



SNS COLLEGE OF ENGINEERING



Kurumbapalayam(Po), Coimbatore – 641 107

Accredited by NAAC-UGC with 'A' Grade

Approved by AICTE, Recognized by UGC & Affiliated to Anna University, Chennai

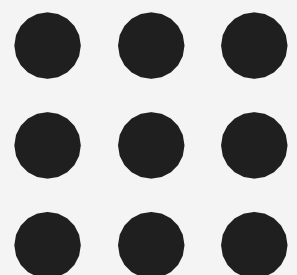
Department of Information Technology

Course Name – 23ADT202 Fundamental of Data
science and Analytics

II Year / IV Semester

Unit 1 – Introduction to Data science

Build in models





Introduction



Introduction

- Building models is a core activity in data science.
- It involves creating algorithms for predictions, classifications, and insights.
- Models are trained on historical data and evaluated for their performance.
- The goal is to meet desired objectives based on real-world data.



Defining the Objective

- The first step is to clearly define the model's goal.
- Objectives may include tasks like predicting sales, classifying emails, or forecasting trends.
- Example: Predicting sales based on marketing spend and seasonal factors.
- A well-defined objective guides the entire model-building process.



Splitting the Data

- Data must be divided into training and testing sets.
- The training set is used to build the model, while the test set evaluates its accuracy.
- Typically, 80% of data is used for training and 20% for testing.
- Proper splitting helps prevent overfitting and underfitting.



Selecting the Model

- Model selection depends on the problem type.
- For continuous outcomes, use regression models (e.g., Linear Regression).
- For categorical outcomes, classification models like Logistic Regression or Decision Trees work well.
- Clustering techniques like K-means are used for grouping data points.



Training the Model:

Once the model is chosen, fit it to the training data. The model learns patterns and relationships from the data.

For example, using `model.fit(x_train,y_train)` in regression tasks.

Training is a key step in allowing the model to make accurate predictions.



Evaluating the Model

- Model evaluation is done using various metrics.
- For regression, metrics include Mean Squared Error (MSE) and R^2 score.
- For classification, key metrics include Accuracy, Precision, Recall, and F1-score.
- Example: A regression model may use MSE to assess prediction accuracy.



Refining the Model

- After evaluation, the model may require optimization.
- Hyperparameter tuning can improve model performance.
- Methods like Grid Search and Random Search are commonly used for this purpose.
- Refining the model ensures it performs well on new, unseen data.



Challenges in Building Models

- **Overfitting:** The model is too complex and performs well on training data but poorly on test data.
- **Underfitting:** The model is too simple and fails to capture important patterns.
- **Data Imbalance:** Unequal class distribution in data can skew predictions.
- Solutions include regularization, more complex models, and oversampling techniques.



Key Takeaways

- Building models is an iterative process requiring constant fine-tuning.
- Achieving a balance between complexity and generalizability is essential for success.
- The chosen model must be suited to the data and problem at hand.
- Model evaluation and refinement are crucial to ensuring its effectiveness.



THANK YOU