

	MAHATMA GANDHI UNIVERSITY Kottayam, Kerala Undergraduate Programmes (HONOURS) 2026 Admission Onwards
---	---

SYLLABUS			
SIGNATURE COURSE			
Name of the College	Alphonsa College, Pala		
Faculty/ Discipline	Mathematics		
Programme	BSc (Hons) Mathematics		
Course Coordinator	Arya S		
Contributors	Sheena Mathew, Sonia K Thomas		
Course Name	Data Mining and Machine Language basics		
Type of Course	DSE		
Specialization title	Data Science		
Course Code	MG3DSEMATA06		
Course Level	200		
Course Summary	This course introduces the fundamental concepts, motivations, and challenges in data mining and knowledge discovery. The course also introduces adversarial and explainable machine learning, emphasizing model interpretability and security.		
Semester	3	Credits	4
Course Details	Learning Approach		Total Hours
	Credits	4	60
Pre-requisites, if any	Nil		

Course Outcomes (CO)

Number of COs		4	
CO No.	Expected Course Outcome	Learning Domains *	PO No
1	Understand the Fundamentals of Data Mining	U	PO1, PO2
2	Apply Techniques for Pattern Discovery	A	PO1, PO2, PO3
3	Implement Similarity Search Methods	A	PO1, PO2, PO10
4	Analyze Model Security and Explainability	An	PO8, PO10

*Remember (K), Understand (U), Apply (A), Analyse (An), Evaluate (E), Create (C), Skill (S), Interest (I) and Appreciation (Ap)

CO-PO Articulation Matrix

CO/PO	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO 10
CO 1	3	3	-	-	-	-	-	-	-	-

CO/PO	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO 10
CO 2	3	3	3	-	-	-	-	-	-	-
CO 3	3	3	-	-	-	-	-	-	-	3
CO 4	-	-	-	-	-	-	-	3	-	3

'0' is No Correlation, '1' is Slight Correlation (Low level), '2' is Moderate Correlation (Medium level) and '3' is Substantial Correlation (High level).

Course Content

Content for Classroom transaction (Units)

Module	Units	Course Description	Hrs	CO No.
1	Introduction to Data Mining			
	1.1	Introduction, Motivation, Plan, Evaluation, Policies	4	["1"]
	1.2	Introduction to Data Mining, What is Data mining, Motivation & challenges, Data Mining Tasks	5	["1"]
	1.3	Data: Types of Data, Data Quality, Data Preprocessing, Exploratory Data Analysis (EDA), Feature Engineering	7	["1"]
2	Mining Frequent Item sets			
	2.1	Application of Frequent Itemset: Association Rule Mining	4	["2"]
	2.2	Association Rule Mining, Apriori and FP Growth algorithm	6	["2"]
	2.3	Confidence, interest, lift measures, Rule Evaluation	6	["2"]
3	Similarity Search and Locality-Sensitive Hashing			
	3.1	Finding Similar Items: Application of Nearest Neighbor Search, Jaccard and Cosine Similarity, Shingles, Min-hashing	7	["3"]
	3.2	Locality Sensitive Hashing (LSH): Distance Measure, Theory of LSH, LSH Families for other Distance Measure, Application of LSH	8	["3"]
4	Adversarial and Explainable Machine Learning			
	4.1	Adversarial Machine Learning: Threat Modeling, Poisoning Attacks, Evasion Attacks	6	["4"]
	4.2	Explainable Machine Learning: Data, Models, Evaluation and Threat modeling	7	["4"]

Teaching and Learning Approach	Classroom Procedure (Mode of transaction) Lecture, Practical, Class with ICT Tools
---------------------------------------	--

Assessment Types	MODE OF ASSESSMENT Mode of Assessment: Theory
	A. Continuous Comprehensive Assessment (CCA) • Theory - 30 Marks Module Test, Seminar, Assignment, Practical, Quiz, Viva voce
	B. End Semester Evaluation (ESE) • Theory - 70 Marks Assessment Methods – End Semester Examination Duration of Examination – 2.00 Hrs Pattern of examination for Theory – Non-MCQ Different parts of written examination – Part - A , B , C Answer Type: ◦ PART - A ◦ One word - (5 out of 8) - $5 \times 2 = 10$ ◦ PART - B ◦ Short Essays/Problems - (5 out of 8) - $5 \times 6 = 30$ ◦ PART - C ◦ Essays - (3 out of 6) - $3 \times 10 = 30$

References

- Introduction to Data Mining. Pang-Ning Tan, Michael Steinbach, and Vipin Kumar, Pearson Education, 2016.

