MK UNIVERSITY

PATAN, GUJARAT

ESTABLISHED BY THE GUJARAT GOVT. **RECOGNIZED BY UGC UNDER SECTION 2(f) OF UGC ACT,1956**

MK University, Patan
Faculty of Engineering Technology,
Department of Data science Engineering

M. TECH (DATA SCIENCE ENGINEERING) SEM-I									
SR NO.	COURSE TYPE	COURSE CODE	COURSE NAME	LECTURE (HRS.)/WEEK	PRACTICAL (HRS.)/WEEK	CREDITS	EXAMINATION		TOTAL MARKS
							INTERNAL	EXTERNAL	
1	MAJOR	MTDSE101	ADVANCED MATHEMATICS FOR ENGINEERS	4	0	4	40	60	100
2	MAJOR	MTDSE102	DATA MANAGEMENT & DATA BASE SYSTEMS	4	2	6	90	60	150
3	MAJOR	MTDSE103	BUSINESS FUNDAMENTALS FOR ANALYTICS	4	2	6	90	60	150
4	MINOR	MTDSE104	RESEARCH METHDOLOGY & TECHNICAL COMMUNICATION	4	0	4	40	60	100
5	SEC	MTDSE105	ENTERPRENURSHIP DEVELOPMENT	4	0	4	40	60	100
TOTAL				20	4	24	300	300	600

M. TECH (DATA SCIENCE ENGINEERING) SEM-II									
SR NO.	COURSE TYPE	COURSE CODE	COURSE NAME	LECTURE (HRS.)/WEEK	PRACTICAL (HRS.)/WEEK	CREDITS	EXAMINATION		TOTAL MARKS
							INTERNAL	EXTERNAL	
1	MAJOR	MTDSE201	MACHINE LEARNING FUNDAMENTALS	4	0	4	40	60	100
2	MAJOR	MTDSE202	BIG DATA TECHNOLOGIES & PROCESSING	4	2	6	90	60	150
3	MAJOR	MTDSE203	ADVANCED VISUALIZATION	4	2	6	90	60	150
4	MINOR	MTDSE204	TIME SERIES ANALYSIS	4	2	6	90	60	150
5	VAC	MTDSE205	BUSINESS COMMUNICATION-I	2	0	2	0	50	50
TOTAL				18	6	24	310	290	600

MK UNIVERSITY

PATAN, GUJARAT

ESTABLISHED BY THE GUJARAT GOVT. **RECOGNIZED BY UGC UNDER SECTION 2(f) OF UGC ACT,1956**

M. TECH (DATA SCIENCE ENGINEERING) SEM-III									
SR NO.	COURSE TYPE	COURSE CODE	COURSE NAME	LECTURE (HRS.)/ WEEK	PRACTICAL (HRS.)/ WEEK	CREDITS	EXAMINATION		TOTAL MARKS
							INTERNAL	EXTERNAL	
1	MAJOR	MTDSE301	NATURAL LANGUAGE PROCESSING	4	2	6	90	60	150
2	MAJOR	MTDSE302	CLOUD COMPUTING FOR DATA SCIENCE	4	2	6	90	60	150
3	MINOR	MTDSE303	MOOC/SWAYAM COURSE	3	0	3	100	00	100
4	VAC	MTDSE304	DISSERTATION PHASE-I	0	8	8	100	100	200
TOTAL				11	12	23	380	220	600

M. TECH (DATA SVIENCE ENGINEERING) SEM-IV									
SR NO.	COURSE TYPE	COURSE CODE	COURSE NAME	LECTURE (HRS.)/ WEEK	PRACTICAL (HRS.)/ WEEK	CREDITS	EXAMINATION		TOTAL MARKS
							INTERNAL	EXTERNAL	
1	MAJOR	MTDSE401	INDUSTRY SEMINARS/WORKSHOPS/INTERNSHIP	0	2	2	50	00	50
2	MINOR	MTDSE402	COMPREHENSIVE VIVA VOCE	0	2	2	50	00	50
3	MAJOR	MTDSE403	DISEERTATION PHASE-II	0	16	16	200	200	400
4	VAC	MTDSE404	BUSINESS COMMUNICATION-II	2	0	2	00	50	50
TOTAL				2	20	22	300	250	550

MK UNIVERSITY

PATAN, GUJARAT

ESTABLISHED BY THE GUJARAT GOVT. RECOGNIZED BY UGC UNDER SECTION 2(f) OF UGC ACT,1956

SEMESTER-I

SUBJECT CODE: MTDSE101

SUBJECT NAME: ADVANCED MATHEMATICS FOR ENGINEERS

Course Objectives:

- To provide a rigorous mathematical foundation for advanced engineering modeling and analysis.
- To bridge theoretical mathematics with practical engineering applications.
- To develop problem-solving skills using analytical and computational tools.
- To prepare students for research and development in engineering domains requiring mathematical sophistication.

Course Outcomes: At the end of the course students shall be able to

CO1	Formulate and solve engineering problems using advanced techniques in linear algebra and tensor analysis.
CO2	Apply partial differential equations (PDEs) and transform methods to model dynamical systems and boundary value problems.
CO3	Use variational calculus and optimization methods for engineering design and control problems.
CO4	Analyze stochastic systems and uncertainty propagation using probability theory and statistical methods.

Unit	Content	Credit	Weightage
I	**Advanced Linear Algebra & Tensors for Engineers** ○ Review of vector spaces, eigenvalues, SVD, Jordan form ○ Matrix decompositions (LU, QR, Cholesky, Schur) ○ Tensor algebra: notation, operations, invariants ○ Tensor applications: stress-strain, inertia, constitutive models ○ Numerical linear algebra (conditioning, iterative solvers) • **Applications:** Structural analysis, continuum mechanics, control systems, data compression.	1	25%
II	**Partial Differential Equations & Transform Methods** ○ Classification of PDEs (elliptic, parabolic, hyperbolic) ○ Separation of variables, eigenfunction expansions ○ Green's functions for ODEs and PDEs ○ Integral transforms (Fourier, Laplace, Hankel) for PDEs ○ Introduction to finite element and finite volume concepts	1	25%

	• **Applications:** Heat transfer, wave propagation, fluid dynamics, signal processing.		
III	**Calculus of Variations & Optimization** ○ Functional derivatives, Euler–Lagrange equation ○ Constraints (Lagrange multipliers, isoperimetric problems) ○ Direct methods (Ritz, Galerkin) ○ Optimal control theory (Pontryagin's principle, Hamiltonian formulation) ○ Convex optimization basics (gradient descent, KKT conditions) • **Applications:** Optimal design, trajectory optimization, energy minimization, control systems.	1	25%
IV	**Stochastic Processes & Uncertainty Quantification** ○ Probability spaces, random variables, distributions ○ Stochastic processes (Brownian motion, Poisson process, Markov chains) ○ Itô calculus basics (stochastic differential equations) ○ Uncertainty quantification (Monte Carlo, polynomial chaos, sensitivity analysis) ○ Statistical estimation and regression for engineering data • **Applications:** Risk analysis, reliability engineering, random vibrations, financial engineering, signal noise modelling	1	25%

TEXT BOOKS:

- Kreyszig, E. – *Advanced Engineering Mathematics* (10th ed.) – Wiley.
- Strang, G. – *Linear Algebra and Its Applications* (5th ed.) – Cengage.
- Arfken, G.B., Weber, H.J., Harris, F.E. – *Mathematical Methods for Physicists* (7th ed.) – Academic Press.
- J.N. Reddy – *Applied Functional Analysis and Variational Methods in Engineering* – McGraw-Hill.
- Papoulis, A., & Pillai, S.U. – *Probability, Random Variables and Stochastic Processes* (4th ed.) – McGraw-Hill.

