JINDAL GLOBAL BUSINESS SCHOOL
INDIA'S FIRST MULTI-DISCIPLINARY GLOBAL BUSINESS SCHOOL

O.P. Jindal Global University
A Private University Promoting Public Service
NAAC Accreditation - 'A' Grade

Jindal Global Business School
Course Outline

Course Title	Data mining and Business Intelligence
Core or Elective	Elective
Program and Batch	BBA-2024, BBA-BA-2024, BBA-FB-2024, BBA-FM-2024
Semester & Academic Year	Fall 2026
Credits	3
Discipline/Area	IS & Analytics
Provide details if this course is a Prerequisite for any course/specialization	N/A
Name of the Faculty Member/Course Instructor	Arpit Singh
Contact Details of the Faculty Member	asingh6@jgu.edu.in +91-7355710937
Contact Details of Support Staff	jgbs-eo@jgu.edu.in
Faculty Member's Open Office Day/s & Time	TBD

Introduction to the Course

The course data mining and business intelligence is designed for BBA-3 and IBM-3 cohorts. This course provides an in-depth understanding of the techniques and tools used to analyse large datasets and extract meaningful insights for making informed business decisions. The course covers a range of topics such as data preparation, data mining techniques, clustering, classification, and association rule mining. The course has a **major pre-requisite** that students must have a detailed understanding of **probability** and **statistics**, including uni-variate and multi-variate statistical distributions, and basic **operations research**. The course is predominantly conceptual in nature with a judicious mix of both theoretical and practical sessions. However, a few examples related to some topics will be discussed in the EL sessions using python and therefore students are advised to have at least the basic understanding of python.

Course Learning Objectives

At the end of the course, students should be able to

1. CLO1- Understand the fundamental concepts including data preparation, data analysis, and data visualization
2. CLO2- learn how to use various tools and techniques to extract valuable insights from complex datasets and make data-driven decisions
3. CLO3- develop the skills to interpret and communicate data analysis results to stakeholders in a clear and concise manner
4. CLO4- learn how to apply data mining algorithms and predictive models to solve business problems and make accurate predictions

Programme Competency Goals

BBA Programme Competency Goals (PCGs)		**BBA Programme Learning Objectives (PLOs)**
		Students will be able to
1	**Responsible Global Citizenship:** Ability to understand the interplay between local and global issues and to act with sensitivity towards ethical and social issues	1. Understand local business issues
		2. Understand global business issues
		3. Demonstrate sensitivity towards ethical issues
		4. Demonstrate sensitivity towards social issues
2	**Effective communication:** Ability to effectively exchange ideas and information	5. Present their ideas with clarity
		6. Write in a coherent manner
		7. Use technology for communication
3	**Critical Thinking:** Ability to identify, analyze business problems and propose effective solutions	8. Identify main issues of business problems
		9. Examine information from different sources
		10. Draw inferences from analysis
4	**Teamwork:** Ability to work and contribute effectively in group -settings	11. Understand the factors to work effectively in groups
		12. Contribute effectively in groups

PLO-PCG Assessments Mapping Matrix

Program Learning Objectives (PLOs)	**Program Competency Goals (PCGs)**	**Course Assessment Item**
This course helps you to develop the following Program Learning Outcomes:	This course helps you to develop the following Program Competency Goals:	This learning outcome will be assessed in the following items
PLO5, PLO6, PLO7, PLO8, PLO9, PLO10	**PCG2, PCG3**	**A1, A2, A3, A4, A5**

Evaluation Schema

The course grade will be determined based on:

Assessment Task	**Weightage (Percentage)**	**Nature (Individual/Group)**	**Week of Assessment**	**PLOs to be Assessed**
A1: Class Participation	10%	Individual	Continuous	PLO-5, PLO-8, PLO-9
A2: Quiz 1 & A4: Quiz 2	15% + 15% = 30%	Individual	4th Week and 12th Week	PLO-5, PLO-6, PLO-9, PLO-10
A3 Short report	30%	Individual	8th Week	PLO-5, PLO-6, PLO-9, PLO-10
A5: Comprehensive Project	30%	Individual	In Examination Week	PLO-5, PLO-6, PLO-7, PLO-9, PLO-10

Description of Assessments:

A1- Class participation: You are expected to contribute in class discussions and participate in activities conducted in class pertaining to the course.

