Tribhuvan University

Faculty of Management

Office of the Dean

Course detail of
MITM (Master of Information Technology Management) 2nd Semester

August 2026

ECO 551: Economics for Business

MITM 2nd Semester

Credits: 2
Lecture Hours: 32

Course Objectives

This course aims to enhance students' understanding of economic concepts, theories, and their practical applications, enabling them to apply cognitive knowledge and practical skills in the fields of information technology and management.

Course Description

This course encompasses market mechanisms, efficiency and welfare, the theory of the firm and pricing, estimation and forecasting in managerial decisions, and the seminar on contemporary issues of digital economics and digitalization.

Course Details

Unit 1: Market mechanism, efficiency, and welfare **8 LHs**

Fundamental economic principles: how people make decisions, how people interact, and how the economy as a whole works; applications of economic principles in managerial decisions; Markets and efficiency: Demand and supply functions, Market equilibrium and welfare, effects of government policies (tax, subsidy, and price control measures) on market equilibrium, efficiency, and welfare; and Market failure: Concepts and causes

Learning Outcomes

1. Examine the relationship between fundamental economic principles and managerial decisions
2. Explain demand and supply functions and the market equilibrium
3. Measure market efficiency and welfare
4. Analyze the effects of government policies on the market equilibrium.
5. Explain the sources of market failure

Unit 2: Theory of the firm and pricing **8 LHs**

Concept of production and cost functions; Goals of the firm and applications: profit maximization, sales revenue maximization, Williamson's model of managerial discretion; Games and strategy: strategy, Nash equilibrium, prisoner's dilemma, and applications of games in strategy formulation; Pricing practices: price discrimination, cost-plus pricing, incremental cost pricing, freemium pricing, usage-based pricing, dynamic pricing, subscription pricing, bundling.

Learning Outcomes

1. Analyze the goals of firms with their practical applications.
2. Explain pricing practices.
3. Apply game theory in strategy formulation.

Unit 3: Estimation and forecasting in managerial decisions 4 LHs

Demand forecasting: concepts and significance, Survey method of forecasting: consumer survey and market survey; Quantitative forecasts: time series analysis, Demand estimation: regression analysis, elasticity of demand.

Learning Outcome

1. Forecast the demand by using the qualitative and quantitative methods

Unit-4: Seminar on Contemporary Issues of Digital Economics and Digitalization 12 LHs

Asymmetric information: concepts, determinants, and applications; Asymmetric information and digitalization; Economics of digital components(platforms, networks, lock-in and switching costs, artificial intelligence); Institutional economics in the digital era; Effects of digitalization on consumer choice and labor markets; Intellectual property and digitalization; E-governance: concepts, process and structure; E-governance practices in Nepal; Implications of digital economy and financial innovation on monetary policy and financial markets; Transformation of economic growth theories in the context of the digital economy; Opportunities and challenges for digital transformation in public sector, trade, tourism, agriculture and SMEs.

Learning Outcome:

1. Analyze the contemporary issues of digital economics and digitalization and their practical implications.

(***Students have to write a seminar paper on contemporary issues in digital economics and digitalization)***

Pedagogical Strategies

- Interactive lectures with demonstration
- Problem-based learning
- Guest lectures from industry experts and practitioners
- Continuous assessment and feedback
- Case analysis with group presentation
- Article review

Assessment Methods

a. Internal Assessment Methods and Types

Assessment Type	Weightage	Details
Class participation & attendance	10%	Contribution to discussions, engagement in class activities
Quizzes/short tests	5%	Periodic quizzes to assess comprehension
Practical/Project	25%	Presentation on case analysis and article review
Seminar Paper Writing	30%	Viva-Voce (Presentation)
Pre-board examination	30%	Comprehensive written test covering all units

b. External Assessment Methods and Types

Department/College will arrange the final Viva-Voce examination for the seminar with the presence of experts from the Business Economics Subject Committee, Faculty of Management, TU, to test students' understanding, application, analysis, evaluation, and creation on research-based learning. Department/College must conduct the seminar before the final examination.

Mapping Course: Learning Outcomes and Program Learning

Course Learning Objective	Knowledge (K)	Skills (S)	Competence (C)
Total Learning	40%	40%	20%
(CLO) Dimensions			

Suggested Readings

Salvatore, D. & Rastogi, S.K. (2017). *Managerial economics*. 8th edit New Delhi: Oxford University Press

Truett, L.J. & Truett, D.B. (2004). *Managerial economics: Analysis, problems, cases*. Hoboken, NJ: Wiley.

McConnell, C. R., Brue, S. L. (2013***).*** *Economics: Principles, problems, and policies*. 19th edition. New Delhi: McGraw-Hill Education.

Hirschey, M. & Pappas, J.L. (1989). *Fundamentals of managerial economics*. New York: The Dryden Press.