REFERENCE BOOKS:

- Riley, K.F., Hobson, M.P., Bence, S.J. – *Mathematical Methods for Physics and Engineering* (3rd ed.) – Cambridge.
- Gelfand, I.M., & Fomin, S.V. – *Calculus of Variations* – Dover.
- Oksendal, B. – *Stochastic Differential Equations: An Introduction with Applications* (6th ed.) – Springer.
- Holmes, M.H. – *Introduction to Numerical Methods in Differential Equations* – Springer.
- Gould, P. – *Introduction to Linear Elasticity* (for tensor applications) – Springer.

ONLINE RESOURCES:

- Coursera:
  - *Mathematics for Engineers Specialization* (The Hong Kong University of Science and Technology)
  - *Data Science Math Skills* (Duke University)

SUBJECT CODE: MTDSE102

SUBJECT NAME: DATA MANAGEMENT AND DATA BASE SYSTEMS

Course Objectives:

- To understand foundational and advanced concepts of database systems, including data models, query languages, and storage structures.
- To design and normalize relational databases using ER modeling and normalization techniques.
- To explore modern database systems such as NoSQL, NewSQL, and distributed databases.
- To implement database systems in practice, including indexing, transaction management, and query optimization.
- To develop skills in using cloud-based and big data platforms for scalable data management.

Course Outcomes: At the end of the course students shall be able to

CO1	Design, implement, and query relational databases using SQL and procedural extensions.
CO2	Apply normalization and indexing techniques for efficient database design.
CO3	Compare and contrast relational and non-relational databases (NoSQL) for different use cases.
CO4	Develop applications using modern database systems, including cloud and distributed databases.

Unit	Content	Credit	Weightage
I	**Foundations of Database Systems** • Introduction to databases and DBMS architecture • Data models: relational, network, hierarchical • Entity-Relationship (ER) modeling and Enhanced ER • Relational algebra and calculus • SQL: DDL, DML, DQL, joins, subqueries, aggregation • **Practical:** SQL exercises on creating tables, queries, and joins (using PostgreSQL/MySQL)	1	25%
II	**Database Design & Advanced SQL** • Functional dependencies and normalization (1NF to BCNF, 4NF) • Indexing structures: B-trees, B+ trees, hash indexing • Query processing and optimization • Views, triggers, stored procedures, cursors • Introduction to PL/SQL or T-SQL • **Practical:** Normalization exercises, writing	1	25%

	advanced SQL queries, creating stored procedures		
III	**Transaction Management & NoSQL Databases** • Transactions: ACID properties, states, schedules • Concurrency control: locks, two-phase locking, timestamps • Recovery: log-based recovery, checkpoints • Introduction to NoSQL databases: types (document, key-value, column-family, graph) • MongoDB basics: CRUD operations, indexing, aggregation framework • **Practical:** Implementing transactions in SQL; CRUD operations in MongoDB	1	25%
IV	**Distributed, Cloud & Modern Data Platforms** • Distributed databases: concepts, fragmentation, replication • NewSQL databases: Google Spanner, CockroachDB • Cloud databases: AWS RDS, DynamoDB, Azure Cosmos DB • Big Data integration: Hadoop HDFS, Hive, Spark SQL • Data warehousing concepts: OLAP, star schema, snowflake schema • **Practical:** Hands-on with cloud database services (AWS/Azure free tier), basics of Hive/Spark SQL	1	25%

TEXT BOOKS:

- *Database System Concepts*, 7th Ed. — Abraham Silberschatz, Henry F. Korth, S. Sudarshan
- *Fundamentals of Database Systems*, 7th Ed. — Ramez Elmasri, Shamkant B. Navathe
- *NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence* — Pramod J. Sadalage, Martin Fowler

REFERENCE BOOKS:

- *Database Management Systems*, 3rd Ed. — Raghu Ramakrishnan, Johannes Gehrke
- *Designing Data-Intensive Applications* — Martin Kleppmann
- *MongoDB: The Definitive Guide*, 3rd Ed. — Shannon Bradshaw, Kristina Chodorow
- *Cassandra: The Definitive Guide*, 2nd Ed. — Jeff Carpenter, Eben Hewitt

ONLINE RESOURCES:

- SQL Practice: SQLZoo, LeetCode, HackerRank, Mode Analytics
- NoSQL Hands-on: MongoDB University, Cassandra Tutorials (DataStax Academy)
- Cloud Platforms: AWS RDS/Aurora, Google Cloud SQL & Firestore, Azure Cosmos DB
- Big Data Platforms: Apache Hadoop, Spark SQL, Databricks Community Edition

PRACTICAL LIST:

- Design an ER model for a given scenario and convert it to relational schema.
- Implement the schema in RDBMS, perform SQL queries for business insights.
- Normalize a given dataset up to BCNF.

- Create stored procedures and triggers for automating database tasks.
- Perform CRUD operations and aggregation pipelines in MongoDB.
- Set up a cloud database instance and connect it with an application.
- Compare performance of simple queries in SQL vs. NoSQL environments.
- Explore a distributed SQL database (e.g., CockroachDB or YugabyteDB).

SUBJECT CODE: MTDSE103

SUBJECT NAME: BUSINESS FUNDAMENTALS FOR ANALYTICS

Course Objectives:

- To bridge the gap between technical data science skills and business decision-making.
- To understand key business domains and processes where analytics creates value.
- To develop the ability to translate business problems into analytical frameworks.
- To learn the principles of data-driven strategy, measurement, and return on investment (ROI).
- To build skills in communicating analytical insights effectively to business stakeholders.

Course Outcomes: At the end of the course students shall be able to

CO1	Identify and articulate business problems that can be solved with analytics across domains like marketing, finance, and operations.
CO2	Apply structured frameworks (e.g., CRISP-DM, OSEMN) to scope analytics projects and define success metrics.
CO3	Understand financial concepts (e.g., NPV, ROI, CAC, LTV) to evaluate the business impact of analytics initiatives.
CO4	Create effective data visualizations and dashboards for business storytelling.

