**INTERSHIP REPORT**

**Date:16-05-2025**

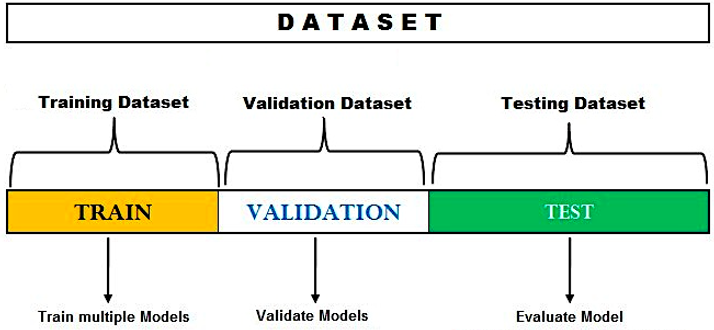
1. **AI MODEL VALIDATION:**

AI Model validation is a process of evaluating the model's performance, which helps to validate how well the model performs on unseen data.

**Types of Model Validation:**

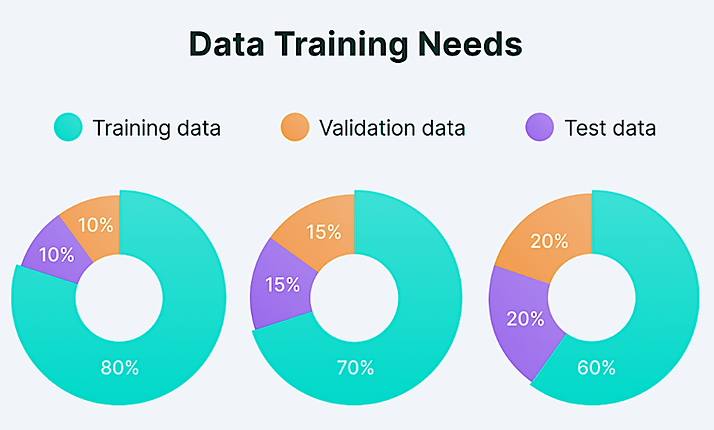
1. **Holdout Method:** The dataset is divided into training and testing sets, used to evaluate the performance of the model.

This method helps us to find the overfitting of the models.



1. **Out-of-Sample Validation**: This approach provides more reliable results to assess the model’s predictions on unseen data.

* **K-Fold Cross-validation:**The data is divided into k number of folds. The model is trained on k-1 folds and tested on the fold that is left. This is repeated k times, each time using a different fold for testing. This offers a more extensive analysis than the holdout method.
* **Leave-One-Out Cross-validation (LOOCV):**This is a form of k-fold cross-validation where k is equal to the number of instances. Only one piece of data is not used to train the model. This is repeated for each data point. Unfortunately, LOOCV is also time consuming when dealing with large datasets.
* **Stratified K-Fold Cross-validation:** k-fold cross-validation in this type of cross-validation each fold has the same ratio of classes/categories as the overall dataset. This is useful, especially where data in one class is very low compared to others.



1. **Ensemble Techniques**

Ensemble methods improve model performance by combining predictions from multiple models using three main approaches: bagging, boosting, and stacking.

Ensemble methods improve model performance by combining predictions from multiple models using three main approaches: bagging, boosting, and stacking.

* **Bagging -** Reduces variance by averaging predictions from models trained on different data subsets. Random Forests, for example, aggregate outputs from multiple decision trees for stability and accuracy.
* **Boosting -** Sequentially trains models to correct errors from earlier ones, enhancing accuracy over iterations. Techniques like Gradient Boosting and AdaBoost excel in high-stakes scenarios requiring precision.
* **Stacking -** Combines predictions from diverse base models (e.g., decision trees, neural networks) using a meta-model, such as linear regression, to optimize the final prediction. Stacking leverages complementary strengths of different models, making it particularly effective for complex datasets.

1. **Metrics:**

**Regression Metrics:**

Mean Squared Error (MSE), Root Mean Squared (RMSE), Mean Absolute Error (MAE)

**Classification Metrics:**

Accuracy, precision, and recall scores.

1. **AI Model Security**

It is the process of using AI to enhance the organisation's security, which can automate threat detection’s prevention and remediation to prevent any data breach.

**Steps to use AI to protect sensitive information**

* 1. **Detect Malicious inputs using AI Models:**In this process, a model or a classifier is trained to detect whether the data provided by the user is malicious or not.

**Ex:** Random Forest Algorithm is trained to detect patterns to detect attacks. SQL injection, Phishing attack, etc

* 1. **Access Control Mechanism:**

AI can enforce an access control mechanism to analyse the behaviour of the users based on Location device it blocks the users if any malicious activity is found.

COLAB FILES:

<https://colab.research.google.com/drive/1q2xrb3QUNoqtv32ENfZpi2SVMEuuFYnx?usp=sharing> – Anomaly Detection using NSL-KDD DATASET

<https://colab.research.google.com/drive/12tGdcEnG65RuamodSCVgECx92_jpy6Gf?usp=sharing> – Implemented how overfitting occurs in the model and computed the cross validation scores.

**References:**

[**What Is Model Validation, and 12 Common Methods to Get it Right**](https://www.citrusx.ai/post/what-is-model-validation-and-12-common-methods-to-get-it-right)

[**Validating Machine Learning Models: A Detailed Overview**](https://www.markovml.com/blog/ml-model-validation)

[**What is AI security? | IBM**](https://www.ibm.com/think/topics/ai-security)

[**AI Validation and Security**](https://chatgpt.com/c/682628ae-5ca0-8013-920b-989394cb626c)