गोंय विद्यापीठ
ताळगांव पठार,
गोंय – ४०३ २०६
फोन : +९१-८६६९६०९०४८

International Year of Cooperatives
Cooperatives Build a Better World

(Accredited by NAAC)

Goa University
Taleigao Plateau, Goa-403 206
Tel : +91-8669609048
Email : registrar@unigoa.ac.in
Website : www.unigoa.ac.in

GU/Acad –PG/BoS -NEP/2025/388 Date: 09.09.2025

CIRCULAR

The University has decided to implement the Curriculum and Credit Framework for the Undergraduate Programme (CCFUP) under the National Education Policy (NEP), 2020 based on All India Council for Technical Education (AICTE) and National Credit Framework (NCrF) Guidelines from the Academic Year 2024-2025 onwards.

The Syllabus of Semester III & IV Courses offered under Specialization **Data Science** of **Bachelor of Engineering in Information Technology/ Bachelor of Engineering in Information Technology & Engineering** Programme is attached.

The Dean, Faculty of Engineering and Principals of affiliated Engineering Colleges are requested to take note of the above and bring the contents of the Circular to the notice of all concerned.

ASHWIN VYAS LAWANDE
Digitally signed by ASHWIN VYAS LAWANDE
Date: 2025.09.09 11:48:28 +05'30'

(Ashwin V. Lawande)
Deputy Registrar – Academic

To,

1. The Dean, Faculty of Engineering, Goa University.
2. The Principals of affiliated Engineering Colleges.

Copy to,

1. The Director, Directorate of Technical Education, Govt. of Goa
2. The Chairperson, BoS in Information Technology Engineering.
3. The Controller of Examinations, Goa University.
4. The Assistant Registrar, Prof. Examinations (Technical and Allied), Goa University.
5. Directorate of Internal Quality Assurance, Goa University for uploading the Syllabus on the University website.

Information Technology/ Information Technology and Engineering
Specialization: Data Science

Sr. No.	Semester	Course Code	Title of the Course	L	T	P	TCr
1	III	ITH-281	Foundations of Data Science	3	0	0	3
		ITH-282	Foundations of Data Science Lab	0	0	1	1
2	IV	ITH-283	Data Transformation and Analytics (Practical)	3	0	0	3
		ITH-284	Data Transformation and Analytics Lab	0	0	1	1
3	V	ITH-381	Neural Networks	3	0	0	3
		ITH-382	Neural Networks Lab	0	0	1	1
4	VI	ITH-383	Deep Learning and Text Processing	3	0	0	3
		ITH-384	Deep Learning and Text Processing Lab	0	0	1	1
5	VII	ITH-481	Applications of Data Science	3	0	0	3
		ITH-482	Applications of Data Science Lab	0	0	1	1
			TOTAL	**15**	**0**	**5**	**20**

Approved by: Standing Committee of the Academic Council on 24th & 25th July 2025

SEMESTER III

Name of the Programme : **Information Technology/ Information Technology and Engineering (Data Science)**
Course Code : **ITH-281**
Title of the Course : **Foundations of Data Science**
Number of Credits : **3**
Effective From AY : **2024-25**

Pre-requisites of the course:	Nil	
Course Objectives	This course will enable the students to: • Describe the data science pipeline, tools, and key concepts. • Analyze data acquisition, preprocessing, and feature engineering techniques. • Interpret data visualization and statistical methods for decision making. • Apply machine learning algorithms for supervised and unsupervised learning tasks.	
Contents:		**No. of Hours**
Unit - 1	**Introduction to Data Science:** Definition, importance, and applications of data science, Data Science pipeline, Different Sectors using Data science. **Tools and technologies used** – Python, R, Jupyter Notebook/Google Colab, Power BI, Tableau. Data types, data sources, and data acquisition techniques.	**10**
Unit - 2	**Data Preprocessing and Cleaning:** Handling missing data, data transformation, normalization, and standardization. **Feature Generation and Feature Selection** (Extracting Meaning from Data) – Motivating application: user (customer) retention- Feature Generation (brainstorming, role of domain expertise, and place for imagination)- Feature Selection algorithms.	**11**
Unit - 3	**Data visualization basics** – Histograms, scatter plots, box plots using Python libraries (Matplotlib, Seaborn). **Foundations of Mathematics for Data Science:** Descriptive statistics, distributions. Probability theory and Bayes Theorem. Hypothesis testing and confidence intervals.	**12**
Unit - 4	**Basics of Machine Learning:** Introduction to supervised and unsupervised learning, Overview of classification, regression, clustering. Applications of Data Science, Data Science and Ethical Issues- Discussions on privacy, security, ethics- A look back at Data Science- Next-generation data scientists- Getting informed consent - The Five Cs – Diversity – Inclusion – Future Trends.	**12**

