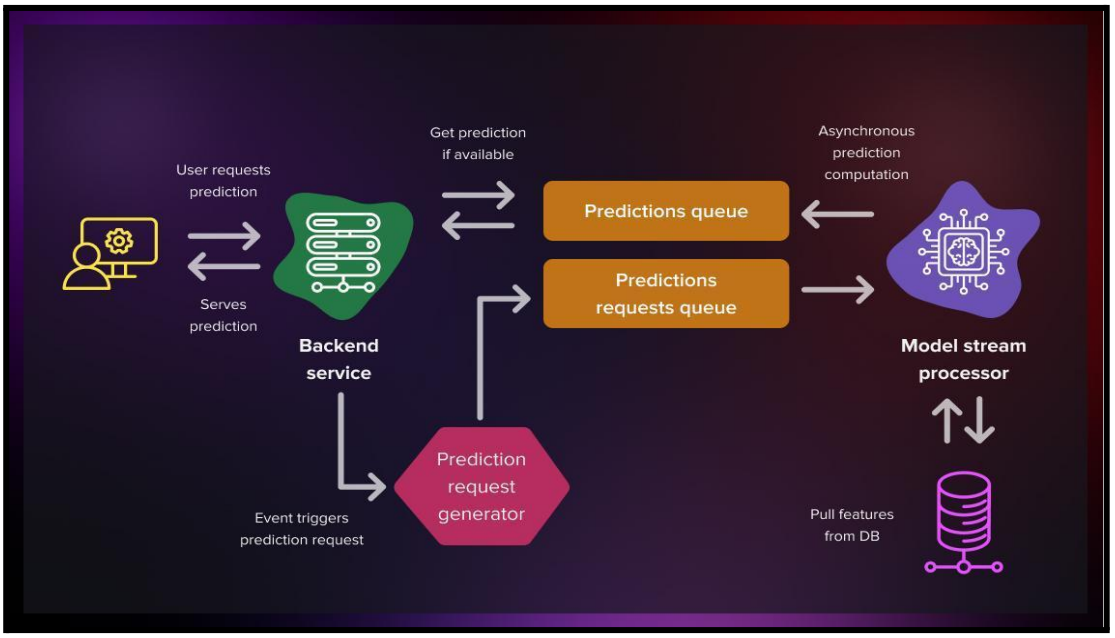
- The mode processes a batch of predictions daily; once completed, they are sent back to the service.
- With this method, the data may become outdated, but it will never be more outdated than the last processed batch, which is generally acceptable for most cases.

Real time



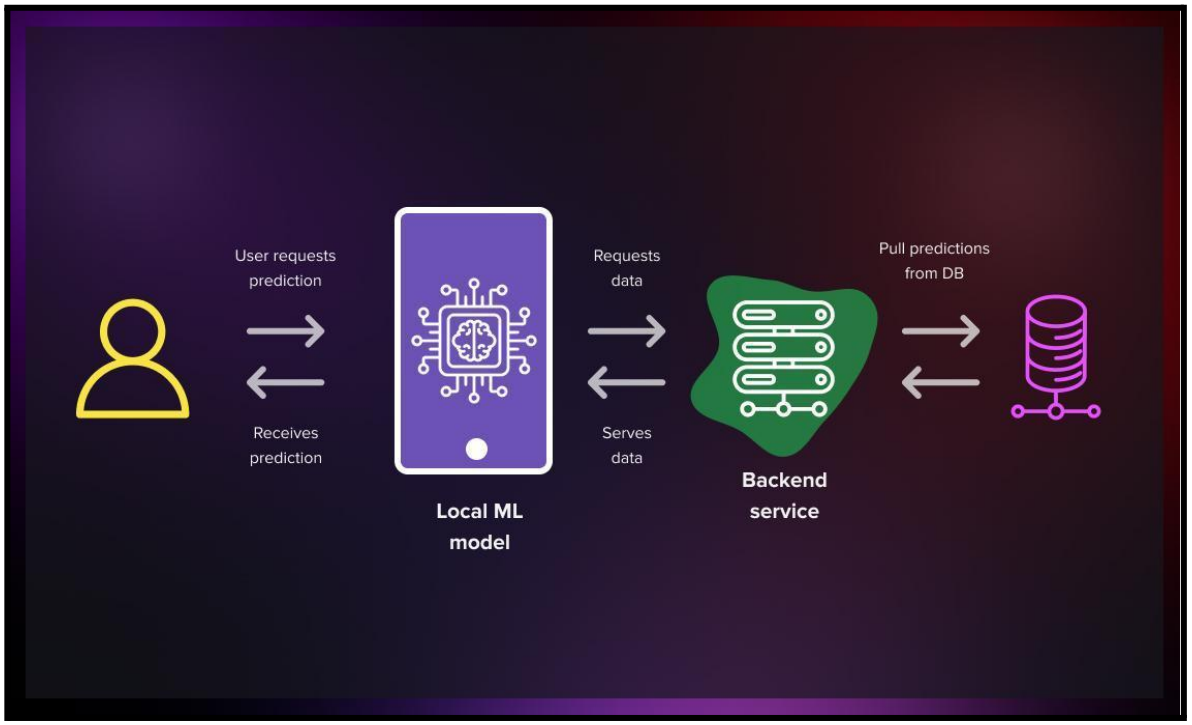
- This method is particularly effective for generating personalized predictions considering recent contextual information, such as the time of day or recent user search queries.
- To accommodate simultaneous requests from multiple users, you need to employ multi-threaded processes and vertical scaling by adding more servers.