A2- Quiz 1: This component is designed to evaluate the comprehension of students on the different concepts, principles, and techniques covered in the course. This component may consist of multiple-choice questions, true/false statements, short answer questions, or a combination of these types of questions.

A3- Short reports: These reports assess students' understanding of key data mining concepts through short written reports on topics like Decision Trees, the Apriori Algorithm, and Naïve Bayes classification, focusing on their principles, strengths, and real-world applications. Students may also be required to analyze research papers, evaluating methodologies, effectiveness, and potential improvements. Assignments may take the form of essay-style reports, case studies, or problem-solving tasks, encouraging critical thinking and technical writing. Rather than relying on objective tests, this component emphasizes deep comprehension, analytical reasoning, and independent research, helping students connect theory with practice..

A4- Quiz 2: This component is designed to evaluate the comprehension of students on the different concepts, principles, and techniques covered in the course. This component may consist of multiple-choice questions, true/false statements, short answer questions, or a combination of these types of questions

A5- Comprehensive Project: Students will work on an end-to-end project, covering data collection, preprocessing, model training, evaluation, and presentation of findings. Evaluation will be based on

implementation, creativity, accuracy, and the depth of insights derived from the analysis. This project encourages students to apply their knowledge, think critically, and showcase their ability to solve complex data-driven problems. Students are encouraged to use tools such as WEKA (in case of complete unfamiliarity with the python programming) or Python with libraries like pandas and numpy for implementation.

Teaching Method

The course will have a judicious mix of lectures, storytelling, experiential exercises, and cases. Here the onus of learning will be with the student, and the instructor will be a facilitator. Instead of learning 'what to do', the cases will also be used as examples of real-world phenomena where issues arise, and good and bad practices are seen. The key to learning this way is to see many examples and situations and learn inductive as well as deductive ways from students' and managers' different experiences.

Textbook / Other Readings

Textbook: Han, J., Pei, J., & Kamber, M. (2012). Data Mining: Concepts and Techniques (3rd ed.). Morgan Kaufmann.

Session Plan

Session Details	**Topics**	**PLOs Covered**
Session 1	**Introduction to Data mining**	PLO-5, PLO-6, PLO-7, PLO-8, PLO-9
Objective of the session	To introduce data mining, this session explains why it is crucial in the modern era, highlighting the shift toward the Information Age and its evolution within information technology. It defines data mining and explores the types of data that can be mined, including database data, data warehouses, transactional data, and other relevant forms.	
Subtopics to be covered	• *Why Data Mining?* • *Moving toward the Information Age* • *Data Mining as the Evolution of Information Technology* • *What Is Data Mining?* • *What Kinds of Data Can Be Mined?* • *Database Data* • *Data Warehouses* • *Transactional Data* • *Other Kinds of Data*	
Case Title & Number	Cases will be shared in the class	
Readings	Textbook: Sections 1.1, 1.2, and 1.3 (pgs 1 – 14)	
Pedagogy	Lectures and class discussions	
Session 2	**Introduction to Data Mining –**	PLO-5, PLO-6, PLO-7, PLO-8,
Objective of the session	To provide an overview of the different types of patterns that	