Pedagogy	Interactive, reflective, and inquiry-based methods, with a strong emphasis on critical thinking and problem-solving skills.
References	**Text Books:** 1. Cathy O'Neil and Rachel Schutt, Doing Data Science, 1st Edition, O'Reilly Media, 2013, ISBN-13: 978-1449367372. 2. Joel Grus, Data Science from Scratch, 2nd Edition, O'Reilly Media, 2019, ISBN-13: 978-1491901427. **Reference Books:** 1. Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, An Introduction to Statistical Learning: with Applications in R, 1st Edition, Springer, 2013, ISBN-13: 978-1461471370. 2. Bart Baesens, Analytics in a Big Data World: The Essential Guide to Data Science and its Applications, Wiley, 2014, ISBN-13: 978-1118876138. 3. Davy Cielen, Arno D. B. Meysman, Mohamed Ali, Introducing Data Science, 1st Edition, Manning Publications Co., 2016, ISBN-13: 978-1617291376.
Course Outcomes	Upon completion of the course, students will be able to: CO 1. Describe the data science pipeline, tools, and key concepts. CO 2. Analyze data acquisition, preprocessing, and feature engineering techniques. CO 3. Interpret data visualization and statistical methods for decision making. CO 4. Apply machine learning algorithms for supervised and unsupervised learning tasks.

Name of the Programme : **Information Technology/ Information Technology and Engineering (Data Science)**
Course Code : **ITH-282**
Title of the Course : **Foundations of Data Science Lab**
Number of Credits : **1**
Effective From AY : **2024-25**

Pre-requisites of the course:	Basics of programming	
Course Objectives:	This course will enable students to: • Use Python libraries like Pandas and NumPy for data manipulation. • Develop Python programs for plotting data with Matplotlib. • Select relevant features using feature selection algorithms. • Test hypotheses using statistical methods.	
Contents:	**Topics**	**No. of Hours**
	1. Develop python program for plots using Matplotlib 2. Implementing Visualization for complex dataset using data visualization Tools- Jupyter Notebook/Google Colab, Microsoft Power BI, Excel/Google Sheets, Tableau, R studio 3. Implementing Feature Selection algorithms. 4. Implementing Data Preprocessing and Cleaning on a complex dataset 5. Implementing of clustering. 6. Implementing the pandas and Numpy Libraries of python in data science. 7. Implement Web Scraping Using Python 8. Implementing Feature Generation. 9. Implementing Bayes Theorem. 10. Implementing Hypothesis testing.	**30**
Pedagogy	Integration of instructional learning, constructive thinking, inquiry-based, collaborative, experiential, and problem-solving approaches.	
References/ Readings:	**Text Books:** 1. Cathy O'Neil and Rachel Schutt, Doing Data Science, 1st Edition, O'Reilly Media, 2013, ISBN-13: 978-1449367372. 2. Joel Grus, Data Science from Scratch, 2nd Edition, O'Reilly Media, 2019, ISBN-13: 978-1491901427. **Reference Books:** 1. Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, An Introduction to Statistical Learning: with Applications in R, 1st Edition, Springer, 2013, ISBN-13: 978-1461471370.	

	2. Bart Baesens, Analytics in a Big Data World: The Essential Guide to Data Science and its Applications, Wiley, 2014, ISBN-13: 978-1118876138. 3. Davy Cielen, Arno D. B. Meysman, Mohamed Ali, Introducing Data Science, 1st Edition, Manning Publications Co., 2016, ISBN-13: 978-1617291376.
Course Outcomes:	Upon completion of the course, students will be able to: CO 1. Demonstrate preprocessing for complex datasets using Python libraries. CO 2. Analyze data patterns using hypothesis testing and Bayesian methods. CO 3. Apply feature selection techniques to improve data modelling. CO 4. Create visualizations of data using Python and other tools.

SEMESTER IV

Name of the Programme : Information Technology/ Information Technology and Engineering (Data Science)
Course Code : ITH-283
Title of the Course : Data Transformation and Analytics
Number of Credits : 3
Effective From AY : 2024-25

Pre-requisites of the course:	Basics of Data Science	
Course Objectives	**This course will enable the students to:** • Describe data collection, cleaning techniques, and handling missing/duplicate data. • Analyze data using descriptive statistics, visualization, and correlation methods. • Evaluate supervised and unsupervised learning models and preprocessing techniques. • Apply feature engineering, clustering, and dimensionality reduction on datasets.	
Contents:		**No. of Hours**
Unit - 1	**Data Collection and Cleanin**g: Data collection techniques Dealing with missing values Handling duplicates and outliers Data types and conversions. Automated data collection with APIs and web scraping tools. Case study: Cleaning raw data from surveys/web scraping	**10**
Unit - 2	**Data Transformation Techniques**: Normalization and standardization, Binning and discretization, One-hot encoding and label encoding, Data aggregation and summarization Transformation using Excel and Python (Pandas). Handling skewed data, Feature scaling methods Feature engineering: creating new features from existing ones.	**11**
Unit - 3	**Exploratory Data Analysis (EDA)**: Introduction to EDA Descriptive statistics: mean, median, mode, variance, standard deviation Frequency distributions and cross-tabulations Correlation analysis Visualization techniques: histograms, boxplots, scatterplots Missing value visualization (using seaborn/matplotlib), Pair plots and heatmaps for correlation visualization, Time series visualization.	**12**
Unit - 4	**Supervised Learning:** Classification and regression, generalization, overfitting and underfitting, supervised machine	**12**