Suggested Readings

- Data Mining: Concepts and Techniques. Jiawei Han, Micheline Kamber, and Jian Pei, Morgan Kaufmann Publishers, 2022.
- Mining of Massive Datasets, Jure Leskovec, Anand Rajaraman, and Jeffrey David Ullman, Dreamtech Press, 2016.

Affidavit

- We, Alphonsa College, Pala and Arya S, retain the copyright of this syllabus and expressly prohibit its distribution in complete form to any institution outside our own.
- We, Alphonsa College, Pala, agree to appoint a new course coordinator for the proposed Data Science in the event of the unavailability of the currently nominated coordinator. This appointment will ensure the continued coordination of course delivery, assessments, and all related academic responsibilities necessary for the successful implementation of the specialization, for as long as the college offers this programme.
- We, Alphonsa College, Pala and Arya S, declare that no part of this signature course submitted here for approval has been taken from the course content developed by, or from any of the course titles prepared by, the BoS/expert committee in the same discipline under our University.

	MAHATMA GANDHI UNIVERSITY Kottayam, Kerala Undergraduate Programmes (HONOURS) 2026 Admission Onwards
---	---

SYLLABUS			
SIGNATURE COURSE			
Name of the College	Alphonsa College, Pala		
Faculty/ Discipline	Mathematics		
Programme	BSc (Hons) Mathematics		
Course Coordinator	Arya S		
Contributors	Sheena Mathew, Sonia K Thomas		
Course Name	Data Analysis and Visualization using Python		
Type of Course	DSE		
Specialization title	Data Science		
Course Code	MG4DSEMATA06		
Course Level	200		
Course Summary	This course is intended to equip students with a solid grounding in Python programming techniques for efficient data processing and analysis. It guides learners in using NumPy for high-speed numerical operations, developing proficiency in data manipulation and transformation with Pandas, and creating insightful visual representations of data using Matplotlib.		
Semester	4	Credits	4
Course Details	Learning Approach		Lecture
	Credits		Total Hours
	4		60
Pre-requisites, if any	Nil		

Course Outcomes (CO)

Number of COs		4	
CO No.	Expected Course Outcome	Learning Domains *	PO No
1	Understand basic Python programming concepts.	U	PO1, PO2
2	Use NumPy arrays for efficient numerical computation.	A	PO1, PO2, PO10
3	Perform Data Manipulation and Pre-processing Using Pandas	An	PO1, PO2, PO3
4	Implement various Python libraries to perform data analysis tasks.	I, S	PO8, PO10

*Remember (K), Understand (U), Apply (A), Analyse (An), Evaluate (E), Create (C), Skill (S), Interest (I) and Appreciation (Ap)

CO-PO Articulation Matrix

CO/PO	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO 10
CO 1	3	3	-	-	-	-	-	-	-	-

CO/PO	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO 10
CO 2	3	3	-	-	-	-	-	-	-	3
CO 3	3	3	3	-	-	-	-	-	-	-
CO 4	-	-	-	-	-	-	-	3	-	3

'0' is No Correlation, '1' is Slight Correlation (Low level), '2' is Moderate Correlation (Medium level) and '3' is Substantial Correlation (High level).