Unit	Content	Credit	Weightage
I	**Introduction to Business Analytics & Value Creation** • The analytics landscape: descriptive, diagnostic, predictive, prescriptive • How analytics creates competitive advantage (Davenport's model) • Key business domains for analytics: Marketing, Finance, Operations, HR • From business problem to analytical solution: The CRISP-DM framework • **Practical:** Analyze a classic business case (e.g., Netflix recommendation engine, Amazon supply chain) to identify the analytical value proposition.	1	25%
II	**Business Metrics, KPIs & Financial Fundamentals** • Key Performance Indicators (KPIs) vs. Metrics • Business measurement frameworks: AARRR (Pirate Metrics), HEART • Financial concepts for analysts: NPV, ROI, IRR, Break-even Analysis • Customer-centric metrics: Customer Acquisition	1	25%

	Cost (CAC), Lifetime Value (LTV), Churn • **Practical:** Build a simple business case with financial justification for an analytics project (e.g., implementing a new CRM system).		
III	**Data-Driven Decision Making & Strategy** • Hypothesis-driven problem solving in business • A/B testing and experimental design for business decisions • Building a data-driven culture: challenges and best practices • Ethics, privacy, and regulatory considerations (GDPR, CCPA) • **Practical:** Design an A/B test for a business scenario (e.g., website redesign, email campaign) with clear hypothesis and success metrics.	1	25%
IV	**Communication, Visualization & Storytelling** • The art of communicating data: Knowing your audience • Principles of effective data visualization (Tufte, Few) • Dashboard design for business users (using Tableau/Power BI) • Crafting the data story: narrative structure and executive summaries • **Practical:** Create a dashboard and a one-page executive summary from a given dataset (e.g., sales performance) to recommend a business action.	1	25%

TEXT BOOKS:

- *Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking* — Foster Provost & Tom Fawcett
- *Competing on Analytics: The New Science of Winning* — Thomas H. Davenport & Jeanne G. Harris

REFERENCE BOOKS:

- *Behind Every Good Decision: How Anyone Can Use Business Analytics to Turn Data into Profitable Insight* — Piyanka Jain & Puneet Sharma
- *The Visual Display of Quantitative Information* — Edward R. Tufte
- *Storytelling with Data: A Data Visualization Guide for Business Professionals* — Cole Nussbaumer Knaflic
- *HBR Guide to Data Analytics Basics for Managers* — Harvard Business Review

ONLINE RESOURCES:

- Case Studies: Harvard Business Publishing Education, Kellogg Case Insights
- Business Simulations: Capsim, Marketplace Live
- Data Visualization: Tableau Public, Power BI, Google Data Studio
- Financial Basics: Khan Academy (Finance & Capital Markets), Corporate Finance Institute
- Analytics Frameworks: CRISP-DM, Google's HEART Framework

PRACTICAL LIST:

- Business Problem Scoping: Write a 2-page project charter for an analytics initiative using the CRISP-DM framework.
- Financial Modeling Lab: Calculate ROI, NPV, and payback period for a proposed data infrastructure investment.
- A/B Test Design: For an e-commerce company, design a complete A/B test to improve checkout conversion, including statistical power considerations.
- Dashboard & Storytelling Project:
    1. Use a dataset (e.g., Superstore sales).
    2. Build an interactive dashboard in Tableau Public/Power BI.
    3. Record a 5-minute video presentation "pitching" your insights to a simulated management team.

SUBJECT CODE: MTDSE104

SUBJECT NAME: RESEARCH METHDOLOGY AND TECHNICAL COMMUNICATION

Course Objectives:

- To equip engineering graduates with a structured approach to scientific inquiry and problem-solving.
- To develop proficiency in selecting and applying appropriate research methods for engineering investigations.
- To enhance technical communication skills for academia and industry.
- To foster an understanding of research ethics, scholarly publishing, and lifelong learning in research.

Course Outcomes: At the end of the course students shall be able to

CO1	Formulate a research problem, conduct systematic literature reviews, and develop a viable research proposal.
CO2	Design and execute appropriate research methodologies (experimental, numerical, analytical) with consideration for ethics and data integrity.
CO3	Apply statistical tools and software for data analysis, interpretation, and validation of research findings.
CO4	Produce high-quality technical documents (research papers, proposals, theses) and deliver effective technical presentations.

Unit	Content	Credit	Weightage
I	**Foundations of Engineering Research & Problem Formulation** • **Topics:** ○ Philosophy of research: inductive vs. deductive reasoning, scientific method in engineering. ○ Types of engineering research: fundamental, applied, experimental, computational, empirical. ○ Problem identification and formulation: research gap analysis. ○ Literature review strategies: databases (Scopus, Web of Science, IEEE Xplore),	1	25%

	citation management tools (Zotero, Mendeley), critical analysis of literature. ○ Developing a research proposal: objectives, scope, significance, and work plan. • **Applications:** Thesis topic selection, grant proposal writing, project planning.		
II	**Research Design, Methods & Ethics** • **Topics:** ○ Research design: experimental, quasi-experimental, case study, modeling & simulation. ○ Data collection methods: sensors, surveys, instrumentation, simulation outputs. ○ Design of Experiments (DoE): factorial design, Taguchi methods, response surface methodology. ○ Research ethics: plagiarism, fabrication/falsification, authorship, informed consent. ○ Ethical approval process and responsible conduct of research (RCR). • **Applications:** Planning a lab/field experiment, setting up a CFD/FEA study, survey design.	1	25%
III	**Data Analysis, Statistics & Software Tools** • **Topics:** ○ Data preprocessing: outlier detection, missing data, normalization. ○ Descriptive and inferential statistics: hypothesis testing (t-test, ANOVA), confidence intervals. ○ Regression analysis: linear, multiple, logistic. ○ Introduction to multivariate analysis and machine learning for engineering data. ○ Software tools: MATLAB/Python (NumPy, SciPy, pandas), R, MiniTab. ○ Data visualization principles: effective graphs, charts, and plots. • **Applications:** Analyzing experimental results, validating computational models, interpreting sensor data.	1	25%
IV	**Technical Communication & Research Dissemination** • **Topics:** ○ Structure of technical documents: research papers, theses, technical reports. ○ Writing strategies: clarity, conciseness, coherence, and argument development.	1	25%

	○ Graphical abstracts, data presentation, and table/figure design. ○ Oral presentations: conference talks, thesis defense, poster design. ○ Publication process: journal selection, peer review, responding to reviewers. ○ Intellectual Property Rights (IPR): patents, copyrights, licensing. ○ Research dissemination: repositories, academic social networks (ResearchGate, LinkedIn), and impact metrics (h-index, citations). • **Applications:** Paper writing, thesis compilation, conference presentation, patent filing.		

TEXT BOOKS:

- Kothari, C.R. – *Research Methodology: Methods and Techniques* (4th ed.) – New Age International.
- Day, R.A., and Gastel, B. – *How to Write and Publish a Scientific Paper* (9th ed.) – Greenwood.
- Montgomery, D.C. – *Design and Analysis of Experiments* (10th ed.) – Wiley.
- Alley, M. – *The Craft of Scientific Writing* (4th ed.) – Springer.

REFERENCE BOOKS:

- Bordens, K.S., and Abbott, B.B. – *Research Design and Methods: A Process Approach* (11th ed.) – McGraw-Hill.
- Wallwork, A. – *English for Writing Research Papers* (2nd ed.) – Springer.
- Box, G.E.P., Hunter, J.S., and Hunter, W.G. – *Statistics for Experimenters* (2nd ed.) – Wiley.
- IEEE Author Center Guides – *IEEE Publication Services and Products Board*.
- Laplante, P.A. – *Technical Writing: A Practical Guide for Engineers and Scientists* – CRC Press.