Streaming deployment



- Streaming deployment facilitates a more asynchronous process, allowing user actions to initiate prediction computations. This is the principle on which most recommender systems are based.
- To achieve this, the process is integrated into a message broker. The machine learning model processes the request when it is ready.
- This approach reduces the server's processing burden and optimizes computational resources through an efficient queuing mechanism. Additionally, the prediction results can be placed in the queue, and the server can use them as needed.

Edge deployment



- Edge deployment enables faster model results and offline predictions.
- The model can be deployed directly on the client device, for example, a smartphone or IoT device.
- To achieve this, models are usually required to be small enough to fit on a compact device

8. INTERPRETABILITY

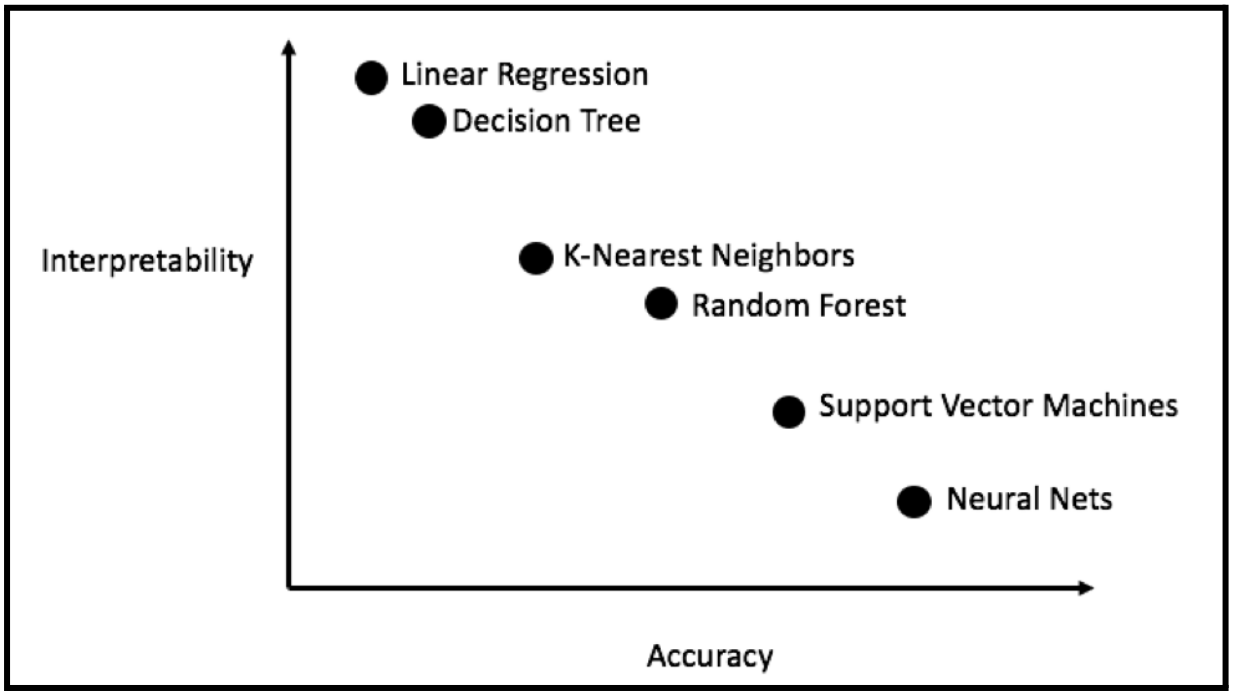
Interpretability is the degree to which a human can understand the cause of a decision. The higher the interpretability of an ML model, the easier it is to comprehend the model's predictions. Interpretability facilitates:

- Understanding
- Debugging and auditing ML model predictions
- Bias detection to ensure fair decision making
- Robustness checks to ensure that small changes in the input do not lead to large changes in the output
- Methods that provide recourse for those who have been adversely affected by model predictions

Model interpretability provides a mechanism to ensure the safety and effectiveness of ML solutions by increasing the transparency around model predictions, as well as the behavior of the underlying algorithm. Promoting transparency is a key aspect of the patient-centered approach, and is especially important for AI/ML-based SaMD, which may learn and change over time.

There is a tradeoff between *what* the model has predicted (model performance) and *why* the model has made such a prediction (model interpretability).

For some solutions, a high model performance is sufficient; in others, the ability to interpret the decisions made by the model is key. The demand for interpretability increases when there is a large cost for incorrect predictions, especially in high-risk applications.



Trade-off between performance and model interpretability

Based on the model complexity, methods for model interpretability can be classified into *intrinsic analysis* and *post hoc* analysis.

- **Intrinsic analysis** can be applied to interpret models that have low complexity (simple relationships between the input variables and the predictions). These models are based on:
    - Algorithms, such as linear regression, where the prediction is the weighted sum of the inputs
    - Decision trees, where the prediction is based on a set of if-then rules

The simple relationship between the inputs and output results in high model interpretability, but often leads to lower model performance, because the algorithms are unable to capture complex non-linear interactions.

- **Post hoc analysis** can be applied to interpret simpler models, as described earlier, as well as more complex models, such as neural networks, which have the ability to capture non-linear interactions. These methods are often model-agnostic and provide mechanisms to interpret a trained model based on the inputs and output predictions. Post hoc analysis can be performed at a *local* level, or at a *global* level.
    - **Local methods** enable you to zoom in on a single data point and observe the behavior of the model in that neighborhood. They are an essential component for debugging and auditing ML model predictions. Examples of local methods include:
    - **Local Interpretable Model-Agnostic Explanations** (LIME), which provides a sparse, linear approximation of the model behavior around a data point
    - **SHapley Additive exPlanations** (SHAP), a game theoretic approach based on Shapley values which computes the marginal contribution of each input variable towards the output

    - **Counterfactual explanations**, which describe the smallest change in the input variables that causes a change in the model's prediction
    - **Integrated gradients**, which provide mechanisms to attribute the model's prediction to specific input variables
    - **Saliency maps**, which are a pixel attribution method to highlight relevant pixels in an image
- **Global methods** enable you to zoom out and provide a holistic view that explains the overall behavior of the model. These methods are helpful for verifying that the model is robust and has the least possible bias to allow for fair decision making. Examples of global methods include:
    - **Aggregating local explanations**, as defined previously, across multiple data points
    - **Permutation feature importance**, which measures the importance of an input variable by computing the change in the model's prediction due to permutations of the input variable
    - **Partial dependence plots**, which plot the relationship and the marginal effect of an input variable on the model's prediction
    - **Surrogate methods**, which are simpler interpretable models that are trained to approximate the behavior of the original complex model

It is recommended to start the ML journey with a simple model that is both inherently interpretable and provides sufficient model performance. In later iterations, if you need to improve the model performance, AWS recommends increasing the model complexity and leveraging post hoc analysis methods to interpret the results.

Selecting both a local method and a global method gives you the ability to interpret the behavior of the model for a single data point, as well as across all data points in the dataset. It is also essential to validate the stability of model explanations, because methods in post-hoc analysis are susceptible to adversarial attacks, where small perturbations in the input could result in large changes in the output prediction and therefore in the model explanations as well.

9. FEEDBACK LOOP

In AI, machines learn how to execute tasks that are typically performed by humans. Like humans, AI systems make mistakes during their infancy and need a feedback loop to confirm or invalidate their decisions.

Feedback loops allow AI systems to know what they did right or wrong, giving them data that enables them to adjust their parameters to perform better in the future. In the C3 AI Reliability application, operators can prioritize maintenance actions based on risk scores and trigger work orders. If users disagree with the application's recommendations, they can log their decisions to help the system do better next time.

AI systems need to adapt to evolving data or new patterns that appear over time. A feedback loop reinforces the model's training with fresh data. In the C3 AI Anti-Money-Laundering application, it is crucial to incorporate the latest typologies and theft modes using a closed-loop workflow to improve predictions.