Thomas, C.R. & Maurice, S.C. (2015). *Managerial economics: Concepts and applications*. New Delhi: McGraw-Hill Education.

Adhikari, G.M., Paudel, R.K., & Regmi, K. (2018). *Managerial economics*. Kathmandu: KEC Publications.

Overby, H. Introduction to Digital Economics, Springer, Latest Edition

Maheshwari, A. Digital Transformation, Wiley, Latest Edition

Reports and publications of Nepal Rastra Bank, National Planning Commission, IIDS, OECD, World Bank,

FIN 551: Financial Management

MITM 2nd Semester

Credits: 2
Lecture Hours: 32

Course Description

This course provides an understanding of financial management with a specific focus on decision-making in technology-driven organizations. The course aims to equip students with analytical tools to evaluate IT investments, manage financial risks, and support strategic decision-making in dynamic technological environments. It covers introduction to financial management, financial statement analysis, capital structure and cost of capital, investment decision, dividend decision, and working capital management.

Learning Outcomes:

Upon completion of this course, students will be able to:

- Understand and apply financial management concepts to IT-based organizations and projects
- Analyze and interpret financial statements using analytical tools and ratios
- Assess risk and return using portfolio and CAPM models
- Estimate cost of capital and apply it in investment decisions
- Evaluate IT investment projects using capital budgeting techniques
- Examine dividend decision, and
- Understand working capital management and analyze working capital requirements.

Unit 1: Financial Management and Environment **4 LHs**

Financial Management: Concept and functions, Objectives of corporation: Value maximization, value maximization and social welfare, Managerial action to maximize shareholder wealth; Role of financial manager; Finance within an organization; Financial environment in Nepal: Financial securities, financial institutions, financial market; tax environment, Emerging trends in financial management and IT.

Unit 2: Financial Statement Analysis **5 LHs**

sFinancial statements and reports; Financial ratio analysis: Liquidity ratios, asset management ratios, debt management ratios, profitability ratios, market value ratios; Trend analysis; Common size analysis; Percentage change analysis; The Du-Pont equation; Uses and limitations of ratio analysis.

Unit 3: Risk-Return and Portfolio Theory **5 LH**

sReturns and risk; Risk in a portfolio context; Efficient portfolio and efficient frontier; The Optimal portfolio; Capital Assets Pricing Model (CAPM); The Capital Market Line (CML) and Security Market Line (SML).

Unit 4: Capital Structure and Cost of Capital **5 LHs**

Raising long-term fund: long-term debt, Preferred stock and common stock; Capital Structure and financial structure; The optimal capital structure; Business bisk and financial risk; Component cost of capital; Weighted Average Cost of Capital (WACC).

Unit 5: Investment Decision **5 LHs**

Capital budgeting decision: Concept and process. Estimating cash flows; Evaluating the project: Payback period, discounted payback period, net present value, internal rate of return, profitability index, modified internal rate of return using MS excel.

Unit 6: Dividend Decision **4 LHs**

sDividends: Concept, types and payment procedures; Residual dividend policy and dividend stability; Stock dividends and Stock splits; Stock repurchase.

Unit 7: Working Capital Management **4 LHs**

Working Capital and working capital management: Concept, importance; factors affecting the size of working capital; The cash conversion cycle; Concept and significance of cash management, inventory management and receivable management.

Suggesting Readings:

Bigham, E. F & Ehrhardt, M. C. ***Financial Management: Theory and Practice.*** New Delhi: Cengage Learning**.**

Ross, S. A., Westerfield, R. W., & Jordan, B. D. **Corporate Finance.** New Delhi: Tata McGraw-Hill.

Gitman, L. J. ***Principles of Managerial Finance***. Delhi: Pearson Education.

MIT 551: Big Data Technologies

MITM 2nd Semester

Credits: 3
Lecture Hours: 48

Course Objectives

By the end of this course, students will be able to understand the fundamentals of Big Data, its characteristics, scope, applications, and challenges, set up a Hadoop cluster and perform basic file operations in HDFS, write and execute MapReduce programs to solve real-world Big Data problems in a Hadoop environment, understand and apply different types of NoSQL databases (including MongoDB and HBase) for structured, unstructured, and semi-structured data analysis, perform real-time and batch analytics using Apache Spark, and execute SQL-like queries on top of Hadoop using Spark SQL, Pig, and Hive.

Course Description

This course introduces students to the fundamental concepts and practical skills required to handle the challenges of Big Data. It begins with an overview of Big Data characteristics, scope, applications, and current trends, then progresses through foundational platforms such as Hadoop, MongoDB, and HBase as NoSQL databases, and advanced processing frameworks including Spark, Hive, and Pig. Students will then have fundamental knowledge on Big Data Analytics to solve various real-world problems posed by Big Data.