	can be uncovered through data mining and to introduce the core technologies—such as statistics, machine learning, and database systems—that support the extraction and analysis of these patterns.	PLO-9
Subtopics to be covered	• *What Kinds of Patterns Can Be Mined?* • *Which Technologies Are Used?*	
Readings	Textbook: Sections 1.4 and 1.5 (pgs 15 – 26)	
Case Title & Number	Cases will be shared in the class	
Pedagogy	Lectures, cases, and class discussions	
Session 3	**Introduction to Data Mining - 3**	PLO-5, PLO-6, PLO-7, PLO-8, PLO-9
Objective of the session	To discuss the practical applications of data mining in areas like business intelligence and web search engines, and to examine the major challenges faced in data mining, including methodology, user interaction, scalability, diverse data types, and societal implications.	
Subtopics to be covered	• *Which Kinds of Applications Are Targeted?* • *Major Issues in Data Mining*	
Readings	Textbook: Sections 1.6 and 1.7 (pgs 27 – 32)	
Case Title & Number	Cases will be shared in the class	
Pedagogy	Lectures, cases, and class discussions	
Session 4	**Getting to know your data - 1**	PLO-5, PLO-6, PLO-7, PLO-8, PLO-9
Objective of the session	To understand different types of data attributes—such as nominal, binary, ordinal, and numeric—and how to distinguish between discrete and continuous attributes, along with gaining foundational knowledge of basic statistical measures like mean, median, mode, range, variance, and standard deviation for describing and summarizing data.	
Subtopics to be covered	• *Data Objects and Attribute Types* • *Basic Statistical Descriptions of Data*	
Readings	Textbook: Sections 2.1 and 2.2 (pgs 40 – 51)	
Case Title & Number	Cases will be shared in the class	
Pedagogy	Lectures, cases, and class discussions	
Session 5	**Getting to know your data - 2**	PLO-5, PLO-6, PLO-7, PLO-8, PLO-9
Objective of the session	To introduce various data visualization techniques for representing complex data and relationships, and to explain how similarity and dissimilarity between data objects are measured using proximity metrics for different attribute types, including nominal, binary, ordinal, numeric, and mixed attributes.	

Subtopics to be covered	• *Data Visualization* • *Measuring Data Similarity and Dissimilarity*	
Readings	Textbook: Sections 2.3 and 2.4 (pgs 56 – 77)	
Case Title & Number	Cases will be shared in the class	
Pedagogy	Lectures, cases, and class discussions	
Session 6	**Data preprocessing - 1**	PLO-5, PLO-6, PLO-7, PLO-8, PLO-9, PLO-10
Objective of the session	To highlight the importance of data preprocessing for ensuring data quality and to explore key preprocessing tasks, focusing on data cleaning methods for handling missing values and noisy data to prepare reliable input for analysis.	
Subtopics to be covered	• *Data Preprocessing* • *Data Cleaning*	
Readings	Textbook: Section 3.1 and 3.2 (pgs 84 – 91)	
Case Title & Number	Cases will be shared in the class	
Pedagogy	Lectures, cases, and class discussions	
Session 7	**Data preprocessing - 2**	PLO-5, PLO-6, PLO-7, PLO-8, PLO-9, PLO-10
Objective of the session	To understand techniques for integrating data from multiple sources while resolving issues like redundancy and conflicts, and to explore various data reduction strategies—such as PCA, sampling, and clustering—that help in compressing large datasets without losing significant information.	
Subtopics to be covered	• *Data Integration* • *Data Reduction*	
Readings	Textbook: Sections 3.3 and 3.4 (pgs 93 – 110)	
Case Title & Number	Cases will be shared in the class	
Pedagogy	Lectures, cases, and class discussions	
Session 8	**Data preprocessing - 3**	PLO-5, PLO-6, PLO-7, PLO-8, PLO-9, PLO-10
Objective of the session	To explore various strategies for transforming data—such as normalization and concept hierarchy generation—and discretization techniques like binning, histogram analysis, clustering, and decision trees, which help in preparing data for more effective analysis and modeling.	
Subtopics to be covered	• *Data Transformation* • *Data Discretization*	
Readings	Textbook: Section 3.5 (pgs 111 – 117)	
Session 9	**Data Warehousing and Online Analytical Processing - 1**	PLO-5, PLO-6, PLO-7, PLO-8,
Objective of the session	To introduce the foundational concepts of data warehousing,	

