Course Code	Course Name	Group	Teaching Scheme	Credits
			Lectures	
RJSPGDSAI101	**Introduction to Supervised Learning**	DSC	3	3

Course Outcome:

1. To understand the features of machine learning to apply it to real world problems.
2. To characterize the machine learning algorithms as supervised learning and unsupervised learning and Apply and analyze the various algorithms of supervised and unsupervised learning.
3. To analyze the concept of neural networks for learning linear and non-linear activation functions.
4. To learn the concepts in Bayesian analysis from probability models and methods

Learning Outcome:

1. Able to understand the basics of Unsupervised Learning.
2. Understand the informed and uninformed problem types and apply search strategies to solve them.
3. Design and evaluate intelligent expert models for perception and prediction from an intelligent environment.

UNIT	TOPICS
UNIT - I	**Introduction to machine learning** How machine learn, Basic structure of Machine Learning Model and comparison to traditional programming, Goal and Applications of Machine learning. Types of Data related to Machine Learning, Introduction to standard machine learning datasets and interpreting the data , Correlation between data, supervised vs Unsupervised learning algorithms.
UNIT - II	**Feature Engineering:** Feature transformation, feature construction, feature scaling(Standardization & Normalization), feature extraction: Principal Component Analysis, t-sne, curse of dimensionality, function and power transformation.
UNIT - III	**Regression :** Linear regression and assumptions, Feature Selection using Correlation Coefficient, Polynomial regression, Multivariate Regression, Logistic regression, Cost functions for Regression(Mean Square Error, Mean Absolute Error, Root Mean Square error, R2 Score) and optimization using gradient descent.

UNIT - IV	**Classification :** Linear & Non-Linear Datasets, Decision Tree Algorithm, Role of Entropy and Information Gain(IG) in decision tree algorithm, CART ,Random forest, Classification Metrics - Type I and Type II Errors and Confusion Matrix, Overfitting vs Under fitting, Bias variance trade off, Naive Bayes Classifier, Support Vector Machine (SVM), Kernels in SVM.

References

1. E. Alpaydin, Introduction to Machine Learning, 3rd Edition, MIT Press, 2015.
2. T Hastie, R Tibshirani and J Friedman, The Elements of Statistical Learning Data Mining, Inference, and Prediction, 2nd Edition, Springer, 2009.
3. C.Bishop, "Pattern Recognition and Machine Learning, Springer.
4. A.F..Vermeulen, "Practical Data Science", APress, 2018.
5. S.Ozdemir, "Principles of Data Science", Packt, 2016.

Course Code	Course Name	Group	Teaching Scheme	Credits
			Lectures	
RJSPGDSAI101P	**Introduction to Supervised Learning**	DSC	4	2

Practical List:

1. Write a python program to import and export data using Pandas library functions
2. Demonstrate various data pre-processing techniques for a given dataset
3. Write a Python program to demonstrate various Data Visualization Techniques.
4. Implement Simple and Multiple Linear Regression Models.
5. Develop Logistic Regression Model for a given dataset
6. Develop Decision Tree Classification model for a given dataset and use it to classify a new sample.
7. Implement Naïve Bayes Classification in Python
8. Build KNN Classification model for a given dataset.
9. Demonstrate Support Vector Machine for a given dataset.
10. Implement Random forest ensemble method on a given dataset.
11. Implement Boosting ensemble method on a given dataset.

Course Code	Course Name	Group	Teaching Scheme	Credits
			Lectures	
RJSPGDSAI102	**Data Analysis and Visualization**	DSC	3	3

Course Outcome:

1. Manipulate large datasets and handle missing or inconsistent values in datasets.
2. Perform statistical analysis using tools.
3. Discover and visualize data sets using PowerBi.

Learning Outcome:

1. Use data analysis tools in the pandas library.
2. Load, clean, transform, merge and reshape data.
3. Create informative visualization and summarize data sets.
4. Analyze and manipulate time series data.
5. Solve real world data analysis problems.

UNIT	TOPICS
UNIT - I	**Introduction to Data Analysis**: Data Analysis - Exploratory Data Analysis and Data Science Process - Responsibilities of a Data Analyst - Data Analytics vs. Data Analysis - Types of Data - Understanding Different Types of File Formats - Sources of Data - Languages for Data Professionals - Overview of Data Repositories - Data Marts, Data Lakes, ETL, and Data Pipelines - Foundations of Big Data - Identifying Data for Analysis
UNIT - II	**Data Sources** - How to gather and Import Data - Data Loading, Storage and File Formats - Reading and Writing Data in Text Format, Web Scraping, Binary Data Formats, interacting with Web APIs, Interacting with Databases – Data Wrangling - Hierarchical Indexing, Combining and Merging Data Sets Reshaping and Pivoting - Tools for Data Wrangling - Data Cleaning and Preparation - Handling Missing Data, Data Transformation, String Manipulation
UNIT - III	**Intro to data visualization** - Introduction to Visualization and Dashboarding Software - Visualization Tools - Getting started with Tableau Desktop – Connecting to the dataset - Creating charts – Creating common visualizations (bar charts, line charts etc.) - Filtering and sorting data - Adding Titles, Labels, and descriptions - Publish your work to Tableau Cloud - Interactivity with text and visual tooltips - Interactivity with actions (filter, highlight, URL) – Assembling dashboards from multiple charts