Course Details

Unit 1: Introduction to Big Data **4 LHs**

Introduction to Big Data; Characteristics of Big Data (Volume, Velocity, Variety, Veracity, Value); Current trends and real-life applications of Big Data; Scope and challenges of Big Data; Tools and technologies used in Big Data.

Learning Outcomes: Understand the concept of Big Data, its defining characteristics, scope, and challenges (Knowledge); Identify and evaluate appropriate tools and technologies for addressing Big Data problems (Skills); Analyze real-world Big Data scenarios to assess their scope and technological requirements (Competency).

Unit 2: Hadoop Ecosystem **10 LHs**

Introduction to Hadoop; History of Hadoop; Hadoop ecosystem overview; Core components of the Hadoop ecosystem: Hadoop Common, Hadoop Distributed File System (HDFS), MapReduce Framework, and YARN (Yet Another Resource Negotiator); Hadoop master/slave architecture; Hadoop daemons; Hadoop configuration modes; Hadoop cluster setup; Hadoop Streaming.

Learning Outcomes: Understand the Hadoop ecosystem, its architecture, core components, and configuration modes (Knowledge); Install, configure, and manage a Hadoop cluster; use YARN and Hadoop daemons effectively (Skills); Deploy and maintain a Hadoop environment capable of distributed data storage and processing (Competency).

Unit 3: Distributed Storage and Processing of Big Data **12 LHs**

HDFS: The design of HDFS, HDFS concepts, command-line interface, Hadoop file system interfaces, data flow, basic HDFS commands. MapReduce: Functional programming, MapReduce fundamentals, execution overview of MapReduce, basic MapReduce API concepts,

setting up the development environment, writing unit tests with MRUnit, running locally on test data, running on a cluster, anatomy of a MapReduce job run, failures, shuffle and sort, task execution, MapReduce types and formats.

Learning Outcomes: Understand HDFS design, concepts, data flow, and the MapReduce programming paradigm (Knowledge); Perform HDFS file operations via the command line; write, test, and execute MapReduce programs in Java/Python on a Hadoop cluster (Skills); Build and run distributed data processing pipelines using HDFS and MapReduce for real-world Big Data problems (Competency).

Unit 4: NoSQL Databases **10 LHs**

Types of data; Introduction to NoSQL; Need for NoSQL; Types of NoSQL databases; NoSQL vs. relational databases; The CAP theorem. MongoDB: Collections, documents, object IDs, queries on MongoDB, aggregation pipeline, nested documents. HBase: Overview, HBase vs. RDBMS, HBase vs. HDFS, HBase architecture, HBase data model, concept of row keys and column families, HBase regions, creating a table, writing queries to insert and retrieve data to and from HBase.

Learning Outcomes: Understand NoSQL database types, the CAP theorem, and the architectures of MongoDB and HBase (Knowledge); Perform CRUD operations in MongoDB and HBase; design schemas using row keys and column families; execute aggregation pipelines (Skills); Select and apply appropriate NoSQL database solutions to diverse Big Data storage and query requirements (Competency).

Unit 5: Data Analytics with Spark **6 LHs**

Introduction to Spark; Need for Spark; Evolution of Spark; Spark shell; Spark context; Resilient Distributed Dataset (RDD); Transformations; Programming with RDD; Spark Core; Spark SQL; MLlib; Spark Streaming; GraphX.

Learning Outcomes: Understand the Spark architecture, RDD model, and its ecosystem components (Knowledge); Write Spark programs using RDDs and DataFrames; perform SQL-based analytics; implement real-time streaming jobs (Skills); Develop scalable, real-time, and batch data analytics applications using Apache Spark (Competency).

Unit 6: Querying Big Data with Pig and Hive **6 LHs**

Pig: Introduction to Pig; execution modes of Pig; comparison of Pig with databases; Grunt; Pig Latin; user-defined functions; data processing operators. Hive: Hive shell; Hive services; Hive Metastore; comparison with traditional databases; HiveQL; tables; querying data; user-defined functions.

Learning Outcomes: Understand the architecture and query paradigms of Apache Pig and Apache Hive (Knowledge); Write Pig Latin and HiveQL scripts to process and query large datasets stored in HDFS; create user-defined functions (Skills); Implement SQL-like analytical workflows on Hadoop using Pig and Hive for large-scale data processing tasks (Competency).

Pedagogical strategies

- Lectures with demonstration
- Hands-on lab sessions
- Problem-based learning

- Guest lectures from industry experts in data engineering
- Continuous assessment and feedback
- Multimedia presentations to visualize concepts
- Mini project

Internal assessment methods and types

Assessment Type	Weightage	Details
Class participation & attendance	10%	Contribution to discussions, engagement in class activities
Quizzes/short tests	15%	Periodic quizzes to assess comprehension
Practical/Project	20%	Lab sessions and mini project
Mid-term examination	25%	Written test
Pre-board examination	30%	Comprehensive written test covering all units

External assessment methods and types

Dean's office will arrange the final/board examination in writing to test understanding, application, analysis, evaluation, and creation.