Course Content

Content for Classroom transaction (Units)

Module	Units	Course Description	Hrs	CO No.
1	Introduction to Data analysis			
	1.1	Understanding structured and unstructured data	4	["1"]
	1.2	Data Analysis process: Defining objectives, Data collection, Data cleaning, Data analysis, Data interpretation and visualization, Types of data analysis	8	["1"]
2	Numerical Computation with NumPy Library			
	2.1	Ndarray, Creating Ndarrays, Data types for Ndarrays, Arithmetic with NumPy Arrays	6	["2"]
	2.2	Initialization methods: zeros, ones, arange Accessing elements and slices, Multi-dimensional arrays, reshape(), flatten(), transpose()	5	["2"]
	2.3	Aggregation: sum(), mean(), std(), axis parameter, Concept of broadcasting	5	["2"]
3	Data Manipulation and Visualization			
	3.1	Data Manipulation with Pandas Library: Introduction to Pandas Object, Pandas data structures- Series, DataFrame, Index object.	6	["3"]
	3.2	Functionalities: Reindexing, Indexing, Selection, Filtering, Sorting	6	["3"]
	3.3	Need for data pre-processing, Data cleaning: removing errors and duplicates Handling missing data: simple methods	5	["3"]
	3.4	Data summarization: mean, median, mode, min, max: Basic data transformation Random Number Generation: np.random.rand(), randint(), Saving and Loading Arrays: .npy and .npz formats	5	["3"]
4	Introduction to Data visualization			
	4.1	Introduction to Data visualization: Matplotlib Library, pyplot	5	["4"]
	4.2	Data visualization using matplotlib: bar plot, line plot, histogram, pie chart, box plots, density plots and scatter plot.	5	["4"]

Teaching and Learning Approach	Classroom Procedure (Mode of transaction) Lecture, Lab Practice, Class Room with ICT Tools
---------------------------------------	--

Assessment Types	MODE OF ASSESSMENT Mode of Assessment: Theory
	A. Continuous Comprehensive Assessment (CCA) • Theory - 30 Marks Module Test, Seminar, Assignment, Practical, Quiz, Viva voce
	B. End Semester Evaluation (ESE) • Theory - 70 Marks Assessment Methods – End Semester Examination Duration of Examination – 2.00 Hrs Pattern of examination for Theory – Non-MCQ Different parts of written examination – Part - A , B , C Answer Type: ◦ PART - A ◦ One or two Sentences - (5 out of 8) – $5 \times 2 = 10$ ◦ PART - B ◦ Short Essays/Problems - (5 out of 8) – $5 \times 6 = 30$ ◦ PART - C ◦ Essays - (3 out of 6) – $3 \times 10 = 30$

References

- Wes Mckinney, "Python for Data Analysis: Data Wrangling with pandas, NumPy, and Jupyter" 3rd Edition, O'Reilly, 2022.
Free online access: <https://wesmckinney.com/book/>
- Richard Szeliski, Computer Vision: Algorithms and Applications, Springer-Verlag London Limited 2011.

Suggested Readings

- Fabio Nelli, "Python Data Analytics Data Analysis and Science Using Pandas, Matplotlib, and the Python Programming Language", Edition 1, 2015, Apress.
- William McKinney, "Python for Data Analysis: Data Wrangling with Pandas, NumPy, and Ipython", Edition 2, 2017, Shroff/O'Reilly.

Affidavit

- We, Alphonsa College, Pala and Arya S, retain the copyright of this syllabus and expressly prohibit its distribution in complete form to any institution outside our own.
- We, Alphonsa College, Pala, agree to appoint a new course coordinator for the proposed Data Science in the event of the unavailability of the currently nominated coordinator. This appointment will ensure the continued coordination of course delivery, assessments, and all related academic responsibilities necessary for the successful implementation of the specialization, for as long as the college offers this programme.
- We, Alphonsa College, Pala and Arya S, declare that no part of this signature course submitted here for approval has been taken from the course content developed by, or from any of the course titles prepared by, the BoS/expert committee in the same discipline under our University.

