



<b>Chapter 1 : Introduction to Machine Learning</b>		<b>1-1 to 1-71</b>
1.1	Introduction to Machine Learning .....	1-1
1.1.1	How does Machine Learning Work ? .....	1-3
1.1.2	Key Terms Associated with Machine Learning.....	1-5
1.1.3	Data Formats (Data and Types) .....	1-6
1.1.4	DIKW Pyramid.....	1-7
1.1.5	Categories of Data Analytics (Learning Tasks).....	1-8
1.2	Types of Machine Learning (Learning Paradigms) .....	1-9
1.2.1	Supervised Learning.....	1-10
1.2.2	Unsupervised Learning .....	1-10
1.2.3	Reinforcement Learning .....	1-11
1.2.4	How to Choose the Right Machine Learning Algorithm ? .....	1-13
1.3	Issues in Machine Learning.....	1-14
1.4	Application of Machine Learning.....	1-16
1.5	Steps in Developing a Machine Learning Application.....	1-23
1.6	Tools and Technology for Machine Learning .....	1-28
1.7	Data Quality and Remediation (Scales of Measurement Data) .....	1-29
1.7.1	Common Data Quality Issues .....	1-29
1.7.2	Remediating (Fixing) Data Quality Issues .....	1-31
1.8	Data and Dimensionality .....	1-33
1.8.1	Types of Dimensionality Reduction Techniques.....	1-35
1.9	Feature and Feature Engineering.....	1-35
1.9.1	Data .....	1-35
1.9.2	Tasks.....	1-36
1.9.3	Models.....	1-37
1.9.4	Features.....	1-37
1.9.5	Feature Engineering.....	1-38
1.9.6	Data Engineering -vs- Feature Engineering .....	1-40
1.10	Feature Transformation.....	1-41



1.10.1	Feature Construction.....	1-41
1.10.2	Quantization or Binning.....	1-41
1.10.3	Log Transform .....	1-43
1.10.4	Feature Scaling or Normalisation.....	1-44
1.10.5	Min-Max Scaling.....	1-44
1.10.6	Standardisation (Variance Scaling).....	1-45
1.10.7	Encoding Categorical Variables.....	1-47
1.10.8	One-Hot Encoding.....	1-47
1.10.9	Dummy Coding.....	1-48
1.10.10	Feature Hashing.....	1-49
1.10.11	Handling Textual Features.....	1-50
1.10.12	Feature Extraction .....	1-51
1.11	Principal Component Analysis (PCA).....	1-52
1.12	Independent Component Analysis (ICA).....	1-54
1.12.1	How ICA Works? .....	1-55
1.12.2	Comparison between PCA and ICA.....	1-57
1.13	Singular Value Decomposition (SVD).....	1-57
1.13.1	How SVD Works?.....	1-58
1.14	Linear Discriminant Analysis (LDA) .....	1-59
1.15	Feature (Subset) Selection.....	1-61
1.16	Key Drivers of Feature Selection .....	1-61
1.16.1	Measures of Feature Relevance.....	1-62
1.16.2	Measures of Feature Redundancy.....	1-63
1.17	Overall Feature Selection Process.....	1-65
1.18	Feature Selection Approaches .....	1-66

**Chapter 2 : Regression****2-1 to 2-54**

2.1	Concept Building - Statistical Tools in Machine Learning.....	2-1
2.1.1	Sampling Distributions .....	2-2
2.1.2	General Statistics .....	2-3
2.1.2(A)	Standard Deviation.....	2-3



2.1.2(B)	Variance .....	2-4
2.1.2(C)	Covariance .....	2-5
2.2	Concepts of Probability .....	2-6
2.2.1	Fundamental Rules of Probability.....	2-7
2.3	Random Variables.....	2-8
2.3.1	Discrete Random Variables .....	2-8
2.3.2	Continuous Random Variables .....	2-9
2.3.3	Multiple Random Variables.....	2-9
2.4	Common Discrete Distributions.....	2-10
2.4.1	The Binomial Distribution.....	2-10
2.4.2	The Bernoulli Distribution.....	2-12
2.4.3	The Multinomial Distribution.....	2-13
2.4.4	Multinoulli Distribution.....	2-15
2.4.5	The Poisson Distribution .....	2-15
2.4.6	The Empirical Distribution .....	2-17
2.5	Common Continuous Distributions .....	2-18
2.5.1	Uniform Distribution .....	2-18
2.5.2	Gaussian (Normal) Distribution.....	2-19
2.5.3	The Laplace Distribution .....	2-21
2.6	Univariate Regression.....	2-22
2.7	Cost Functions.....	2-22
2.7.1	Mean Error (ME).....	2-23
2.7.2	Mean Squared Error (MSE) .....	2-23
2.7.3	Mean Absolute Error (MAE) .....	2-24
2.7.4	Root Mean Squared Error (RMSE).....	2-24
2.7.5	R-Squared.....	2-25
2.8	Optimising Simple Linear Regression with Gradient Descent Algorithm.....	2-26
2.9	Derivative-based Optimisation.....	2-27
2.9.1	Characteristics of Derivative-based Optimisation Techniques.....	2-28
2.9.2	Steepest Descent.....	2-29