Mapping course: Learning outcomes and program learning

Course Learning Objective (CLO) Dimensions	Knowledge (K)	Skills (S)	Competence (C)
Total Learning	35%	40%	25%

Laboratory Work

Students will gain hands-on experience through practical exercises that help strengthen their understanding of Big Data technologies. This includes setting up Hadoop clusters and performing HDFS operations, writing and running MapReduce programs, performing CRUD operations on MongoDB and HBase, analysing data using Spark SQL and Spark Streaming, running Pig and Hive queries, and completing a project that applies multiple Big Data technologies to real-world datasets. These exercises aim to provide practical experience in implementing Big Data solutions across multiple platforms.

Suggested Readings

White, T. (2023). Hadoop: The Definitive Guide (5th ed.). O'Reilly Media.

Chambers, B., & Zaharia, M. (2018). Spark: The Definitive Guide: Big Data Processing Made Simple. O'Reilly Media.

Lam, C. (2023). MongoDB: The Definitive Guide (3rd ed.). O'Reilly Media.

Frampton, M. (2022). Mastering Apache Spark 3: Build structured and unstructured data pipelines. Packt Publishing.

Capriolo, E., Wampler, D., & Rutherglen, J. (2022). Programming Hive: Data Warehouse and Query Language for Hadoop (2nd ed.). O'Reilly Media.

Gates, A., & Noll, D. (2020). Programming Pig: Dataflow Scripting with Hadoop (2nd ed.). O'Reilly Media.

Tiwari, S. (2023). Professional NoSQL. Wiley.

Hurwitz, J., Nugent, A., Halper, F., & Kaufman, M. (2021). Big Data for Dummies (2nd ed.). Wiley.

***

MIT 552: Applied Data Science

MITM 2nd Semester

Credits: *3*
Lecture ***Hours:*** *48*

Course Objectives

By the end of this course, students will be able to understand the core concepts of data science and its role in organizational decision-making, apply Python-based tools such as NumPy, Pandas, Matplotlib, and Seaborn for data manipulation and visualization, perform exploratory data analysis (EDA) to extract insights from real-world datasets, understand the fundamentals of statistical thinking and apply it to business problems, grasp the basic principles of machine learning and how predictive models work, and practice responsible and ethical use of data in organizational contexts.

Course Description

This course introduces students to applied data science from a management and business perspective. It covers the fundamentals of data handling and analysis using Python, including data collection, cleaning, and exploratory analysis. Students will learn to visualize data effectively and apply basic statistical thinking to interpret business data. The course also introduces the foundational concepts of machine learning: how models learn from data and make predictions. Emphasis is placed on practical skills, ethical data use, and applying data science thinking to technology management contexts.

Course Details

Unit 1: Introduction to Data Science **6 LHs**

Introduction to Data Science; Data-Driven Decision Making; Data Science Workflow: Business Problem, Data, Exploration and Analysis, Modeling, Insight, Decision; Data Science life-cycle: CRISP-DM, TDSP; Key Roles: Data Engineer, Data Analyst, Machine Learning Engineer, Data Scientist; Overview of Tools: Python Ecosystem, Jupyter Notebook, Common Libraries; Setting up the Environement; Anaconda, Jupyter Notebook, Installing Libraries-Numpy, Pandas, Matplolib, Seaborn, Scikit-Learn;

Learning Outcomes: Understand data science concepts, how it relates to business intelligence and AI, and the typical data science workflow and team roles (Knowledge); Set up a Python data science environment and write basic Python code for data handling tasks (Skills); Appreciate the strategic value of data science in technology-enabled organizations (Competency).

Unit 2: Data Collection and Preparation **10 LHs**

Types of Data: Structured, Semi-structured and Unstructured; Commonly used Data Formats: CSV files, Excel, JSON, web APIs, Databases; Loading and Inspecting data with Pandas (read_csv, head, info, describe); Data Quality;
Common Issues with Real World Data; Data Cleaning Techniques; Data Validation;

Data Transformation: Filtering, Sorting, Merging, Normalization, Standardization, Encoding Categorical Variables; Feature Engineering: Feature Generation, Feature Selection, Feature Extraction;

Learning Outcomes: Understand the types and sources of data, and the dimensions of data quality; recognize common data quality problems (Knowledge); Load datasets from various sources, inspect and profile data, handle missing values and duplicates, transform and prepare data for analysis using Pandas (Skills); Produce clean, analysis-ready datasets from raw real-world data independently (Competency).