	highlighting its architecture, types, and key components like ETL and metadata, while differentiating it from operational databases and explaining its role in strategic decision-making.	PLO-9, PLO-10
Subtopics to be covered	• *Data Warehouse: Basic Concepts*	
Readings	Textbook: Section 4.1 (pgs 125 – 134)	
Case Title & Number	Cases will be shared in the class	
Pedagogy	Lectures, cases, and class discussions	
Session 10	**Data Warehousing and Online Analytical Processing - 2**	PLO-5, PLO-6, PLO-7, PLO-8, PLO-9
Objective of the session	To explore the concepts of multidimensional data models and OLAP, including data cubes, schemas (star, snowflake, and fact constellations), and the role of dimensions and measures in OLAP operations, providing a framework for analyzing complex data in data warehouses.	
Subtopics to be covered	• *Data Warehouse Modeling: Data Cube and OLAP*	
Readings	Textbook: Section 4.2 (pgs 135 – 146)	
Case Title & Number	Cases will be shared in the class	
Pedagogy	Lectures, cases, and class discussions	
Session 11	**Data Warehousing and Online Analytical Processing - 3**	PLO-5, PLO-6, PLO-7, PLO-8, PLO-9
Objective of the session	To understand the process of designing a data warehouse using a business analysis framework, and explore how data warehouses are utilized for information processing, transitioning from OLAP to multidimensional data mining for deeper insights.	
Subtopics to be covered	• *Data Warehouse Design and Usage*	
Readings	Textbook: Section 4.3 (pgs 150 – 155)	
Case Title & Number	Cases will be shared in the class	
Pedagogy	Lectures, cases, and class discussions	
Session 12	Hands on session	
Subtopics to be covered	The Hands on session for the Data Mining course provides students with a hands-on programming experience in Python. This session allows students to apply the theoretical concepts and techniques learned in the course to practical scenarios and gain valuable experience in data mining programming.	
Readings	Sessions covered upto session 11	
Case Title & Number	N/A	
Pedagogy	Hands on programming sessions	

Session 13	**Mining Frequent Patterns, Associations, and Correlations: Basic Concepts and Methods - 1**	
Objective of the session	To introduce the fundamental concepts of frequent itemsets, closed itemsets, and association rules, with a focus on market basket analysis and the Apriori algorithm for mining frequent itemsets through candidate generation methods.	PLO-5, PLO-6, PLO-7, PLO-8, PLO-9
Subtopics to be covered	• *Mining Frequent Patterns, Associations, and Correlations: Basic Concepts and Methods*	
Readings	Textbook: Sections 6.1 and 6.2 (pgs 243 – 253)	
Case Title & Number	Cases will be shared in the class	
Pedagogy	Lectures, cases, and class discussions	
Session 14	**Mining Frequent Patterns, Associations, and Correlations: Basic Concepts and Methods - 2**	PLO-5, PLO-6, PLO-7, PLO-8, PLO-9, PLO-10
Objective of the session	To explore methods for generating association rules from frequent itemsets, focusing on improving Apriori's efficiency, pattern-growth approaches, and vertical data format techniques, while also discussing the mining of closed and max patterns for more insightful rule generation.	
Subtopics to be covered	• *Generating Association Rules from Frequent Itemsets*	
Readings	Textbook: Section 6.2 (pgs 254 – 262)	
Case Title & Number	Cases will be shared in the class	
Pedagogy	Lectures, cases, and class discussions	
Session 15:	**Guest lecture 1: Mr. Vikram Arora (Lead consultant, ITC Infotech)**	
Objective of the session	Guest lecture on understanding real life application of Data mining in industries	
Subtopics to be covered	Data mining role in businesses to understand customer behavior, optimizing processes, and improving decision-making, ultimately driving innovation and efficiency.	
Readings	N/A	
Case Title & Number	N/A	
Pedagogy	N/A	
Session 16	**Mining Frequent Patterns, Associations, and Correlations: Basic Concepts and Methods – 3**	
Objective of the session	To discuss how strong association rules may not always be interesting and to shift focus from association analysis to	PLO-5, PLO-6, PLO-7, PLO-8,