UNIT - IV	**Introduction to Power BI** - Understanding Desktop - Understanding Power BI Report Designer - Report Canvas, Report Pages: Creation, Renames - Report Visuals, Fields and UI Options - Experimenting Visual Interactions, Advantages - Reports with Multiple Pages and Advantages - Pages with Multiple Visualizations - PUBLISH Options and Report Verification in Cloud - Adding Report Titles. Report Format Options - Introduction to data storytelling - Creating a data story

References

1. Python for Data Analysis: Data Wrangling with Pandas, NumPy and IPython by McKinney, W., 2nd edition. O"Reilly Media, 2017
2. Doing Data Science: Straight Talk from the Frontline by O"Neil, C., & Schutt, R, O"Reilly Media, 2013
3. The Big Book of Dashboards by Steve Wexler, Jeffrey Shaffer, Andy Cotgreave, John Wiley & Sons, 2017
4. Practical Tableau by Ryan Sleeper, O"Reilly Media, 2018
5. Power BI. Book-1, Business Intelligence Clinic: Create and Learn by Roger F Silva, 2018
6. Introducing Microsoft Power BI by Alberto Ferrari and Marco Russo, Microsoft Press, Washington, 2016

Course Code	Course Name	Group	Teaching Scheme	Credits
			Lectures	
RJSPGDSAI102P	**Data Analysis and Visualization**	DSC	4	2

Practical List:

1. Show Basic Visualization in Python
2. Show Basic Visualization in R.
3. Connecting to Data and preparing data for visualization in Tableau
4. Use Data aggregation and statistical functions in Tableau.
5. Show Data Visualization using Tableau.
6. Use dashboards of Tableau.
7. Show Data Visualization using PowerBi.
8. Show Data Visualization using DataWrapper.
9. Show Data Visualization using Gnatt Chart.
10. Show Data Visualization using Zoho.
11. Publish visualised data on Cloud in PowerBi

Course Code	Course Name	Group	Teaching Scheme	Credits
			Lectures	
RJSPGDSAI201	**Introduction to Unsupervised Learning**	DSC	3	3

Course Outcome:

1. Understand how to evaluate clustering results using a variety of metrics.
2. Understand the tradeoffs and assumptions inherent in different clustering techniques.
3. Understand how unsupervised learning can be used to improve supervised prediction.
4. Perform density estimation using a kernel, with a single random variable.
5. Interpret a biplot result from principal components analysis (PCA).

Learning Outcome:

1. Correctly apply and interpret results from clustering methods in scikit-learn, including k-means, agglomerative clustering, hierarchical clustering, and DBSCAN.
2. Correctly apply and interpret results from manifold learning methods, including multidimensional scaling
3. Build awareness of other advanced methods like kernel PCA and spectral clustering.

UNIT	TOPICS
UNIT - I	**Introduction:** What Is Unsupervised Learning? Application of Unsupervised Learning, Types of Unsupervised Learning, Advantages and Disadvantages of Unsupervised Learning **Beyond Supervised Learning:** Data Augmentation, Transfer Learning,: Fine Tuning, Unsupervised pre-training, Semi-supervised learning, Meta-Learning, Few-shot Learning.
UNIT - II	**Dimensionality Reduction :**Curse of Dimensionality, Principal Component Analysis(PCA), Computational issues, choosing number of latent dimensions, t-SNE, PCA vs t-SNE, Singular Value Decomposition(SVD)
UNIT - III	**Introduction to Clustering :** Evaluation the output of clustering methods, Partition Based Clustering, K-Means clustering, elbow methods to find number of clusters, K-Medoids clustering, Hierarchical Clustering, Agglomerative and Divisive Approach, Dendrogram, linkage, Density-based spatial clustering of applications with noise(DBSCAN)
UNIT - IV	**Association rule learning :** Working with association rule learning, Support ,Confidence, Lift, Apriori Algorithm, Frequent Itemset, Steps of Apriori Algorithm, Fp-Growth Algorithm, Frequent Pattern(FP) Tree, Difference Between Fp-Growth and Apriori, Application of Association rule Learning,

Reference:

1. Probabilistic Machine learning An Introduction, Kevi P. Murphy.
2. Machine Learning: Master Supervised and Unsupervised Learning Algorithms with Real Examples by Dr Ruchi Doshi, Dr Kamal Kant Hiran, Ritesh Kumar Jain, Dr Kamlesh Lakhwani**.**

Course Code	Course Name	Group	Teaching Scheme	Credits
			Lectures	
RJSPGDSAI201P	**Introduction to Unsupervised Learning**	DSC	4	2

Practical List:

1. Implementation of data augmentation.
2. Build a model using transfer learning.
3. Demonstration of few shot learning.
4. Perform dimensionality reduction using principal component analysis.
5. Implementation of singular value decomposition.
6. Implementation of partition based clustering.
7. Implementation of density based clustering.
8. Implementation of hierarchical clustering.
9. Perform market basket analysis using apriori algorithm.
10. Implementation of FP-Growth algorithm.

Course Code	Course Name	Group	Teaching Scheme	Credits
			Lectures	
RJSPGDSAI202	**Data Engineering**	DSC	3	3

Course Outcome:

1. To develop the skills of managing the data with respect to knowledge generation.
2. Provide the ability to design the data engineering process
3. To propose the data reliability models
4. To define how to use Machine learning model

Learning Outcome:

1. Building the storage system with appropriate data technologies
2. Designing the data pipelines and data flow 3 1,4
3. Processing the data infrastructure 4 3
4. Investigate possible diagnostics by designing Databases for Reliability, Scalability, and Availability, Understanding Data Operations for Flexibility
5. Training and measuring the serving Infrastructure for Machine Learning Models

UNIT	TOPICS
UNIT - I	**Selecting Appropriate Storage Technologies:** From Business Requirements to Storage Systems, Technical Aspects of Data, Types Of Structure, Schema Design Consideration Building and **Operationalizing Storage Systems:** Cloud SQL, Cloud Spanner, Cloud Bigtable, Cloud Firestore, BigQuery, Cloud Memorystore, Cloud Storage, Unmanaged Databases **Designing Data Pipelines:** Overview Of Data Pipelines, GCP Pipeline Components, Migrating Hadoop and Spark To GCP **Designing a Data Processing Solution:** Designing Infrastructure, Designing for Distributed Processing, Migrating a Data Warehouse
UNIT - II	**Building and Operationalizing Processing Infrastructure:** Provisioning and Adjusting Processing Resources, Monitoring Processing Resources **Designing for Security and Compliance:** Identity and Access Management with Cloud IAM, Using IAM with Storage and Processing Services, Data Security, Ensuring Privacy with the Data Loss Prevention API, Legal Compliance
UNIT - III	**Designing Databases for Reliability, Scalability, and Availability:** Designing Cloud Bigtable Databases for Scalability and Reliability, Designing Cloud Spanner Databases for Scalability and Reliability, Designing BigQuery Databases for Data Warehousing **Understanding Data Operations for Flexibility and Portability:** Cataloging and Discovery with Data Catalog,

	Data Preprocessing with 12 Dataprep, Visualizing with Data Studio, Exploring Data with Cloud Datalab, Orchestrating Workflows with Cloud Composer **Deploying Machine Learning Pipelines:** Structure of ML Pipelines, GCP Options for Deploying Machine Learning Pipeline
UNIT - IV	**Choosing Training and Serving Infrastructure:** Hardware Accelerators, Distributed and Single Machine Infrastructure, Edge Computing with GCP Measuring, Monitoring, and **Troubleshooting Machine Learning Models:** Three Types of Machine Learning Algorithms, Deep Learning, Engineering Machine Learning Models, Common Sources of Error in Machine Learning Models **Leveraging Prebuilt Models as a Service:** Sight, Conversation, Language, Structured Data

Reference:

1. Professional Data Engineer DAN SULLIVAN Sybex - Wiley 3 rd 2020.
2. Data Driven Science and Engineering STEVEN L. BRUNTON, J. NATHAN KUTZ
3. Cambridge University Press 2nd 2019.
4. Data Security in Cloud Computing Vimal Kumar, Sivadon Chaisiri and Ryan Ko The
5. Institution of Engineering and Technology 2020.
6. 4. Data Engineering on Azure Vlad Riscutia Manning Publications 2021.

Course Code	Course Name	Group	Teaching Scheme	Credits
			Lectures	
RJSPGDSAI202P	**Data Engineering**	DSC	4	2

Practical List:

1. Collection of raw data from various sources
2. Implement of data preprocessing pipeline
3. Perform ETL operations using tool
4. Build GCP machine learning pipeline
5. Integration of data into different file formats
6. Designing Cloud Bigtable Databases for Scalability and Reliability
7. Troubleshooting machine learning model
8. Perform data visualization with tools.