Unit 3: Exploratory Data Analysis and Visualization **8 LHs**

Purpose and Process of Exploratory Data Analysis (EDA), Descriptive Statistics: Mean, Median, Mode, Variance, Standard Deviation; Univariate Analysis: Histograms, Box Plots, Frequency Tables; Bivariate Analysis: Scatter Plots, Correlation, Cross-Tabulation;

Data Visualization; Visualizing Distributions and Comparisons: Bar Charts, Pie Charts, Line Plots, Histograms, Box Plots, Scatter Plots, Violin Plots, Pair plots and Interpretations; Principles of Effective Data Visualization; Chart Selection, Labeling, Color; Interpreting EDA findings for Business Insight

Learning Outcomes: Understand the goals of EDA, key descriptive statistics, and the principles of effective data visualization (Knowledge); Perform univariate and bivariate analysis on datasets; create and customize plots using Matplotlib and Seaborn; interpret visualizations to identify patterns and outliers (Skills); Conduct a complete exploratory analysis of a real-world dataset and communicate findings visually (Competency).

Unit 4: Statistics for Data Science **8 LHs**

Statistics and its Importance for Data Science; Probability Basics: Events, Outcomes and Probability Rules; Common Distributions: Normal Distribution and its Properties, Skewness; Sampling: Population vs Sample, Sampling Methods;

Confidence Intervals: Concept and Business Interpretation; Hypothesis Testing: Null and alternative Hypothesis, p-value, Significance Level; Practical Tests: t-test and chi-square test with SciPy; Interpreting Relationships between Variables: Correlation and Simple Linear Regression

Learning Outcomes: Understand key probability and statistics concepts including distributions, sampling, hypothesis testing, and correlation (Knowledge); Apply t-tests and chi-square tests using Python; interpret p-values, confidence intervals, and correlation coefficients in business contexts (Skills); Use statistical reasoning to evaluate data-driven claims and support business decisions (Competency).

Unit 5: Machine Learning **10 LHs**

Understanding Predictive analytics and Machine Learning; Artificial Intelligence vs Machine Learning; Machine Learning Techniques: Supervised, Unsupervised and Reinforcement learning and their types.

Regression Techniques: Linear Regression, Polynomial Regression; Classification Techniques: Logistic Regression, KNN, Decision Tree;

Model Evaluation: Train/test Split; Root Mean Squared Error, Mean Absolute Error, Mean Percentage Error; Confusion Matrix; Accuracy, Precision and Recall;
Unsupervised Learning: Clustering Intuition and K-Means

Learning Outcomes: Understand the core concepts of machine learning, the different types of ML, the model development workflow, and basic evaluation metrics (Knowledge); Build and evaluate a simple regression model and classification model using Scikit-learn; interpret model outputs in business terms (Skills); Explain how machine learning models work and apply basic predictive modeling to a business dataset using Python (Competency).

Unit 6: Ethical and Responsible Data Science **6 LHs**

Ethics for Data Scientist; Real-world Cases of Algorithmic Harm; Bias in Data and Models: Sources and Consequences; Common Cognitive Biases; Addressing Cognitive Bias; Data Privacy Principles and Regulations: GDPR overview, Right to Privacy

Responsible Data Collection: Consent, Anonymization, and Data Minimization

Transparency and Explainability; Ethical guidelines for AI and data science (OECD overview)

Learning Outcomes: Understand the ethical, privacy, and legal dimensions of data science and recognize sources of bias and harm in data-driven systems (Knowledge); Identify ethical risks in a data science scenario, evaluate data privacy compliance needs, and apply a responsible data handling checklist (Skills); Practice data science responsibly and advise organizations on ethical data governance as an IT manager (Competency).

Pedagogical Strategies

- Lectures with live coding demonstrations
- Hands-on lab sessions using Jupyter Notebook
- Case study analysis using real business datasets
- Problem-based mini assignments
- Guest lectures from industry practitioners
- Group discussions on data ethics and current trends
- Mini project covering the end-to-end data science workflow

Internal assessment methods and types

Assessment Type	Weightage	Details
Class participation & attendance	10%	Contribution to discussions and engagement in lab activities
Assignments	15%	Individual programming and analysis task
Practical / Mini Project	20%	Lab sessions and an end-to-end mini data science project on a real dataset
Mid-term Examination	25%	Written test
Pre-board Examination	30%	Comprehensive written test covering all units

External assessment methods and types

Dean's office will arrange the final/board examination in writing to test understanding, application, analysis, evaluation, and creation.