ONLINE RESOURCES:

- edX Courses:
  1. *"Principles of Statistical Analysis"* (Microsoft)
  2. *"How to Write and Publish a Scientific Paper"* (KU Leuven)

SUBJECT CODE: MTDSE105

SUBJECT NAME: ENTERPRENURSHIP DEVELOPMENT

Course Objectives:

- To cultivate an entrepreneurial mindset among engineering graduates.
- To provide practical tools for transforming technical ideas into viable business ventures.
- To develop skills in business modeling, financial planning, and venture funding.
- To prepare students for startup creation, intrapreneurship, or technology commercialization roles.

Course Outcomes: At the end of the course students shall be able to

CO1	Identify and evaluate entrepreneurial opportunities emerging from technological trends and market gaps.

CO2	Develop a comprehensive business model and validate it using lean startup methodologies and customer discovery.
CO3	Create financial projections, evaluate startup costs, and understand funding mechanisms for technology ventures.
CO4	Formulate a complete investor-ready business plan and deliver an effective pitch to potential stakeholders.

Unit	Content	Credit	Weightage
I	**Entrepreneurial Mindset & Opportunity Identification** • **Topics:** ○ The Entrepreneurial Engineer: Mindset, traits, and role in economic development. ○ Sources of Innovation: Technology push vs. market pull, disruptive vs. sustaining innovation. ○ Opportunity Recognition: Identifying problems worth solving, trend analysis (STEEP), blue ocean strategy. ○ Idea Validation: Lean canvas, hypothesis testing, conducting problem-solution interviews. ○ Intellectual Property Strategy for Startups: Patents, trademarks, trade secrets, and licensing basics. • **Applications:** Spotting opportunities in cleantech, Industry 4.0, medtech, and digital transformation.	1	25%
II	**Business Model Design & Customer Development** • **Topics:** ○ Business Model Innovation: Business Model Canvas (Osterwalder), Value Proposition Canvas. ○ Customer Discovery & Validation: The "Get Out of the Building" approach, creating MVP (Minimum Viable Product). ○ Market Analysis: TAM, SAM, SOM, competitive analysis, positioning. ○ Pricing Strategies for Tech Products: Cost-plus, value-based, subscription, freemium models. ○ Go-to-Market Strategy: Sales channels, partnerships, digital marketing fundamentals. • **Applications:** Designing scalable models for SaaS, hardware-as-a-service, platform businesses.	1	25%

III	**Startup Finance, Funding & Legal Foundations** • **Topics:** ○ Startup Financials: Building financial models, unit economics, burn rate, runway. ○ Pro Forma Statements: Income statement, cash flow, balance sheet projections. ○ Funding Sources: Bootstrapping, angels, venture capital, crowdfunding, government grants (SBIR, DST). ○ Valuation Methods for Early-Stage Startups: Scorecard, Berkus, risk factor summation. ○ Legal Structures & Compliance: Company registration (LLP, Pvt Ltd), shareholder agreements, ESOPs, compliance essentials. ○ Term Sheet Fundamentals: Key clauses, negotiation basics. • **Applications:** Preparing for seed funding, managing cash flow, cap table management.	1	25%
IV	**Business Planning, Pitching & Scaling Ventures** • **Topics:** ○ The Business Plan: Executive summary, company description, product/service, market analysis, marketing plan, management team, financial projections. ○ The Art of Pitching: Investor pitch deck structure, storytelling, demo preparation. ○ Building the Team: Co-founder selection, hiring early employees, advisory boards. ○ Operational Planning: Supply chain, quality, scaling production. ○ Growth Strategies: Scaling challenges, pivoting, exit strategies (acquisition, IPO). ○ Social Entrepreneurship & Ethics: Creating social impact, ethical leadership. • **Applications:** Crafting investor pitches, developing operational roadmaps, planning for scale.	1	25%

TEXT BOOKS:

- Osterwalder, A., and Pigneur, Y. – *Business Model Generation* – Wiley.
- Ries, E. – *The Lean Startup* – Penguin.
- Blank, S., and Dorf, B. – *The Startup Owner's Manual* – K & S Ranch.
- Barrow, C., Barrow, P., and Brown, R. – *The Business Plan Workbook* (10th ed.) – Kogan Page.

REFERENCE BOOKS:

- Aulet, B. – *Disciplined Entrepreneurship* – Wiley.

- Mullins, J. – *The New Business Road Test* (5th ed.) – FT Publishing.
- Kawasaki, G. – *The Art of the Start 2.0* – Portfolio Penguin.
- Thiel, P. – *Zero to One* – Crown Business.
- Maurya, A. – *Running Lean* (2nd ed.) – O'Reilly.

ONLINE RESOURCES:

- Coursera: *"Entrepreneurship Specialization"* (Wharton), *"Startup Entrepreneurship"* (Technion).
- edX: *"Entrepreneurship in Emerging Economies"* (HarvardX), *"Innovation and Entrepreneurship"* (DelftX).

Udemy: Courses on Business Plan Writing, Startup Funding, and Digital Marketing.

MK UNIVERSITY

PATAN, GUJARAT

ESTABLISHED BY THE GUJARAT GOVT. RECOGNIZED BY UGC UNDER SECTION 2(f) OF UGC ACT,1956

SEMESTER-II

SUBJECT CODE: MTDSE201

SUBJECT NAME: MACHINE LEARNING FUNDAMENTALS

Course Objectives:

- To provide a rigorous mathematical foundation for understanding ML algorithms
- To implement and evaluate supervised and unsupervised learning techniques
- To develop skills in model selection, validation, and performance optimization
- To introduce modern ML paradigms including ensemble methods and neural networks
- To cultivate best practices in ML engineering and deployment considerations

Course Outcomes: At the end of the course students shall be able to

CO1	**Mathematically Derive** and implement fundamental ML algorithms from scratch
CO2	**Select Appropriate Models** for different problem types (regression, classification, clustering)
CO3	**Comprehensively Evaluate** model performance using appropriate metrics and validation techniques
CO4	**Optimize Models** through feature engineering, hyperparameter tuning, and regularization

Unit	Content	Credit	Weightage
I	**Foundations & Mathematical Preliminaries** **Topics:** • Introduction to ML: Types (Supervised, Unsupervised, Reinforcement) • Probability & Statistics for ML: Distributions, Bayes' Theorem, MLE, MAP • Linear Algebra: Vectors, Matrices, Eigenvalues, SVD • Calculus: Gradients, Optimization (Gradient Descent variants) • Bias-Variance Tradeoff, No Free Lunch Theorem • **Tools:** Python scientific stack (NumPy, SciPy, Pandas), Mathematical derivations	1	25%
II	**Supervised Learning** **Topics:** • **Linear Models:** Simple/Multiple Linear Regression, Logistic Regression • **Regularization:** Ridge (L2), Lasso (L1), Elastic Net • **Non-linear Models:** Polynomial Regression, Splines • **Tree-based Models:** Decision Trees, Random Forests, Gradient Boosting (XGBoost, LightGBM)	1	25%