	correlation analysis, while comparing various pattern evaluation measures to identify truly valuable patterns.	PLO-9
Subtopics to be covered	• *Which Patterns Are Interesting?—Pattern Evaluation Methods*	
Readings	Textbook: Section 6.3 (pgs 264 – 270)	
Case Title & Number	Cases will be shared in the class	
Pedagogy	Lectures, cases, and class discussions	
Session 17	**Classification – Basic concepts - 1**	PLO-5, PLO-6, PLO-7, PLO-8, PLO-9
Objective of the session	To introduce the basic concepts of classification and the general approach to it, with a focus on decision tree induction, including attribute selection measures, tree pruning, scalability concerns, and visual mining for decision tree induction.	
Subtopics to be covered	• *Fundamental concepts of Classification*	
Readings	Textbook: Sections 8.1 and 8.2 (pgs 327 – 349)	
Case Title & Number	Cases will be shared in the class	
Pedagogy	Lectures, cases, and class discussions	
Session 18	**Classification – Basic concepts - 2**	PLO-5, PLO-6, PLO-7, PLO-8, PLO-9
Objective of the session	To explore Bayes' Theorem and its application in Naïve Bayesian classification, along with rule-based classification methods, including using IF-THEN rules, extracting rules from decision trees, and employing sequential covering algorithms for rule induction.	
Subtopics to be covered	• *Bayes Classification Methods*	
Readings	Textbook: Section 8.4 (pgs 355 – 363)	
Case Title & Number	Cases will be shared in the class	
Pedagogy	Lectures, cases, and class discussions	
Session 19	**Classification – Basic concepts - 3**	PLO-5, PLO-6, PLO-7, PLO-8, PLO-9
Objective of the session	To discuss various metrics for evaluating classifier performance, including the holdout method, random subsampling, cross-validation, bootstrap, and statistical tests of significance. Additionally, the session covers model selection through cost-benefit analysis and ROC curves.	
Subtopics to be covered	• *Model Evaluation and Selection*	
Readings	Textbook: Section 8.5 (pgs 364 – 373)	
Case Title & Number	Cases will be shared in the class	
Pedagogy	Lectures, cases, and class discussions	
Session 20	**Cluster analysis - I**	PLO-5, PLO-6,

Objective of the session	To introduce the concept and objectives of cluster analysis, explore the essential requirements for effective clustering, and provide an overview of basic clustering methods. The session emphasizes partitioning techniques, particularly the k-Means algorithm, explaining how it partitions data into clusters based on centroid optimization.	PLO-7, PLO-8, PLO-9
Subtopics to be covered	• *Cluster Analysis* • *Partitioning Methods* • *k-Means*	
Readings	Textbook: Section 10.1 and 10.2 (pgs 444 – 453)	
Case Title & Number	Cases will be shared in the class	
Pedagogy	Lectures, cases, and class discussions	
Session 21	Hands on session	PLO-5, PLO-6, PLO-7, PLO-8, PLO-9
Objective of the session	This session for the Data Mining course provides students with a hands-on programming experience in Python. This session allows students to apply the theoretical concepts and techniques learned in the course to practical scenarios and gain valuable experience in data mining programming.	
Subtopics to be covered	Sessions covered upto session 20	
Readings	Sessions covered upto session 20	
Case Title & Number	N/A	
Pedagogy	Hands on programming sessions	
Session 22	**Guest lecture 1: Mr. Dhruv Singla (Office Sahayogi and JaiParkashSingla Legal & Consulting LLP)**	
Objective of the session	Guest lecture on understanding real life application of Data mining in industries	
Subtopics to be covered	Data mining role in businesses to understand customer behavior, optimizing processes, and improving decision-making, ultimately driving innovation and efficiency.	
Readings	N/A	
Case Title & Number	N/A	
Pedagogy	N/A	
Session 23	**Cluster analysis - 2**	PLO-5, PLO-6, PLO-7, PLO-8, PLO-9
Objective of the session	To compare two main approaches in hierarchical clustering—agglomerative (bottom-up) and divisive (top-down)—highlighting their processes and use cases. The session also aims to familiarize students with different distance measures used in clustering algorithms, which play a crucial role in determining the similarity between data points and thus	