Mapping course: Learning outcomes and program learning

Course Learning Objective (CLO) Dimensions	Knowledge (K)	Skills (S)	Competence (C)
Total Learning	35%	40%	25%

Laboratory Work

Students will get hands-on experience through practical exercises that reinforce lecture content. Lab sessions are conducted in Jupyter Notebook using Python. Approximately 14 lab sessions are planned:

Lab #	Unit	Activity	Tools/Deliverable
L1	Unit 1	Environment setup; Introduction to Jupyter Notebook and Python review for data science	Anaconda, Jupyter; Setup verification notebook
L2	Unit 2	Loading data from CSV, Excel, and JSON files; initial data inspection and profiling with Pandas	Pandas; Inspection report
L3	Unit 2	Handling missing values and duplicates; correcting data types on a business dataset	Pandas; Cleaned dataset
L4	Unit 2	Data transformation: normalization, encoding categorical variables, feature creation	Pandas, Scikit-learn; Transformed dataset
L5	Unit 3	Descriptive statistics and univariate analysis: histograms, box plots, frequency tables	Pandas, Matplotlib; Summary notebook
L6	Unit 3	Bivariate analysis and correlation: scatter plots, correlation matrix, heatmap	Seaborn, Matplotlib; Visualization gallery
L7	Unit 3	Full EDA on a real business dataset: deriving and presenting insights visually	Pandas, Seaborn; EDA report
L8	Unit 4	Descriptive statistics and probability with Python; normal distribution visualization	NumPy, SciPy, Matplotlib; Lab notebook
L9	Unit 4	Hypothesis testing: t-test and chi-square test on business data; interpreting results	SciPy, Pandas; Statistical analysis report
L10	Unit 4	Correlation and simple linear regression: fitting a line, interpreting slope and R^2	Scikit-learn, Matplotlib; Regression notebook
L11	Unit 5	ML workflow: train/test split, fitting a Linear Regression model, evaluating MAE and R^2	Scikit-learn; Model evaluation notebook
L12	Unit 5	Classification with Logistic Regression:	Scikit-learn;

		predicting a business outcome; confusion matrix and accuracy	Classification report
L13	Unit 5	Unsupervised learning: K-Means clustering for customer segmentation; visualizing clusters	Scikit-learn, Seaborn; Clustering notebook
L14	All Units	Mini project: end-to-end data science project - data cleaning, EDA, a simple predictive model, and a summary report with visualizations	Full Python stack; Project report + presentation

Suggested Readings

Grus, J. (2019). *Data science from scratch: First principles with Python* (2nd ed.). O'Reilly Media.

Bruce, P., Bruce, A., & Gedeck, P. (2020). *Practical statistics for data scientists: 50+ essential concepts using R and Python* (2nd ed.). O'Reilly Media.

McKinney, W. (2022). *Python for data analysis: Data wrangling with pandas, NumPy, and Jupyter* (3rd ed.). O'Reilly Media.

Provost, F., & Fawcett, T. (2013). *Data science for business: What you need to know about data mining and data-analytic thinking*. O'Reilly Media.

VanderPlas, J. (2023). *Python data science handbook: Essential tools for working with data* (2nd ed.). O'Reilly Media.

Géron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*. O'Reilly.

***

MIT 553: Information Retrieval and Web Search
MITM 2nd Semester

***Credits**: 3*
***Lecture Hours**: 48*

Course Objectives

To equip students with a deep understanding of how modern search engines index, retrieve and rank the textual and semi-structured information. By the end of this course, graduate students should be able to

- Build and compress highly efficient inverted indices.
- Implement classical (Vector Space, Probabilistic) and modern (Transformer-based) retrieval models.
- Implement robust evaluation metrics on benchmark test collections.

Course Description

This course introduces students to the concepts of modern search engines and AI assistants that can find, rank, and deliver relevant information from massive collections of data. Students will be able to learn the entire pipeline of data retrieval: from classic text matching and web-crawling algorithm to modern, AI - powered semantic search. The curriculum bridges traditional mathematical ranking with AI technologies, including vector search, neural ranking, and Retrieval-Augmented Generation (RAG) for Large Language Models. Ultimately, it teaches students how to build systems that don't just match literal keywords, but truly understand the meaning and intent behind a user's query.

Course Details

Unit 1: Classical IR Foundations **12 LHs**

Introduction and Text Processing: IR Architecture, Information Need VS Queries, Tokenization, Lemmatization, stemming (Porter Stemmer), Stop Word Removal

Inverted Indexing: Inverted Index Construction, Skip List, Zipf's Law

IR Models: Boolean Retrieval Model, Term Frequency (TF), Inverse Document Frequency (IDF), TF – IDF, Cosine Similarity, Vector Space Model

Probabilistic Model: The Binary Independence Model (BIM), Probability Ranking Principle (PRP), BM25 Ranking Function, Probabilistic VS likelihood Models

Learning Outcome: Understand the basic underling foundation of information retrieval engine. Develop an idea how classical IR models are implemented for information retrieval.