	• **Support Vector Machines:** Hard/Soft Margin, Kernels (Linear, RBF, Polynomial) • **Evaluation Metrics:** MSE/MAE, Accuracy/Precision/Recall/F1, ROC-AUC, Confusion Matrix • **Practicals:** Implementing from scratch (linear models, trees) + scikit-learn		
III	**Unsupervised Learning & Model Optimization** **Topics:** • **Clustering:** K-Means, Hierarchical, DBSCAN, Gaussian Mixture Models • **Dimensionality Reduction:** PCA, t-SNE, UMAP, Autoencoders introduction • **Model Selection:** Train/Test Split, Cross-Validation (k-fold, stratified) • **Hyperparameter Tuning:** Grid Search, Random Search, Bayesian Optimization • **Feature Engineering:** Encoding, Scaling, Transformation, Selection methods • **Ensemble Methods:** Bagging, Boosting, Stacking, Voting • **Practicals:** Clustering customer segmentation, Dimensionality reduction visualization	1	25%
IV	**Introduction to Neural Networks & ML Engineering** **Topics:** • **Neural Network Basics:** Perceptron, Multi-layer Perceptron, Activation Functions • **Training NNs:** Backpropagation, Optimization Algorithms (Adam, RMSprop) • **Deep Learning Introduction:** CNNs for images, RNNs for sequences (conceptual) • **ML Pipelines:** Scikit-learn Pipelines, MLflow for experiment tracking • **Model Deployment Basics:** Serialization (pickle, joblib), ONNX format • **Ethics in ML:** Fairness, Bias Detection, Model Cards • **Practicals:** Building ML pipeline with preprocessing, training, evaluation, and logging	1	25%

TEXT BOOKS:

- Primary: *Pattern Recognition and Machine Learning* — Christopher M. Bishop
- Primary: *Hands-On Machine Learning with Scikit-Learn, Keras & TensorFlow* — Aurélien Géron (3rd Edition)

REFERENCE BOOKS:

- *The Elements of Statistical Learning* — Trevor Hastie, Robert Tibshirani, Jerome Friedman
- *Machine Learning: A Probabilistic Perspective* — Kevin P. Murphy

- *Deep Learning* — Ian Goodfellow, Yoshua Bengio, Aaron Courville
- *Python Machine Learning* — Sebastian Raschka & Vahid Mirjalili (3rd Edition)

ONLINE RESOURCES:

- Kaggle: Competitions, Datasets, Notebooks (Kaggle Learn)
- Coursera: Machine Learning Specialization (Andrew Ng)

SUBJECT CODE: MTDSE202

SUBJECT NAME: BIG DATA TECHNOLOGIES AND PROCESSING

Course Objectives:

- To provide a comprehensive understanding of big data technologies, distributed computing frameworks, and cloud-based AI/ML services.
- To develop expertise in processing, analyzing, and deriving insights from massive-scale datasets using scalable cloud platforms.
- To implement end-to-end machine learning pipelines on cloud infrastructure for production-level AI systems.
- To prepare students for roles in data engineering, cloud AI solution architecture, and large-scale AI system deployment.

Course Outcomes: At the end of the course students shall be able to

CO1	Design and implement scalable data processing pipelines using big data technologies.
CO2	Develop and deploy machine learning models on cloud platforms using managed AI services.
CO3	Optimize distributed ML training and inference for large-scale datasets.
CO4	Architect and implement production-grade cloud-native AI systems with MLOps principles.

Unit	Content	Credit	Weightage
I	**Big Data Foundations and Distributed Processing** **Topics:** • Introduction to Big Data: 5 Vs (Volume, Velocity, Variety, Veracity, Value) • Hadoop ecosystem: HDFS, MapReduce, YARN • Apache Spark: RDDs, DataFrames, Spark SQL, Spark MLlib • Distributed data processing patterns and optimizations • NoSQL databases: MongoDB, Cassandra, Redis • Data lakes vs. data warehouses • Real-time stream processing: Apache Kafka, Apache Flink • **Applications:** Log analysis, clickstream processing, IoT data aggregation	1	25%
II	**Cloud Computing for AI** • Cloud computing fundamentals: IaaS, PaaS, SaaS,	1	25%

	FaaS • Major cloud platforms: AWS, Azure, Google Cloud Platform • Cloud storage solutions: S3, Azure Blob, Google Cloud Storage • Containerization and orchestration: Docker, Kubernetes for AI workloads • Serverless computing for AI: AWS Lambda, Azure Functions • Cloud-based ML platforms: SageMaker, Azure ML, Vertex AI • Cost optimization and resource management in cloud AI • **Applications:** Scalable model training, automated ML pipelines, batch inference		
III	**Distributed Machine Learning and MLOps** **Topics:** • Distributed training frameworks: Horovod, PyTorch DDP, TensorFlow Distributed • Model parallelism and data parallelism • Hyperparameter tuning at scale: Ray Tune, Optuna on cloud • Feature stores: Feast, Tecton • ML pipeline orchestration: Apache Airflow, Kubeflow Pipelines • Model versioning and registry: MLflow, DVC • Monitoring and logging for ML systems: Prometheus, Grafana • CI/CD for machine learning (MLOps) • **Applications:** Large language model training, recommendation systems at scale	1	25%
IV	**Advanced Analytics and Cloud AI Services** **Topics:** • Big data analytics tools: Apache Hive, Presto, Apache Druid • Data visualization at scale: Tableau, Power BI, Apache Superset • Managed AI services: Computer Vision APIs, NLP services, speech recognition • Vector databases and similarity search: Pinecone, Weaviate, Milvus • Graph analytics and processing: Neo4j, Amazon Neptune • Edge-cloud AI integration • Security, compliance, and governance in cloud AI • Ethical considerations: Bias in large-scale AI, data privacy	1	25%

	• **Applications:** Real-time recommendation engines, fraud detection, predictive maintenance		

TEXT BOOKS:

- Chambers, B., & Zaharia, M. – *Spark: The Definitive Guide* – O'Reilly.
- Lakshmanan, V., Robinson, S., & Munn, M. – *Machine Learning Design Patterns* – O'Reilly.
- Kimball, R., & Ross, M. – *The Data Warehouse Toolkit* (3rd ed.) – Wiley.
- Huyen, C. – *Designing Machine Learning Systems* – O'Reilly.

REFERENCE BOOKS:

- White, T. – *Hadoop: The Definitive Guide* (4th ed.) – O'Reilly.
- Guller, M. – *Big Data Analytics with Spark* – Apress.
- Kleppmann, M. – *Designing Data-Intensive Applications* – O'Reilly.
- AWS/Azure/GCP – *Official certification guides and whitepapers*.

ONLINE RESOURCES:

- Databricks Community Edition: Free Spark cluster
- Google Colab Pro: GPU access for distributed ML experiments
- AWS Educate/Azure for Students: Free cloud credits
- Coursera: *Big Data Specialization* (University of California San Diego)
- edX: *Data Science and Machine Learning in the Cloud* (Microsoft)

PRACTICAL LIST:

1. **Big Data Processing with Apache Spark**

Task: Process a multi-terabyte dataset (e.g., Wikipedia dump or Twitter stream) using PySpark on Databricks. Perform data cleaning, aggregation, and analysis. Compare performance between RDD and Data Frame APIs.