2.9.3	Newton's Method.....	2-38
2.10	Multivariate Regression - Model Representation.....	2-42
2.11	Polynomial Regression.....	2-43
2.12	Performance Enhancement with Regularisation Techniques.....	2-45
2.12.1	Comparison between L1 and L2 Regularisation Techniques.....	2-45
2.13	Model Representation and Interpretability (Generalisation Issues).....	2-46
2.13.1	Overfitting.....	2-46
2.13.2	Underfitting.....	2-47
2.13.3	Bias vs Variance.....	2-47
2.13.4	Bias-Variance Trade-Off.....	2-49
2.13.5	Characteristics (Detection) of a High Bias Model.....	2-51
2.13.6	Characteristics (Detection) of a High Variance Model.....	2-51

**Chapter 3 : Classification****3-1 to 3-30**

3.1	Regression Analysis.....	3-2
3.1.1	Linear Regression.....	3-3
3.1.1(A)	Use Cases (or Applications of) for Linear Regression.....	3-8
3.1.2	Logistic Regression.....	3-9
3.1.2(A)	Use Cases (or Applications of) for Logistic Regression.....	3-11
3.2	Reasons to Choose and Cautions.....	3-11
3.3	Additional Regression Models.....	3-12
3.4	Performance Evaluation.....	3-12
3.5	Support Vector Machines (SVM).....	3-12
3.5.1	Maximum Margin Linear Separators.....	3-13
3.5.2	Quadratic Programming Solution to Finding Maximum Margin Separators.....	3-14
3.5.3	Kernels for Learning Non-Linear Functions (Kernel Trick).....	3-17
3.5.4	Comparison between Logistic Regression and SVM.....	3-18
3.6	Multi-class Classification Techniques.....	3-18
3.6.1	One vs One (OvO).....	3-19
3.6.2	One vs Rest (OvR) (Ove vs All).....	3-20
3.6.3	Comparison between OvO and OvR.....	3-20
3.7	Enhancing Performance of Classification.....	3-20



3.7.1	Cross-Validation.....	3-21
3.7.1(A)	Holdout Method .....	3-21
3.7.1(B)	k-Fold Cross-Validation.....	3-21
3.7.1(C)	Leave-P-Out Cross-Validation (LpOCV).....	3-22
3.7.2	Sub-Sampling.....	3-23
3.7.3	Hyperparameter Tuning Techniques .....	3-24
3.7.3(A)	What Do Hyperparameters Do? .....	3-24
3.7.3(B)	How is Hyperparameter Tuning Carried out?.....	3-24
3.7.3(C)	Hyperparameter Tuning Algorithms .....	3-25

**Chapter 4 : Tree Based and Probabilistic Models****4-1 to 4-72**

4.1	Importance of Bayesian Methods.....	4-1
4.2	Bayes' Algorithm (Theorem) .....	4-2
4.3	Bayes' Theorem and Concept Learning.....	4-4
4.4	Bayesian Belief Network.....	4-5
4.5	Probability Density Estimation.....	4-7
4.6	Maximum A Posteriori (MAP) Estimation.....	4-9
4.7	Comparison between MLE and MAP.....	4-10
4.8	Classification Models .....	4-10
4.8.1	Decision Trees.....	4-11
4.8.2	Key Terms and Concepts .....	4-12
4.8.2(A)	Entropy.....	4-12
4.8.2(B)	Information Gain.....	4-15
4.8.2(C)	Gain Ratio.....	4-20
4.8.2(D)	Gini Index.....	4-24
4.9	Decision Tree Algorithms.....	4-38
4.9.1	The General Algorithm.....	4-38
4.9.2	ID3 Algorithm.....	4-38
4.9.3	C4.5 Algorithm.....	4-43
4.9.4	CART Algorithm .....	4-43
4.9.5	Evaluating a Decision Tree .....	4-44