	MAHATMA GANDHI UNIVERSITY Kottayam, Kerala Undergraduate Programmes (HONOURS) 2026 Admission Onwards
---	---

SYLLABUS			
SIGNATURE COURSE			
Name of the College	Alphonsa College, Pala		
Faculty/ Discipline	Mathematics		
Programme	BSc (Hons) Mathematics		
Course Coordinator	Arya S		
Contributors	Sheena Mathew, Sonia K Thomas		
Course Name	Applied Statistical Methods using Python		
Type of Course	DSE		
Specialization title	Data Science		
Course Code	MG5DSEMATA06		
Course Level	300		
Course Summary	This course introduces core methods in data analysis, including clustering, classification, and regression, with emphasis on both unsupervised and supervised learning. It covers statistical inference, sampling distributions, estimation, hypothesis testing, and applications to proportions and variance analysis. The course also develops skills in experimental design, ANOVA, and linear regression for prediction and model evaluation.		
Semester	5	Credits	4
Course Details	Learning Approach		Total Hours
	Credits	4	60
Pre-requisites, if any	Basic probability concepts, random variables, probability distributions, and moments		

Course Outcomes (CO)

Number of COs		4	
CO No.	Expected Course Outcome	Learning Domains *	PO No
1	Understand clustering and classification methods	U	PO1, PO2
2	Apply statistical inference techniques	A	PO1, PO2, PO10
3	Analyze proportions, ANOVA, and simple regression	An	PO1, PO2, PO10
4	Evaluate multivariate and logistic regression models.	E	PO2, PO3

*Remember (K), Understand (U), Apply (A), Analyse (An), Evaluate (E), Create (C), Skill (S), Interest (I) and Appreciation (Ap)

CO-PO Articulation Matrix

CO/PO	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO 10
CO 1	3	3	-	-	-	-	-	-	-	-
CO 2	3	3	-	-	-	-	-	-	-	3
CO 3	3	3	-	-	-	-	-	-	-	3
CO 4	-	3	3	-	-	-	-	-	-	-

'0' is No Correlation, '1' is Slight Correlation (Low level), '2' is Moderate Correlation (Medium level) and '3' is Substantial Correlation (High level).

Course Content

Content for Classroom transaction (Units)

Module	Units	Course Description	Hrs	CO No.
1	Cluster Analysis			
	1.1	Cluster Analysis: Overview, K-Means, Agglomerative and Divisive clustering	6	["1"]
	1.2	DBSCAN, BIRCH Algorithm, CURE Algorithm, Cluster evaluation	4	["1"]
	1.3	Classification: Overview, Decision Tree, Model underfitting and overfitting,	4	["1"]
	1.4	Evaluating the performance of a Classifier, Rule-based Classifier, Nearest-neighbour Classifier, Bayesian Classifier,	5	["1"]
	1.5	Support Vector Machine, Ensemble Methods	5	["1"]
2	Sampling Distributions and Statistical Inference			
	2.1	Introduction, Review of sampling, selecting a sample, Sampling from a finite and infinite population, Point Estimation, Sampling distribution of sample means and properties of point estimators	3	["2"]
	2.2	Interval estimation, Population mean when σ is known, Population mean σ is unknown, Determining the sample size, Introduction to hypothesis testing, Types of errors, Inferences about the single population mean.	4	["2"]
	2.3	Inferences about the difference between two means, population variances, and Interval estimation	4	["2"]
3	Statistical Methods for Proportions, ANOVA, and Regression			
	3.1	Testing the equality of population proportions, Test of independence, Goodness of fit test, Introduction to experimental design, and Analysis of Variance (ANOVA) (Completely randomized design, Multiple comparison procedures, Randomized block design)	6	["3"]
	3.2	Simple linear regression model, Least squares method, Coefficient of determination, Model assumptions, Test for significance, Using the estimated regression equation for estimation and prediction	5	["3"]
	3.3	Residual analysis: Validating model assumptions, outliers, and influential observations.	4	["3"]
4	Multiple Regression and Logistic Regression			
	4.1	Multiple regression model, Least squares method, Multiple coefficients of determination, Model assumptions, Testing for significance, Multicollinearity.	3	["4"]
	4.2	The regression equation for estimation and prediction, residual analysis, Discuss case studies	4	["4"]
	4.3	The categorical independent variable and Logistic regression	3	["4"]