Unit 2: Neural Information Retrieval **20 LHs**

Foundation: General architecture of Transformer, Large Language Models for IR, BERT as Encoder Only Model

Text Tokenization: BPE (Byte Pair Encoding), Word Piece, Unigrams

Text Representations for Ranking: BOW Encodings, Word Embedding (Word2Vec), Contextual Embedding (BERT Embedding), Sentence Embedding (Sentence BERT Embedding), Position Embedding, Attention, Re - ranking Retrieval, Embedding Models VS Re – ranker Models

Interaction Focused Systems: Pre – trained language models, Fine – tuning interaction focus systems, Prompt Optimization, Dealing with Long Texts

RAG (Retrieval Augmented Generation): Novelty of RAG, Design of RAG Systems, Indexing Pipeline (Data Loading, Data Chunking, Data Embedding, Vector Database), Generation Pipeline (Retrieval, Augmented, Generation), RAG Evaluations (Retrieval Metrics, RAG Specific Metrics), Hallucination as a Challenge

Learning Outcome: Describe the modern information retrieval model as large language models. How transformer based model can be used in text representation and ranking the documents? Learn the need of semantics over syntactic aspects. Why do we need RAG to improve the quality of Large Language Model (LLM) responses by grounding the model on external, real-time data sources?

Unit 3: Evaluation Information Retrieval **5 LHs**

Unranked Metrics (Precision, Recall), Average Precision, Discounted Cumulated Gain, Ranked Metrics (Mean Reciprocal Rank, Normalized Discounted Cumulative Gain), Faithfulness, Relevance, ANOVA (Analysis of Variance), A/B Testing, Interleaving

Learning Outcome: Explain the evaluation methods to express the accuracy.

Unit 4: Adaptations and Concerns **5 LHs**

CLIR (Cross Language Information Retrieval): Introduction, Some use cases, The Core Technology of CLIR

Bias in Retrieval Systems: Pre – existing bias, Stakeholder bias, Data bias, Algorithmic bias

Privacy in IR: Privacy for Searchers, Privacy for Search Engines, Privacy for Document Owners

Learning Outcome: Why do we need multilingual machine translation model? Describe the different privacy issues related to information retrieval

Unit 5: Web Search **6 LHs**

Search engines (working principle), Spidering (Structure of a spider, Simple spidering algorithm, multithreaded spidering, Bot), Directed spidering (Topic directed, Link directed), Crawlers (Basic crawler architecture), Link analysis (HITS, Page ranking), Query log analysis, Handling “invisible” Web – Snippet generation

Learning Outcome: Describe the working mechanism of search engine spider. Apply the Link analysis approach for ranking the pages. What are the advantages of having Query Logs?

Pedagogical Strategies

- Lectures with demonstration
- Hands-on lab sessions
- Problem-based learning
- Guest lectures from tech industry experts
- Continuous assessment and feedback
- Multimedia presentations to visualize concepts
- Mini project

Assessment Methods

a. Internal Assessment Methods and Types

Assessment Type	Weightage	Details
Class participation & attendance	10%	Contribution to discussions, engagement in class activities
Quizzes/short tests	15%	Periodic quizzes to assess comprehension
Practical/Project	20%	Lab sessions and mini project
Mid-term examination	25%	Written test
Pre-board examination	30%	Comprehensive written test covering all units

b. External Assessment Methods and Types

Dean's office will arrange the final/board examination in writing to test understanding, application, analysis, evaluation, and creation.

Mapping Course: Learning Outcomes and Program Learning

Course Learning Objective (CLO) Dimensions	Knowledge (K)	Skills (S)	Competence (C)
Total Learning	40%	40%	20%

Laboratory Works

The laboratory work includes writing programs for implementing the concepts of text normalization pipeline and construct an Inverted Index from raw text documents, to implement classic ranking and scoring algorithms to return documents sorted by relevance, to build a multithreaded web crawler and implementing semantic machine learning models and generative systems using a programming language like Python.

Suggested Readings

Alonso, O., & Baeza - Yates, R. (2025). *Information Retrieval Advanced Topics and Techniques.* Association For Computing Machinery.

Kimothi, A. (2025). *A Simple Guide To Retrieval Augmented Generation.* Manning .

W. White, R., & Shah, C. (2025). *Information Access in the Era of Generative AI.* Springer.

STT 551 Statistics & Data Management

MITM 2nd semester

Credits: 2

Lecture Hours: 32

Course Objectives

The objective of this course is to acquaint students with essential statistical techniques required for managerial decision-making and business problem-solving. It aims to develop their ability to analyze and interpret data in a meaningful manner to support organizational goals.

Course Description

This course delivers a rigorous exploration of statistical learning methods tailored for modern business applications. Students will master advanced inferential statistics, parametric and non-parametric tests, multiple correlation analysis, and multiple regression analysis as well as data management. The curriculum prioritizes high-volume corporate data interpretation utilizing scripting languages. Emphasizing model building and validation, students to solve real-world problems such as customer churn, market optimization, and financial risk mitigation.