4.10	Naïve Bayes (Classification by Bayesian Belief Networks) .....	4-44
4.10.1	Naïve Bayes Classifier .....	4-45
4.10.2	Smoothing .....	4-53
4.10.3	Advantages of Naïve Bayes Classifier .....	4-55
4.10.4	Disadvantages of Naïve Bayes Classifier.....	4-55
4.11	Diagnostics (Evaluation Measures) of Classifiers .....	4-56
4.11.1	ROC Curve .....	4-59
4.11.2	Area Under the Curve (AUC) .....	4-60
4.12	Additional Classification Methods.....	4-60
4.12.1	Bagging.....	4-60
4.12.2	Boosting .....	4-63
4.12.3	Random Forests .....	4-64

---

**Chapter 5 : Distance and Rule Based Models**
**5-1 to 5-66**

5.1	Distance Based Models (Distance Metrics) .....	5-1
5.1.1	Euclidean Distance.....	5-1
5.1.2	Manhattan Distance .....	5-2
5.1.3	Hamming Distance.....	5-3
5.1.4	Minkowski Distance .....	5-4
5.2	Clustering .....	5-5
5.2.1	Properties of a Cluster .....	5-7
5.2.2	Types of Clustering.....	5-7
5.2.3	Use Cases (Applications) of Clustering.....	5-7
5.2.4	K-means .....	5-10
5.2.5	Determining the Number of Clusters.....	5-19
5.2.6	Diagnostics (Performance Measures).....	5-23
5.2.7	Reasons to Choose and Cautions (Drawbacks / Challenges).....	5-24
5.3	k-Nearest Neighbours (kNN) Classification Algorithm.....	5-25
5.4	Hierarchical Clustering.....	5-27
5.4.1	Dendrogram .....	5-28
5.4.2	Hierarchical Clustering Strategies (Algorithms) .....	5-29



5.4.2(A)	Agglomerative Hierarchical Clustering.....	5-29
5.4.2(B)	Divisive Hierarchical Clustering.....	5-31
5.4.3	Agglomeration (Linkage) Methods.....	5-33
5.5	Association Rules.....	5-33
5.5.1	Key Terms and Properties of Association Rules .....	5-34
5.5.1(A)	Itemset.....	5-35
5.5.1(B)	Support.....	5-36
5.5.1(C)	Confidence .....	5-37
5.5.1(D)	Lift.....	5-38
5.5.1(E)	Leverage.....	5-41
5.5.1(F)	Subsets.....	5-42
5.5.2	Rule Learning for Subgroup Discovery.....	5-44
5.5.3	Apriori Algorithm.....	5-44
5.5.3(A)	How Apriori Algorithm Works.....	5-45
5.5.3(B)	Case Study – Transactions in Grocery Store .....	5-50
5.5.3(C)	Validation and Testing.....	5-55
5.5.3(D)	Diagnostics.....	5-55

---

**Chapter 6 : Introduction to Artificial Neural Network**
**6-1 to 6-32**

6.1	Introduction to Neural Networks .....	6-1
6.1.1	Fundamental Concept .....	6-1
6.1.2	Biological Neuron.....	6-3
6.1.3	Evolution of Neural Networks (Hebb’s Rule) .....	6-4
6.2	Artificial Neural Networks (ANN) .....	6-5
6.2.1	The Perceptron.....	6-10
6.2.2	The Bias Input.....	6-11
6.2.3	Multilayer Perceptron.....	6-12
6.2.4	Shallow and Deep Neural Network (Concept of Deep Learning) .....	6-13
6.3	Neural Network (NN) Architecture .....	6-14
6.4	Activation Functions .....	6-17
6.4.1	Types of Activation Functions.....	6-17



---

6.4.2	Loss Function - Mean Square Error .....	6-21
6.5	Learning Process (Learning Algorithm).....	6-22
6.6	Backpropagation.....	6-22
6.6.1	How Backpropagation Works? .....	6-25
6.6.2	Forward Propagation .....	6-27
6.6.3	Error Derivative.....	6-28
6.6.4	Additional Derivatives .....	6-29
6.6.5	Backpropagation.....	6-30

---