2**. End-to-End Cloud ML Pipeline**

Task: Build a complete ML pipeline on AWS SageMaker or Azure ML including data ingestion, feature engineering, model training (XGBoost/Deep Learning), hyperparameter tuning, and deployment. Implement automated retraining with Airflow.

3, **Distributed Deep Learning Training**

Task: Train a large vision or language model (ResNet50 or BERT) using distributed training across multiple GPUs/instances on cloud. Implement with Horovod or PyTorch DDP. Compare training time and cost vs. single instance.

4. **Real-time Analytics and AI Serving**

Task: Create a real-time fraud detection system using Kafka for streaming, Spark Streaming/Flink for processing, and a pre-trained model served via TensorFlow Serving on Kubernetes. Build a monitoring dashboard with Grafana.

MK UNIVERSITY

PATAN, GUJARAT

ESTABLISHED BY THE GUJARAT GOVT. **RECOGNIZED BY UGC UNDER SECTION 2(f) OF UGC ACT,1956**

SUBJECT CODE: MTDSE203

SUBJECT NAME: ADVANCED VISUALIZATION

Course Objectives:

- To move beyond basic charts to advanced visualization techniques for complex data
- To understand perceptual principles and cognitive science behind effective visual encoding
- To master interactive, geospatial, temporal, and high-dimensional visualization methods
- To develop skills in creating production-grade dashboards and visual analytics systems
- To explore cutting-edge visualization techniques including network graphs, 3D, and AR/VR applications

Course Outcomes: At the end of the course students shall be able to

CO1	**Design & Evaluate** visualizations using perceptual and cognitive principles
CO2	**Implement Interactive** dashboards using modern libraries (Plotly, D3.js, Altair)
CO3	**Visualize Complex** data types: geospatial, temporal, hierarchical, network, high-dimensional
CO4	**Create Production** visual analytics systems with real-time capabilities

Unit	Content	Credit	Weightage
I	**Theoretical Foundations & Perceptual Principles** **Topics:** • History and theory of data visualization • Gestalt principles of visual perception • Color theory for visualization: color spaces, palettes, accessibility • Visual encoding: marks, channels, effectiveness ranking • Cognitive load and pre-attentive processing • Visualization criticism and evaluation frameworks • **Practical:** Critical analysis of existing visualizations using Munzner's framework	1	25%
II	**Advanced Chart Types & Interactive Systems** **Topics:** • Beyond bar charts: small multiples, trellis plots, parallel coordinates • Hierarchical data: treemaps, sunburst, icicle, dendrograms • Network/graph visualization: force-directed layouts, arc diagrams, matrix plots • Time series: horizon charts, calendar heatmaps, streamgraphs • Interactive principles: brushing & linking, coordinated multiple views	1	25%

	• Dashboard design patterns and best practices • **Practical:** Build an interactive dashboard with multiple linked views using Dash/Plotly		
III	**Geospatial, Temporal & High-Dimensional Visualization** **Topics:** • Geospatial: choropleth maps, heatmaps, point maps, flow maps • Web mapping: tile layers, projections, Mapbox/Leaflet integration • Temporal visualization: Gantt charts, timelines, animation principles • High-dimensional data: t-SNE, UMAP, PCA visualization • Text visualization: word clouds, sentiment maps, topic modeling visualizations • Streaming/real-time data visualization techniques • **Practical:** Create an animated geospatial visualization of temporal data	1	25%
IV	**Production Systems & Emerging Trends** **Topics:** • Visualization pipelines: data preparation to rendering optimization • Scalable visualization: level-of-detail, aggregation strategies • Web deployment: containerization, performance optimization • 3D visualization: principles and pitfalls, WebGL applications • Immersive visualization: AR/VR for data exploration • Ethical visualization: misleading charts, bias in visualization • Automated visualization and generative approaches • **Practical:** Deploy a scalable visualization application to cloud with real-time capabilities	1	25%

TEXT BOOKS:

- Primary: *The Visual Display of Quantitative Information* (2nd Ed.) — Edward R. Tufte
- Primary: *Interactive Data Visualization for the Web* (2nd Ed.) — Scott Murray
- Primary: *Visualization Analysis and Design* — Tamara Munzner

REFERENCE BOOKS:

- *Storytelling with Data: A Data Visualization Guide for Business Professionals* — Cole Nussbaumer Knaflic
- *Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures* — Claus O. Wilke
- *Design for Information: An Introduction to the Histories, Theories, and Best Practices*

Behind Effective Information Visualizations — Isabel Meirelles
- *Data Visualization: A Practical Introduction* — Kieran Healy
- *Python Data Visualization Cookbook* — Igor Milovanović, Dimitry Foures, Giuseppe Vettigli

ONLINE RESOURCES:

- MOOCs: Data Visualization Specialization (Coursera - UC Davis), Kaggle Visualization courses
- Communities: Observable Community, Reddit's r/dataisbeautiful

PRACTICAL LIST:

1. **Visualization Critique & Redesign**
   - Select 3 poor visualizations from media
   - Redesign them using perceptual principles
   - Present before/after with justification
2. **Interactive Dashboard Development**
   - Choose a complex dataset (e.g., NYC taxi trips, COVID-19 data)
   - Build a multi-view dashboard with at least 4 coordinated visualizations
   - Implement filtering, tooltips, and export functionality
3. **Geospatial Storytelling Project**
   - Use geospatial dataset (e.g., climate data, urban mobility)
   - Create a scrollytelling narrative visualization
   - Integrate maps with supplementary charts
   - Publish as an interactive web article
4. **Advanced Technique Implementation**
   - Option A: Network visualization of social/media data
   - Option B: Real-time streaming visualization (e.g., stock prices, IoT sensor data)
   - Option C: 3D visualization of multidimensional dataset
   - Document the implementation challenges and solutions
5. **Capstone: Production Visualization System**
   - End-to-end project: From data ingestion to deployed application
   - Must include: Data pipeline, interactive visualizations, user authentication (basic)
   - Performance optimization for large datasets
   - Documentation and user guide

MK UNIVERSITY

PATAN, GUJARAT

ESTABLISHED BY THE GUJARAT GOVT. RECOGNIZED BY UGC UNDER SECTION 2(f) OF UGC ACT,1956

SUBJECT CODE: MTDSE204

SUBJECT NAME: TIME SERIES ANALYSIS

Course Objectives:

- To understand fundamental concepts, components, and characteristics of time series data
- To master classical statistical methods for time series analysis and forecasting
- To implement modern machine learning and deep learning approaches for time series
- To develop skills in handling real-world challenges: seasonality, missing data, multi-series forecasting
- To build end-to-end forecasting pipelines with evaluation, validation, and deployment considerations

Course Outcomes: At the end of the course students shall be able to

CO1	**Preprocess** and analyze time series data, identifying trends, seasonality, and stationarity
CO2	**Implement** classical statistical models (ARIMA, SARIMA, ETS) for forecasting
CO3	**Apply** machine learning and deep learning models (LSTMs, Transformers) to time series problems
CO4	**Evaluate** and compare models using appropriate time series-specific validation techniques