	learning algorithms, uncertainty estimates from classifiers. (Common models: Linear Regression, Logistic Regression, Decision Trees, Random Forests, SVM, KNN). **Unsupervised Learning and Preprocessing**: Types of unsupervised learning, challenges in unsupervised learning, preprocessing and scaling, dimensionality reduction, feature extraction, and manifold learning, clustering. (K-Means, DBSCAN, and Hierarchical Clustering), Principal Component Analysis (PCA),	
Pedagogy	Interactive, reflective, and inquiry-based methods, with a strong emphasis on critical thinking and problem-solving skills.	
References	**Text Books:** 1. Wes McKinney, Python for Data Analysis 2nd edition O'Reilly Media, 2017, ISBN-13: 978-1491957660. 2. Andreas C. Muller and Sara Guido, Introduction to Machine Learning with Python 1st Edition, O'Reilly Media, 2016, ISBN-13: 978-1449369415. **Reference Books:** 1. Aurélien Géron, Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2nd Edition, O'Reilly Media, 2019, ISBN-13: 978-1492032649. 2. Jake VanderPlas, Python Data Science Handbook, 1st Edition, O'Reilly Media, 2016, ISBN-13: 978-1491912058. 3. U Dinesh Kumar, Business Analytics: The Science of Data - Driven Decision Making, 2nd Edition, John Wiley & Sons, 2021, ISBN-13: 978-9354246197.	
Course Outcomes	Upon completion of the course, students will be able to: CO 1. Describe methods for data collection, cleaning, and handling missing or duplicate data. CO 2. Analyze datasets using descriptive statistics, visualizations, and correlation techniques. CO 3. Evaluate different supervised and unsupervised learning algorithms and preprocessing methods. CO 4. Apply feature engineering, clustering, and dimensionality reduction to real-world datasets.	

Name of the Programme : **Information Technology/ Information Technology and Engineering (Data Science)**
Course Code : **ITH-284**
Title of the Course : **Data Transformation and Analytics Lab**
Number of Credits : **1**
Effective From AY : **2024-25**

Pre-requisites of the course:	Basics of programming	
Course Objectives:	This course will enable students to: • Develop skills in handling in collecting, cleaning, and preprocessing real-world datasets using tools like Python and Excel. • Apply data transformation techniques such as normalization, encoding, and aggregation. • Conduct exploratory data analysis (EDA) using statistical methods and data visualization. • Implement the machine learning models using supervised and unsupervised techniques.	
Contents:	**List of Programs**	**No. of Hours**
	1. Program to handle missing values using deletion, imputation (mean/median/mode), detect and remove duplicate entries and interpolation. 2. Perform normalization and standardization on numerical features. (Apply MinMaxScaler and StandardScaler and compare their impact.) 3. Perform data binning and discretization. (Convert a continuous variable into discrete bins (e.g., age groups). 4. Visualize data distributions using histograms, boxplots, and scatterplots. 5. Train a supervised model (e.g., Linear Regression or Decision Tree) with clean and transformed data. 6. Apply one-hot encoding and label encoding on categorical data. 7. Generate descriptive statistics for numerical columns. (Calculate mean, median, mode, variance, and standard deviation) 8. Apply an unsupervised learning algorithm (e.g., K-Means clustering). 9. Perform dimensionality reduction using PCA. 10. End-to-end data pipeline — from raw dataset to EDA and machine learning model.	**30**

Pedagogy:	Integration of instructional learning, constructive thinking, inquiry-based, collaborative, experiential, and problem-solving approaches.
References/ Readings:	**Text Books:** 1. Wes McKinney, Python for Data Analysis 2nd edition O'Reilly Media, 2017, ISBN-13: 978-1491957660. 2. Andreas C. Muller and Sara Guido, Introduction to Machine Learning with Python 1st Edition, O'Reilly Media, 2016, ISBN-13: 978-1449369415. **Reference Books:** 1. Aurélien Géron, Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2nd Edition, O'Reilly Media, 2019, ISBN-13: 978-1492032649. 2. Cathy O'Neil and Rachel Schutt, Doing Data Science, 1st Edition, O'Reilly Media, 2013, ISBN-13: 978-1449367372. 3. U Dinesh Kumar, Business Analytics: The Science of Data - Driven Decision Making, 2nd Edition, John Wiley & Sons, 2021, ISBN-13: 978-9354246197.
Course Outcomes:	Upon completion of the course, students will be able to: CO 1. Clean datasets by handling missing values and outliers CO 2. Transform data using tools like Pandas, Scikit-learn, and Excel CO 3. Visualize data trends and patterns using histograms, boxplots, scatterplots, and heatmaps CO 4. Develop and evaluate machine learning models for classification, regression, and clustering