

Branch Name:	Bachelor of Computer Application (BCA)
Program Code:	CS101
Course Title:	Feature Engineering
Course Code:	23-DSE203-2C
Pre-requisite Course:	Basic knowledge of data analytics/machine learning and familiarity with any programming language.

Course Objective (CO):

This course is intended to make the students:

- To understand the role of feature engineering in machine learning and data analytics.
- To apply data preprocessing and transformation techniques.
- To learn feature selection and extraction methods.
- To develop optimized datasets for improving model performance.

Teaching and Examination Scheme:

Teaching Scheme (Hours per week)				Evaluation Scheme (Marks)								
Lecture (L)	Tutorial (T)	Practical (P)	Credit	Theory				Practical				Total
				University Assessment		Continuous Assessment		University Assessment		Continuous Assessment		
				Min. Marks	Mix. Marks	Min. Marks	Mix. Marks	Min. Marks	Max. Marks	Min. Marks	Max. Marks	
2	-	2	3	09	25	09	25	-	-	09	25	75

Course Contents:

Sr. No	Topic	Total Hours	Weightage (%)
1.	Introduction to Data and Features: Importance of Features in Machine Learning, Role of Features in Prediction Accuracy and Model Performance, Types of Data and Features: Numerical Data Categorical Data, Ordinal Data, Discrete Data, Continuous Data, Interval Data, Ratio Data. Feature Characteristics, Representation of Real-World Data in Machine Learning Systems.	12	25
2.	Basic Feature Preprocessing: Introduction to Feature Preprocessing Handling Missing Data: Deletion Methods, Imputation Techniques, Data Cleaning Techniques, Noise and Inconsistency Handling, Feature Scaling: Normalization Standardization Data Transformation Techniques Importance of Data Preprocessing in Machine Learning.	12	25
3.	Feature Engineering and Feature Selection Techniques: Feature Engineering for Numerical Data: Binning and Discretization, Polynomial Features, Interaction Features. Feature Engineering for Categorical Data: One-Hot Encoding, Label Encoding. Feature Extraction vs Feature Selection. Steps in Feature Selection. Feature Selection Methods: Filter Methods, Wrapper Methods, Hybrid Methods.	12	25
4.	Feature Reduction and Dimensionality Reduction: Introduction to Feature Reduction, Need for Dimensionality Reduction, Concept of Dimensionality Reduction, Principal Component Analysis (PCA): Basic Concept, Working Principle, Applications of PCA. Advantages of Feature Reduction, Improving Model Performance using Reduced Features.	12	25

Course Learning Outcomes (CLO):

At the end of this course, the student would be able

- CLO1:** Understand the importance of features in machine learning and differentiate between various types of data and features (structured vs. unstructured, categorical, numerical, text, and date-time).
- CLO2:** Apply basic feature preprocessing techniques such as handling missing data, data cleaning, and feature scaling and normalization.
- CLO3:** Implement feature engineering techniques for numerical data, including binning, discretization, polynomial and interaction features, and log transformation.
- CLO4:** Utilize categorical data techniques, such as one-hot encoding and label encoding, and understand feature selection methods, including filter and wrapper methods.
- CLO5:** Perform feature transformation using techniques like Principal Component Analysis (PCA) and understand its application in machine learning.

List of Text book:

- M.C. Trivedi, Data Science and Data Analytics Using Python Programming, Khanna Publishing House, 2024.
- Zheng, Alice, & Casari, Amanda. (2018). Feature engineering for machine learning: Principles and techniques for data scientists. O'Reilly Media, Inc.
- Kalita, J. K., Bhattacharyya, D. K., & Roy, S. (2023). Fundamentals of Data Science: Theory and Practice. Elsevier. ISBN-13: 9780323917780.

List of Reference book:

- Duda, R. O., Hart, P. E., Stork, D (2007). Pattern classification (2Ed), John Wiley & Sons, ISBN-13: 978-8126511167.
- N. Bhaskar, Vasundhara, Machine Learning, Khanna Publishing House, 2024.
- M.C. Trivedi, Deep Learning and Neural Network_MC Trivedi, Khanna Publishing House, 2024.
- Ng, Andrew. (2018). Machine learning yearning (Draft, MIT Licensed). GitHub. ISBN- 10: 199957950X, ISBN-13: 978-1999579500.
- Han, Jiawei, Kamber, Micheline, & Pei, Jian. (2011). Data mining: Concepts and techniques (3rd ed.). Morgan Kaufmann Publishers. ISBN 978-0123814791.

E-Resource / Web Links:

- <https://www.coursera.org>
- <https://www.edx.org>
- <https://nptel.ac.in>

List of Experiments:

Note: The experiment list provided beneath is for reference only. The course teacher may change/formulate it as per his/her methodology and requirement.

Practical List:

No.	Program Description
1.	Handle missing values in column(s) of a dataset. For example, fill missing values with the mean/median/mode of the columns such as 'Age', 'Height', 'Weight', 'Grade' for a dataset.
2.	Clean a dataset by identifying and removing invalid data entries. For example, a dataset having columns 'Name', 'Gender' and 'Age' where 'Name' contains 'invalid data'.
3.	Scale numerical features using Min-Max normalization for a dataset with columns like 'Height', 'Weight'.
4.	Perform exploratory data analysis and visualize data distributions using histograms and boxplots.
5.	Compute and visualize the correlation matrix of a dataset with 2 or more columns.
6.	Bin numerical data into discrete intervals for a dataset with a column containing numerical values.
7.	Create polynomial and interaction features from numerical data in a dataset with two columns.
8.	Apply logarithmic transformation to skewed numerical features in a dataset with column 'Distance'.
9.	Perform one-hot encoding on categorical features in a dataset with column 'Category' containing categorical values. The distinct values in the Category feature are [Good, Better, Best] and Gender [Male, Female].
10.	Preprocess text data (tokenization) for a dataset with a column 'Text'.
11.	Preprocess text data (stemming) for a dataset with a column 'Text'.
12.	Preprocess text data (lemmatization) for a dataset with a column 'Text'.
13.	Convert text data into a Bag-of-Words representation for a dataset with a column 'Text'.

14.	Apply TF-IDF transformation to text data for a column 'Text'.
15.	Perform image augmentation (resizing, normalization, rotation, translation) for a set of images.
16.	Perform image augmentation resizing for a set of images.
17.	Perform image augmentation normalization for a set of images.
18.	Perform image augmentation rotation for a set of images.
19.	Perform image augmentation translation for a set of images.
20.	Decompose a time series into trend, seasonal, and residual components for a dataset with a column 'Time Series'.
21.	Perform Principal Component Analysis (PCA) on a dataset and visualize the first two principal components.

Mapping of CLO with PO and PSO

Course Learning Outcomes	Program Outcomes (POs)												Program Specific Outcomes (PSOs)		
	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CLO1	L	H		H	L	L				L	L	M	M	L	M
CLO2	L	H		H	H	M	M	L		M	M	M	H	M	H
CLO3		H		H	H	H	M	M	L	M	H	M	H	M	H
CLO4	L	M	L	H	H	H	M	M	L	H	M	M	H	M	H
CLO5		H		H	M	H	H	H	H	M	M	H	H	3	H

H: High, M: Medium, L: Low