Unit	Content	Credit	Weightage
I	**Foundations & Classical Decomposition** **Topics:** • Characteristics of time series: trend, seasonality, cyclicity, noise • Time series components: Additive vs Multiplicative models • Stationarity: tests (ADF, KPSS) and transformations (differencing, Box-Cox) • Exploratory analysis: autocorrelation (ACF), partial autocorrelation (PACF) • Classical decomposition: moving averages, STL decomposition • Basic forecasting methods: Naïve, Seasonal Naïve, Exponential Smoothing (SES) • **Practical:** Exploratory analysis of real datasets, stationarity testing, decomposition	1	25%
II	**Statistical Forecasting Models** **Topics:** • Exponential Smoothing methods: Holt, Holt-Winters (ETS models) • ARIMA family: AR, MA, ARMA, ARIMA, SARIMA models • Model identification: using ACF/PACF, auto-	1	25%

	ARIMA approaches • Model diagnostics: residual analysis, Ljung-Box test • Advanced statistical models: VAR (Vector Autoregression), GARCH for volatility • Forecasting evaluation: Time series cross-validation, metrics (MSE, MAE, MAPE, MASE) • **Practical:** Building and tuning ARIMA/SARIMA models, comparing with ETS		
III	**Machine Learning for Time Series** **Topics:** • Feature engineering for time series: lag features, rolling statistics, Fourier terms • Time series as supervised learning problem: sliding window approach • Tree-based models: Random Forests, Gradient Boosting for forecasting • Advanced feature extraction: tsfresh library • Automated forecasting: prophet (Facebook) for business time series • Multivariate forecasting and exogenous variables • **Practical:** Feature engineering pipeline, ML model implementation, prophet for business data	1	25%
IV	**Deep Learning & Advanced Topics** **Topics:** • Neural networks for sequences: RNNs, LSTMs, GRUs • Attention mechanisms and Transformers for time series • Architecture patterns: Seq2Seq, N-BEATS, DeepAR, TFT (Temporal Fusion Transformers) • Probabilistic forecasting and uncertainty quantification • Hierarchical and multi-series forecasting • Time series anomaly detection: statistical and ML approaches • Deployment considerations: model refresh, monitoring drift • **Practical:** Build LSTM/Transformer model, probabilistic forecasting, anomaly detection	1	25%

TEXT BOOKS:

- Primary: *Forecasting: Principles and Practice* (3rd Ed.) — Rob J. Hyndman & George Athanasopoulos (Open online textbook)
- Primary: *Time Series Analysis and Its Applications: With R Examples* (4th Ed.) — Robert H. Shumway & David S. Stoffer

REFERENCE BOOKS:

- *Practical Time Series Analysis: Prediction with Statistics and Machine Learning* — Aileen Nielsen
- *Deep Learning for Time Series Forecasting* — Jason Brownlee
- *Time Series Analysis: Forecasting and Control* (5th Ed.) — George E. P. Box, Gwilym M. Jenkins, Gregory C. Reinsel, Greta M. Ljung
- *An Introduction to State Space Time Series Analysis* — Jacques J.F. Commandeur & Siem Jan Koopman
- *Machine Learning for Time Series with Python* — Ben Auffarth

ONLINE RESOURCES:

- MOOCs: Time Series Courses (Coursera - UC Santa Cruz, Data camp, Udacity)
- Documentation: Official docs for sk time, stats models, prophet
- Communities: Towards Data Science TS articles, r/data science, Kaggle Time Series forums

PRACTICAL LIST:

1. **Exploratory Analysis & Classical Forecasting**
   - Analyze a dataset (e.g., air passengers, retail sales)
   - Perform decomposition, stationarity tests
   - Implement and compare naive, exponential smoothing, and ARIMA models
   - Report with visualizations and metric comparisons
2. **Statistical vs ML Forecasting**
   - Use complex dataset with seasonality and trends (e.g., electricity demand)
   - Build SARIMA model and diagnostic checks
   - Create ML pipeline with feature engineering (lag, rolling features)
   - Compare Random Forest/XGBoost with SARIMA
3. **Deep Learning for Forecasting**
   - Choose dataset suitable for deep learning (high frequency, multivariate)
   - Implement LSTM and Transformer-based model
   - Incorporate exogenous variables (e.g., weather in energy forecasting)
   - Compare performance with classical and ML approaches
4. **End-to-End Forecasting System (Capstone)**
   - Option A: Multi-series forecasting (e.g., retail products across stores)
   - Option B: Probabilistic forecasting with uncertainty intervals
   - Option C: Real-time anomaly detection system
   - Requirements: Data pipeline, model training, API endpoint, dashboard visualization

MK UNIVERSITY

PATAN, GUJARAT

ESTABLISHED BY THE GUJARAT GOVT. RECOGNIZED BY UGC UNDER SECTION 2(f) OF UGC ACT,1956

SEMESTER-III

SUBJECT CODE: MTDSE301

SUBJECT NAME: NATURAL LANGUAGE PROCESSING

Course Objectives:

- To provide a comprehensive understanding of linguistic foundations and computational techniques for natural language understanding and generation.
- To develop expertise in modern NLP techniques including statistical models, neural approaches, and transformer architectures.
- To implement and evaluate NLP systems for real-world applications such as text classification, machine translation, and conversational AI.
- To prepare students for research and development in cutting-edge NLP technologies and emerging language AI applications.

Course Outcomes: At the end of the course students shall be able to

CO1	Understand linguistic structures and apply preprocessing techniques for text data.
CO2	Design and implement statistical and neural NLP models for various language tasks.
CO3	Build and fine-tune transformer-based models for advanced NLP applications.
CO4	Develop end-to-end NLP systems and evaluate their performance with appropriate metrics.

Unit	Content	Credit	Weightage
I	**Foundations of NLP and Text Processing** **Topics:** • Introduction to linguistics: Morphology, syntax, semantics, pragmatics • Text preprocessing: Tokenization, stemming, lemmatization, stop-word removal • Regular expressions and pattern matching • N-gram language models and smoothing techniques • Word representations: One-hot encoding, TF-IDF, co-occurrence matrices • Collocations and statistical measures (PMI, chi-square) • Named Entity Recognition (NER) using rule-based and statistical methods • **Applications:** Text normalization, spell checking, basic information extraction	1	25%
II	**Statistical NLP and Classical Approaches** **Topics:** • Part-of-Speech tagging: Hidden Markov Models, Viterbi algorithm • Context-Free Grammars and parsing algorithms (CKY, Earley)	1	25%