Teaching and Learning Approach	Classroom Procedure (Mode of transaction) Lecture, Demonstration through ICT tools and video, Practical
---------------------------------------	---

Assessment Types	MODE OF ASSESSMENT Mode of Assessment: Theory
	A. Continuous Comprehensive Assessment (CCA) • Theory - 30 Marks Module Test, Seminar, Assignment, Quiz, Viva voce
	B. End Semester Evaluation (ESE) • Theory - 70 Marks Assessment Methods – End Semester Examination Duration of Examination – 2.00 Hrs Pattern of examination for Theory – Non-MCQ Different parts of written examination – Part - A , B , C Answer Type: ◦ PART - A ◦ Short answer - (5 out of 8) – 5 × 2 = 10 ◦ PART - B ◦ Short Essays/Problems - (5 out of 8) – 5 × 6 = 30 ◦ PART - C ◦ Essays - (3 out of 6) – 3 × 10 = 30

References

- Introduction to Data Mining. Pang-Ning Tan, Michael Steinbach, and Vipin Kumar, Pearson Education, 2016
- David R Anderson, Dennis J Sweeney, Thomas A Williams, Jeffrey D. Camm and James J. Cochran, Statistics for Business and Economics, 12th Edition, Cengage Learning, 2014.

Suggested Readings

- Mining of Massive Datasets, Jure Leskovec, Anand Rajaraman, and Jeffrey David Ullman, Dreamtech Press, 2016.
- Data Mining: Concepts and Techniques. Jiawei Han, Micheline Kamber, and Jian Pei, Morgan Kaufmann Publishers, 2022.
- Sheldon M. Ross, Introductory Statistics, Academic Press, 2005

Affidavit

- We, Alphonsa College, Pala and Arya S, retain the copyright of this syllabus and expressly prohibit its distribution in complete form to any institution outside our own.
- We, Alphonsa College, Pala, agree to appoint a new course coordinator for the proposed Data Science in the event of the unavailability of the currently nominated coordinator. This appointment will ensure the continued coordination of course delivery, assessments, and all related academic responsibilities necessary for the successful implementation of the specialization, for as long as the college offers this programme.
- We, Alphonsa College, Pala and Arya S, declare that no part of this signature course submitted here for approval has been taken from the course content developed by, or from any of the course titles prepared by, the BoS/expert committee in the same discipline under our University.

	MAHATMA GANDHI UNIVERSITY Kottayam, Kerala Undergraduate Programmes (HONOURS) 2026 Admission Onwards
---	---

SYLLABUS			
SIGNATURE COURSE			
Name of the College	Alphonsa College, Pala		
Faculty/ Discipline	Mathematics		
Programme	BSc (Hons) Mathematics		
Course Coordinator	Arya S		
Contributors	Sheena Mathew, Sonia K Thomas		
Course Name	Data Visualization using Power BI and SQL		
Type of Course	DSE		
Specialization title	Data Science		
Course Code	MG6DSEMATA06		
Course Level	300		
Course Summary	This course introduces techniques for transforming raw data into meaningful visual insights using Power BI and SQL. It covers data extraction, modeling, and visualization for effective decision-making.		
Semester	6	Credits	4
Course Details	Learning Approach		Lecture
	Credits		4
			Total Hours
			60
Pre-requisites, if any	NIL		

Course Outcomes (CO)

Number of COs		4	
CO No.	Expected Course Outcome	Learning Domains *	PO No
1	Explain different categories of data analytics and recognize commonly used tools for data visualization	U	PO1, PO2, PO3
2	Create visualizations using Power BI.	C	PO1, PO2, PO3
3	Design SQL queries for data definition and data manipulation	A	PO2, PO3, PO9, PO10
4	Apply SQL aggregation and grouping functions to summarize datasets	A	PO1, PO2, PO9, PO10

*Remember (K), Understand (U), Apply (A), Analyse (An), Evaluate (E), Create (C), Skill (S), Interest (I) and Appreciation (Ap)

CO-PO Articulation Matrix

CO/PO	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO 10
CO 1	3	3	3	-	-	-	-	-	-	-

CO/PO	PO 1	PO 2	PO 3	PO 4	PO 5	PO 6	PO 7	PO 8	PO 9	PO 10
CO 2	3	3	3	-	-	-	-	-	-	-
CO 3	-	3	3	-	-	-	-	-	3	3
CO 4	3	3	-	-	-	-	-	-	3	3

'0' is No Correlation, '1' is Slight Correlation (Low level), '2' is Moderate Correlation (Medium level) and '3' is Substantial Correlation (High level).

