**AIML TERMINOLOGIES**

**Considered by taking Diabetes Prediction data set as sample data set.**

**FEATURE**

* In Artificial Intelligence and Machine Learning, a Feature refers to an individual measurable property or characteristic of a phenomenon being observed.
* Features are the inputs to machine learning models and are used to make predictions or classifications.
* A feature can also be called an attribute or a variable.
* **Features include glucose concentration, blood pressure, insulin level, BMI, age.**

**LABEL**

* In Artificial Intelligence and Machine Learning (AI/ML), the term "label" refers to the target variable or the output that the model aims to predict.
* The label is the value or outcome that the model is trying to predict.
* It is also known as the dependent variable in statistical terms.
* **According to above data set we can consider Diabetes Outcome as label.**

**PREDICTIONS**

* In Artificial Intelligence and Machine Learning (AI/ML), the term "prediction" refers to the output or result generated by a machine learning model after it has been trained.
* This output is an estimate or forecast of the target variable (label) based on the input features.
* **The prediction target is typically whether or not a patient has diabetes.**

**OUTLIER**

* In Artificial Intelligence and Machine Learning (AI/ML), an "outlier" refers to an observation or data point that significantly deviates from the other observations in the dataset.
* Outliers can affect the performance of machine learning models and the insights drawn from data analysis.
* An outlier is a data point that is markedly different from the rest of the data.
* It lies far away from the mean or median of the dataset, indicating that it does not follow the expected pattern.
* **Extremely high BMI (e.g., over 50) or very low BMI (e.g., under 15) could be an outlier.**

**TEST DATA**

* In Artificial Intelligence and Machine Learning (AI/ML), "test data" refers to a subset of the dataset that is used to evaluate the performance of a trained machine learning model.
* It is distinct from the training data.
* Test data is the dataset used to assess the model after it has been trained.
* It consists of data that the model has not encountered during the training process, ensuring that the evaluation is based on new, unseen data.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| gender | age | hypertension | heart\_disease | smoking\_history | bmi | HbA1c\_level | blood\_glucose\_level | diabetes |
| Female | 26 | 0 | 0 | No Info | 27.32 | 5 | 158 | 0 |
| Male | 39 | 0 | 0 | No Info | 27.32 | 6.1 | 100 | 0 |
| Male | 22 | 0 | 0 | current | 29.65 | 6 | 80 | 0 |
| Female | 26 | 0 | 0 | never | 34.34 | 6.5 | 160 | 0 |
| Female | 40 | 0 | 0 | never | 40.69 | 3.5 | 155 | 0 |
| Female | 36 | 0 | 0 | No Info | 24.6 | 4.8 | 145 | 0 |
| Female | 80 | 0 | 0 | No Info | 27.32 | 6.2 | 90 | 0 |
| Female | 2 | 0 | 0 | No Info | 17.37 | 6.5 | 100 | 0 |
| Male | 66 | 0 | 0 | former | 27.83 | 5.7 | 155 | 0 |
| Female | 24 | 0 | 0 | never | 35.42 | 4 | 100 | 0 |
| Female | 57 | 0 | 0 | current | 22.43 | 6.6 | 90 | 0 |

**TRAINING DATA**

* In Artificial Intelligence and Machine Learning (AI/ML), "training data" refers to the dataset used to train a machine learning model.
* This data includes both input features and corresponding labels (in supervised learning) or only input features (in unsupervised learning).
* The model learns patterns, relationships, and representations from this data, which it uses to make predictions or decisions on new, unseen data.

**MODEL**

* In Artificial Intelligence and Machine Learning (AI/ML), a "model" refers to a mathematical representation or algorithm that is trained on data to perform a specific task, such as classification, regression, clustering, or anomaly detection.
* The model learns patterns and relationships from the data, enabling it to make predictions or decisions based on new, unseen data.
* It can be thought of as an algorithm that processes input data to generate predictions or classifications.

**VALIDATION DATA**

* In Artificial Intelligence and Machine Learning (AI/ML), "validation data" refers to a subset of the dataset that is used to tune the model's hyperparameters and to evaluate its performance during the training process.
* It helps in assessing how well the model generalizes to unseen data and in preventing overfitting.
* It serves as a check on the model’s performance at various stages of training, helping to prevent overfitting.

**HYPERPARAMETER**

* In Artificial Intelligence and Machine Learning (AI/ML), a "hyperparameter" refers to a parameter whose value is set before the learning process begins and is not learned from the data.
* Hyperparameters are used to control the training process and the structure of the machine learning model.
* Hyperparameters are set manually or through optimization methods before the training of the model starts.
* Hyperparameters are not learned during training but are specified by the practitioner.