	influence the clustering results.	
Subtopics to be covered	• *Agglomerative vs. Divisive Hierarchical Clustering* • *Distance Measures in Algorithmic Methods*	
Readings	Textbook: Section 10.3 (pgs 459 – 463)	
Case Title & Number	Cases will be shared in the class	
Pedagogy	Lectures, cases, and class discussions	
Session 24	**Cluster analysis - 3**	PLO-5, PLO-6, PLO-7, PLO-8, PLO-9
Objective of the session	The session introduces DBSCAN, a powerful density-based clustering algorithm that identifies clusters of arbitrary shapes and handles noise effectively.	
Subtopics to be covered	• *DBSCAN: Density-Based Clustering Based on Connected Regions with High Density*	
Readings	Textbook: Section 10.4 (pgs 471 – 473)	
Case Title & Number	Cases will be shared in the class	
Pedagogy	Lectures, cases, and class discussions	
Session 25	**Cluster analysis - 4**	PLO-5, PLO-6, PLO-7, PLO-8, PLO-9
Objective of the session	To explores key aspects of evaluating clustering outcomes, such as assessing whether data has a natural tendency to cluster, selecting the optimal number of clusters, and measuring the quality of the clustering results.	
Subtopics to be covered	• *Evaluation of Clustering*	
Readings	Textbook: Section 10.6 (pgs 483 – 489)	
Case Title & Number	Cases will be shared in the class	
Pedagogy	Lectures, cases, and class discussions	
Session 26	**Outlier detection - 1**	PLO-5, PLO-6, PLO-7, PLO-8, PLO-9
Objective of the session	This session focuses on understanding outliers—data points that deviate significantly from the rest of the dataset. It explores the different types of outliers, such as point, contextual, and collective outliers, and delves into the challenges associated with their detection.	
Subtopics to be covered	• *Outliers and Outlier Analysis*	
Readings	Textbook: Section 12.1 (pgs 544 – 549)	
Case Title & Number	Cases will be shared in the class	
Pedagogy	Lectures, cases, and class discussions	
Session 27	**Outlier detection - 2**	PLO-5, PLO-6, PLO-7, PLO-8,
Objective of the session	This session introduces various methods for detecting outliers in data. It covers the distinctions between supervised, semi-	

	supervised, and unsupervised approaches, along with commonly used techniques such as statistical, proximity-based, and clustering-based methods.	PLO-9
Subtopics to be covered	• *Outlier Detection Methods*	
Readings	Textbook: Section 12.2 (pgs 549 – 552)	
Case Title & Number	Cases will be shared in the class	
Pedagogy	Lectures, cases, and class discussions	
Session 28	**Hands on session**	
Objective of the session	This session for the Data Mining course provides students with a hands-on programming experience in Python. This session allows students to apply the theoretical concepts and techniques learned in the course to practical scenarios and gain valuable experience in data mining programming.	PLO-5, PLO-6, PLO-7, PLO-8, PLO-9
Subtopics to be covered	Sessions covered upto session 28	
Readings	Sessions covered upto session 28	
Case Title & Number	N/A	
Pedagogy	Hands on programming sessions	
Session 29	Reading and Revision Period	N/A
Objective of the session	**Course Revision**	
Subtopics to be covered	N/A	
Readings	N/A	
Case Title & Number	N/A	
Pedagogy	N/A	
Session 30	Reading and Revision Period	N/A
Objective of the session	**Course Revision**	
Subtopics to be covered	N/A	
Readings	N/A	
Case Title & Number	N/A	
Pedagogy	N/A	

Disability Support

JGU endeavours to make all its courses accessible to students. The Disability Support Committee (DSC) has identified conditions that could hinder a student's overall wellbeing. These include physical and mobility-related difficulties, visual impairment, hearing impairment, mental health conditions, and intellectual/learning difficulties, e.g., dyslexia and dyscalculia. Students with any known disability needing academic and other support are required to register with the Disability Support Committee (DSC) by following the procedure specified at **https://jgu.edu.in/disability-support-committee/**

Students who need support may register any time during the semester up until a month before the end semester exam begins. Those students who wish to continue receiving support from the previous semester, must re-register within the first month of a semester. Last-minute registrations and support might not be possible as sufficient time is required to make the arrangements for support.

The DSC maintains strict confidentiality about the identity of the student and the nature of their disability and the same is requested from faculty members and staff as well. The DSC takes a strong stance against in-class and out-of-class references made about a student's disability without their consent and disrespectful comments referring to a student's disability.

All general queries are to be addressed to **disabilitysupportcommittee@jgu.edu.in**