Course Content

Content for Classroom transaction (Units)

Module	Units	Course Description	Hrs	CO No.
1	Introduction to Data Visualization			
	1.1	Data Visualization Principles: Importance of data visualization, Choosing appropriate chart types, Color theory and design principles	6	["1"]
	1.2	Data Visualization Tools - Google Data Studio, QlikView, Tableau, Power BI.	6	["1"]
	1.3	Features- Common and salient features of popular Data Visualization Tool	6	["1"]
2	Power BI - Unlock the power of business analytics			
	2.1	Introduction to Power BI, Installing Power BI Desktop, interface tour, basic terminology, reports, dashboards, datasets, workspaces	8	["2"]
	2.2	Data visualization using Power BI- Creating simple visualization, Map Visualization, Combination Charts	6	["2"]
	2.3	Dashboard in Power BI - creating, sharing, Tiles in Dashboard.	6	["2"]
3	Introduction to SQL			
	3.1	Introduction- Features, advantages	5	["3"]
	3.2	Datatypes: numeric-string-date /time-boolean, Variables, MYSQL database-table-view.	5	["3"]
4	MYSQL			
	4.1	MYSQL Commands- DDL-Create-Alter, Drop, DML-Insert, Select, Update, Delete DCL-Grant, Revoke	6	["4"]
	4.2	Command clauses- where, order by, group by, having, like, between, MYSQL functions-aggregate functions.	6	["4"]

Teaching and Learning Approach	Classroom Procedure (Mode of transaction) Use of ICT tools in conjunction with traditional classroom teaching methods, Interactive sessions, Class discussions, Lab exercises
---------------------------------------	---

Assessment Types	MODE OF ASSESSMENT Mode of Assessment: Theory
	A. Continuous Comprehensive Assessment (CCA) • Theory - 30 Marks Module Test, Seminar, Assignment, Practical, Quiz, Viva voce
	B. End Semester Evaluation (ESE) • Theory - 70 Marks Assessment Methods - End Semester Examination Duration of Examination - 2.00 Hrs Pattern of examination for Theory - Non-MCQ Different parts of written examination - Part - A , B , C Answer Type: ◦ PART - A ◦ One or two Sentences - (5 out of 8) - $5 \times 2 = 10$ ◦ PART - B ◦ Short Essays/Problems - (5 out of 8) - $5 \times 6 = 30$ ◦ PART - C ◦ Essays - (3 out of 6) - $3 \times 10 = 30$

References

- Vikram Vaswani, "MySQL: The Complete Reference", 1st Edition, McGraw Hill Education, 2004 (Module 3 & 4).
- Claus O. Wilke, Fundamentals of Data Visualization, O'Reilly, April 2019
- https://www.tutorialspoint.com/power_bi/power_bi_visualization_options.htm

Suggested Readings

- Ajith Kumar B P. Python for Education, Inter University Accelerator Centre - New Delhi ,2010

Affidavit

- We, Alphonsa College, Pala and Arya S, retain the copyright of this syllabus and expressly prohibit its distribution in complete form to any institution outside our own.
- We, Alphonsa College, Pala, agree to appoint a new course coordinator for the proposed Data Science in the event of the unavailability of the currently nominated coordinator. This appointment will ensure the continued coordination of course delivery, assessments, and all related academic responsibilities necessary for the successful implementation of the specialization, for as long as the college offers this programme.
- We, Alphonsa College, Pala and Arya S, declare that no part of this signature course submitted here for approval has been taken from the course content developed by, or from any of the course titles prepared by, the BoS/expert committee in the same discipline under our University.