**EPOCH**

* In Artificial Intelligence and Machine Learning (AI/ML), an "epoch" refers to a single pass through the entire training dataset by the learning algorithm during the training process.
* An epoch represents one complete cycle through the full training dataset.
* During an epoch, the model processes each training example and updates its parameters based on the gradients computed from the loss function.

**LOSS FUNCTION**

* In Artificial Intelligence and Machine Learning (AI/ML), a "loss function" is a mathematical function used to measure how well a model's predictions match the actual target values.
* It quantifies the error or difference between the predicted values and the true values, providing a way to evaluate and optimize the model during training.
* The loss function calculates the error between the predicted outputs and the actual target values.
* The goal is to minimize this loss function during training, meaning the model’s predictions are as close as possible to the actual values.

**LEARNING RATE**

* In Artificial Intelligence and Machine Learning (AI/ML), the **"learning rate"** is a hyperparameter that controls the size of the steps taken during the optimization process to minimize the loss function.
* It determines how much the model’s parameters are adjusted in response to the calculated gradients during training.
* The learning rate is a scalar value that scales the magnitude of the updates made to the model’s parameters.
* It controls how quickly or slowly the optimization algorithm converges towards the minimum of the loss function.

**OVERFITTING**

* In Artificial Intelligence and Machine Learning (AI/ML), **"overfitting"** refers to a situation where a model learns the training data too well, capturing not only the underlying patterns but also the noise and anomalies.
* As a result, the model performs well on the training data but poorly on new, unseen data.
* Overfitting occurs when a model learns the details and noise in the training data to an extent that it negatively impacts its performance on new data.
* The model becomes too specialized to the training data and fails to generalize well to unseen data.

**UNDERFITTING**

* In Artificial Intelligence and Machine Learning (AI/ML), **"underfitting"** refers to a situation where a model is too simple to capture the underlying patterns in the training data.
* As a result, the model performs poorly on both the training data and new, unseen data.
* Underfitting occurs when the model cannot learn the relationships in the data effectively.

**REGULARIZATION**

* In Artificial Intelligence and Machine Learning (AI/ML), **"regularization"** refers to techniques used to prevent overfitting by adding constraints or penalties to the model’s complexity.
* Regularization helps ensure that the model generalizes well to new, unseen data rather than just fitting the training data too closely.
* Regularization methods are employed to limit the complexity of the model and prevent it from fitting the noise or random fluctuations in the training data.
* By introducing additional terms in the loss function, regularization discourages the model from becoming too complex.

**CROSS VALIDATION**

* In Artificial Intelligence and Machine Learning (AI/ML), **"cross-validation"** is a technique used to assess the performance and generalization ability of a model by partitioning the dataset into multiple subsets.
* It helps to ensure that the model performs well across different subsets of data and not just on a specific split.
* Cross-validation is a statistical technique used to evaluate and compare the performance of a model by dividing the dataset into multiple folds or subsets.
* It provides an estimate of how well the model will perform on unseen data by using different subsets for training and validation.

**FEATURE ENGINEERING**

* In Artificial Intelligence and Machine Learning (AI/ML), **"feature engineering"** refers to the process of using domain knowledge to create, transform, or select features (variables) that improve the performance of a machine learning model.
* It involves designing and creating new features from raw data to make the model more effective at learning and making predictions.
* Involves generating new features from existing data that can better represent the underlying patterns or relationships in the data.

**DIMENSIONALITY REDUCTION**

* In Artificial Intelligence and Machine Learning (AI/ML), **"dimensionality reduction"** refers to techniques used to reduce the number of features or dimensions in a dataset while preserving as much of the important information as possible.
* It helps in simplifying models, improving computational efficiency, and mitigating issues related to high-dimensional data, such as the curse of dimensionality.

**BIAS**

* Bias refers to the systematic error that a model introduces due to assumptions made during its design or training.
* It represents the model's tendency to consistently deviate from the true values or patterns in the data.
* Bias often arises because the model makes simplifying assumptions about the data, which may lead to inaccuracies in its predictions.

**VARIANCE**

* In Artificial Intelligence and Machine Learning (AI/ML), **"variance"** refers to the amount by which a model's predictions change when trained on different subsets of the training data.
* It is a measure of the model's sensitivity to fluctuations or noise in the training data.
* High variance indicates that the model learns the details and noise of the training data too well, which can lead to poor generalization to new, unseen data.

**BY**

**DEEPIKA S**

**22CSR040**