Course Details

Unit 1: Theory of Probability and Probability Distributions 6 LHs

Theory of Probability: Concepts, Approaches to calculation of probability, Rules of addition & multiplication, marginal, joint conditional probability, and Bayes' theorem; Expected value and standard deviation of probability distribution.
Probability Distribution: Discrete distribution (Binomial distribution and Poisson distribution),
Continuous Distributions **(**Normal Distribution and its application).

Learning Outcomes: By the end of this chapter, students will be able **to** master probability concepts & approaches, apply probability rules, implement Bayes' theorem, evaluate expected value & standard deviation, analyze discrete probability distributions, and utilize continuous distributions.

Unit 2: Sampling and Estimation 6 LHs.

Sampling Methods: Objective of sampling, Random sampling (Simple random, Systematic, Stratified, and Cluster, Multi-stage sampling, Non-random sampling (Quota, Convenience, Judgment, Snow-ball), Central Limit Theorem and Standard Error.
Estimation: Point and Interval estimation, Confidence Intervals, Characteristics of a good estimator, Interval estimate for mean and proportion, Determination of sample size.

Learning Outcomes: By this chapter, students will be able to master how to choose samples using Simple, Systematic, Stratified, Cluster, and Multi-stage methods, identify and evaluate Quota, Convenience, Judgment, and Snowball methods., apply the Central Limit Theorem and calculate standard error for means and proportions., differentiate single-

value estimates from range-based confidence intervals, identify the four criteria: unbiasedness, consistency, efficiency, and sufficiency, build and interpret confidence ranges for population means and proportions, calculate the exact sample size needed based on desired precision and confidence levels.

Unit 3: Hypothesis Testing **10 LHs**

Fundamentals: Null and Alternative hypothesis, Type I and Type II Errors, Level of Significance.
Parametric Tests: Z-test (Single mean and double means, Single proportion and double proportion), t-test (One-sample, Independent, and Paired), and Analysis of Variance (One way ANOVA and two way ANOVA).
Non-Parametric Tests: Chi-square test (Test of independence and goodness of fit), Kruskal -Wallis test, Mann -Whitney Test.

Learning Outcomes: Apply Z-tests to compare single/double means and proportions for large samples, execute one-sample, independent, and paired t-tests for small sample sizes, perform one way and two-way ANOVA to evaluate means across multiple groups, conduct Chi-square tests for independence and goodness of fit, Implement the Kruskal-Wallis test for multi-group non-parametric analysis.

Unit 4: Correlation and Regression Analysis **6 LHs**

Correlation: Multiple and partial correlation, Coefficient of multiple and partial determination.
Regression: Multiple regression model, Residual analysis; Inferences regarding population regression coefficients, Introduction to multi-variable influence analysis.

Learning Outcomes: Differentiate between multiple correlation (overall linear relationship) and partial correlation (relationship between two variables while controlling for others), calculate and interpret the coefficient of multiple determination, and the coefficient of partial determination to assess how much variation model explains. Construct and formulate a multiple linear regression model using several independent variables to predict a single outcome, execute residual analysis to validate standard regression assumptions like linearity, homoscedasticity, and normality.

Unit 5: Foundational Data Management **4 LHs**

Data Life Cycle: Concept and stages of the data life cycle including collection, storage, analysis, archiving, and disposal.
Data Quality: Importance of data integrity and accuracy; techniques for managing data inconsistencies, missing values, and outliers.
Data Privacy and Ethics: Overview of data protection regulations such as GDPR and CCPA; ethical principles in data usage and transparency.
Data Security: Fundamentals of data protection; access control, encryption, and strategies for safeguarding organizational data from unauthorized access.

Learning Outcomes: Students will learn to manage information assets effectively from creation to deletion, develop the technical skills needed to maintain reliable, clean, and accurate datasets; evaluate legal compliance and apply moral standards to data handling, design defense strategies to protect organizational data from breaches.

Suggested Readings:

Aczel, A. D., Sounderpandian, J., Saavanan, P. & Joshi, R. (2012). Complete Business Statistics, McGraw Hill Education (India) Pvt. Ltd.: New Delhi.

Anderson, D. R., Sweeney, D. J., Williams, T. A., Camm, J. D. & Cochran, J. J. (2014). Statistics for Business & Economics, Cengage Learning Custom Publishing.

Black, K. (2012). Applied Business Statistics. Wiley: New Jersey.

Coleman, L. S.,(2018). Navigating the Labyrinth: An Executive Guide to Data Management, Technics Publications LLC: USA.

Levin, R.I. & Rubin, D.S. (1997). Statistics for Management, Prentice-Hall of India.

Vohra, N. D. (2012). Business Statistics. McGraw Hill Education (India) Pvt. Ltd.: New Delhi.