	• Dependency parsing and constituency parsing • Semantic analysis: Word sense disambiguation, semantic roles • Sentiment analysis using lexicon-based and machine learning approaches • Topic modeling: Latent Dirichlet Allocation (LDA) • Text classification using Naïve Bayes, SVM, and logistic regression • **Applications:** Document classification, opinion mining, topic extraction		
III	**Neural NLP and Word Embeddings** **Topics:** • Distributed representations: Word2Vec (CBOW, Skip-gram), GloVe • Contextual embeddings: ELMo, ULMFiT • Neural sequence models for NLP tasks • Sequence-to-sequence models with attention mechanism • Neural machine translation • Text generation with RNNs and LSTMs • Neural text classification and sentiment analysis • **Applications:** Semantic similarity, document clustering, basic chatbots	1	25%
IV	**Transformer Models and Advanced NLP** **Topics:** • Transformer architecture: Self-attention, multi-head attention, positional encoding • BERT and its variants (RoBERTa, DistilBERT, ALBERT) • GPT models and autoregressive language modeling • Fine-tuning pre-trained language models • Question Answering systems (SQuAD) • Text summarization: Extractive and abstractive approaches • Dialogue systems and conversational AI • Ethical considerations: Bias in language models, fairness, interpretability • **Applications:** Chatbots, document summarization, intelligent search systems	1	25%

TEXT BOOKS:

- Jurafsky, D., & Martin, J.H. – *Speech and Language Processing* (3rd ed. draft) – Pearson.
- Eisenstein, J. – *Introduction to Natural Language Processing* – MIT Press.
- Goldberg, Y. – *Neural Network Methods for Natural Language Processing* – Morgan & Claypool.
- Clark, K., Khandelwal, U., Levy, O., & Manning, C.D. – *What Does BERT Look At?* – ACL.

REFERENCE BOOKS:

- Manning, C.D., & Schütze, H. – *Foundations of Statistical Natural Language Processing* – MIT Press.
- Bird, S., Klein, E., & Loper, E. – *Natural Language Processing with Python* – O'Reilly.
- Sarkar, D. – *Text Analytics with Python* – Apress.
- Howard, J., & Ruder, S. – *Universal Language Model Fine-tuning for Text Classification* – ACL.

ONLINE RESOURCES:

- Coursera: *Natural Language Processing Specialization*
- Stanford Online: *CS224N: Natural Language Processing with Deep Learning*
- Kaggle: NLP competitions and datasets
- Google Colab: GPU access for training NLP models

SUBJECT CODE: MTDSE302

SUBJECT NAME: CLOUD COMPUTING FOR DATA SCIENCE

Course Objectives:

- To understand cloud computing concepts, service models (IaaS, PaaS, SaaS), and deployment models
- To develop skills in deploying and managing data science workloads on major cloud platforms (AWS, Azure, GCP)
- To implement scalable data processing pipelines using cloud-native services
- To design and deploy machine learning models in production cloud environments
- To understand cloud security, cost optimization, and MLOps practices for data science

Course Outcomes: At the end of the course students shall be able to

CO1	**Compare and select** appropriate cloud services for different data science tasks
CO2	**Deploy scalable** data pipelines using cloud storage, databases, and processing services
CO3	**Containerize** data science applications using Docker and orchestrate with Kubernetes
CO4	**Implement end-to-end** ML pipelines using cloud ML services (SageMaker, Vertex AI, Azure ML)

Unit	Content	Credit	Weightage
I	**Cloud Foundations & Core Services** **Topics:** • Cloud computing fundamentals: IaaS, PaaS, SaaS, FaaS • Comparison of major providers: AWS vs Azure vs GCP • Core services: Compute (EC2/VMs, Lambda/Functions), Storage (S3/Blob/Cloud Storage) • Cloud databases: RDS, DynamoDB, Cosmos DB, Cloud SQL	1	25%

	• Networking basics: VPCs, security groups, load balancers • Identity and Access Management (IAM) • **Practical:** Set up cloud account, deploy VM, configure storage, implement basic IAM policies		
II	**Data Engineering in the Cloud** **Topics:** • Data lakes vs data warehouses on cloud (S3 + Redshift vs ADLS + Synapse) • Cloud ETL/ELT services: AWS Glue, Azure Data Factory, Dataflow • Big data processing: EMR, Databricks on cloud, HDInsight • Streaming data: Kinesis, Event Hubs, Pub/Sub • Serverless data processing: AWS Lambda, Azure Functions • Infrastructure-as-Code: Terraform/CloudFormation basics • **Practical:** Build cloud ETL pipeline from S3 to Redshift/ Big Query, implement with IaC	1	25%
III	**Machine Learning on Cloud Platforms** **Topics:** • Managed ML services: SageMaker, Azure ML, Vertex AI • ML pipeline components: data prep, training, tuning, deployment • AutoML services and managed notebooks • Distributed training on cloud (multi-node, GPU clusters) • Model deployment options: real-time endpoints, batch transforms • Model registry and versioning • **Practical:** Build and deploy ML model using SageMaker/Vertex AI, implement AutoML solution	1	25%
IV	**MLOps & Advanced Architectures** **Topics:** • Containerization for ML: Docker for data science environments • Container orchestration: Kubernetes basics for ML workloads • MLOps pipelines: CI/CD for ML, model monitoring, drift detection • Advanced architectures: microservices for ML, event-driven ML systems • Cost optimization: right-sizing, spot instances, auto-scaling • Security best practices: encryption, VPC endpoints,	1	25%

	private subnets • **Practical:** Containerize ML application, deploy on Kubernetes, implement monitoring dashboard		

TEXT BOOKS:

- Primary: *Cloud Computing for Science and Engineering* — Ian Foster & Dennis B. Gannon
- Primary: *Data Science on AWS* — Chris Fregly & Antje Barth

REFERENCE BOOKS:

- *Building Machine Learning Pipelines* — Hannes Hapke & Catherine Nelson
- *Kubernetes for Developers* — Joseph Heck
- *Cloud Native Data Center Networking* — Dinesh G. Dutt
- *MLOps Engineering on AWS* — Sireesha Muppala, Randy DeFauw, Shelbee Eigenbrode
- *The Azure Data Engineer's Cookbook* — Ahmad Osama

ONLINE RESOURCES:

- Certification paths: AWS/Azure/GCP Data & ML certifications
- Official Documentation: AWS Docs, Azure Architecture Center, Google Cloud Architecture Framework
- Communities: r/aws, Azure Reddit, Google Cloud Community, DevOps/SRE forums

PRACTICAL LIST:

1. **Cloud Infrastructure Setup & Data Pipeline**
   - Provision cloud resources using Terraform/CloudFormation
   - Set up data lake architecture (raw/processed/curated zones)
   - Implement ETL pipeline using cloud-native services (Glue/Data Factory)
   - Load data into cloud data warehouse
   - Document infrastructure and cost estimation
2. **Cloud-based ML Development**
   - Use cloud notebook service (SageMaker Studio/Vertex Notebooks)
   - Train model using managed training service
   - Perform hyperparameter tuning at scale
   - Deploy model as REST endpoint
3. **Containerization & Orchestration Project**
   - Dockerize a data science application with all dependencies
   - Create Dockerfile with optimized layers
   - Deploy on Kubernetes cluster (EKS/AKS/GKE)
   - Implement auto-scaling based on CPU/memory
   - Set up logging and basic monitoring
4. **End-to-End MLOps Pipeline (Capstone)**
   - Option A: Real-time fraud detection system on cloud
   - Option B: Recommender system with continuous retraining
   - Option C: Computer vision pipeline with batch/stream processing