IPS Academy, Institute of Engineering & Science

(A UGC Autonomous Institute, Affiliated to RGPV, Bhopal)

Department of Computer Science & Engineering (Data Science)

Bachelor of Technology (B.Tech.)

Honors Certification in Data Science and Machine Learning (DS & ML)

(For students of Computer Science & Engineering (Data Science))

S.No.	Subject Code	Semester	Subject Name	Contact Hours per week			Total Credits
				L	T	P	
1.	**HOCSDS-IML501**	**V**	**Statistics and Exploratory Data Analytics**	2	1	2	**4**
2.	**HOCSDS-IML601**	**VI**	**Data Analysis & Business Intelligence**	2	1	2	**4**
3.	**HOCSDS-IML701**	**VII**	**Deep Learning**	2	1	2	**4**
4.	**HOCSDS-IML801**	**VIII**	***Practical Machine Learning with Tensorflow**	2	1	-	**3**
			Total	**8**	**4**	**6**	**15**

1 Hr Lecture	1 Hr Tutorial	2 Hr Practical
1 Credit	1 Credit	1 Credit

Note: *VIII semester subject (Practical Machine Learning with Tensorflow or any other course equivalent to Practical Machine Learning with Tensorflow) can also be done from MOOC courses (NPTEL etc.) with minimum credit 3.

IPS Academy
Institute of Engineering & Science
Department of Computer Science & Engineering
(Data Science)
Bachelor of Technology (B.Tech.)

Honors Certification in Data Science & Machine Learning
(DS & ML)

HOCSDS-DML501	**Statistics and Exploratory Data Analytics**	**2L: 1T: 2P (5 hrs.)**	**Credits:04**

Course Objective:

Course Objective: The main objective of this course is to provide students with the foundations of probabilistic and statistical analysis using R Programming which is mostly used in varied applications in engineering

Module 1: **(08 hrs.)**

Introduction of R Software and Probability theory

Introduction to basic concepts of R programming language, Introduction to Probability , Sample Space and Events , Set Theory and Events using Venn Diagrams , Relative Frequency and Probability , Probability and Relative Frequency - An Example , Axiomatic Definition of Probability , Some Rules of Probability , Basic Principles of Counting- Ordered Set, Unordered Set, and Permutations , Basic Principles of Counting- Combination , Conditional Probability , Multiplication Theorem of Probability, Bayes' Theorem , Independent Events ..

Module 2: **(08 hrs.)**

Random Variables

Random Variables - Discrete and Continuous , Cumulative Distribution and Probability Density Function , Discrete Random Variables, Probability Mass Function and Cumulative Distribution Function , Expectation of Variables, Moments and Variance , Skewness and Kurtosis , Quantiles and Tschebyschevâ€™s Inequality .

Module 3: **(08 hrs.)**

Probability Distributions

Degenerate and Discrete Uniform Distributions, Bernoulli and Binomial Distribution, Poisson Distribution , Geometric Distribution , Continuous Random Variables and Uniform Distribution, Normal Distribution, Exponential Distribution , Bivariate Probability Distribution for Discrete

Random Variables, Bivariate Probability Distribution for Continuous Random Variables, Examples in Bivariate Probability Distribution Functions .

Module 4: **(08 hrs.)**

Testing of hypothesis-I

Covariance and Correlation , Bivariate Normal Distribution, Chi square Distribution, t - Distribution , F – Distribution, Distribution of Sample Mean, Convergence in Probability and Weak Law of Large Numbers , Central Limit Theorem , Needs for Drawing Statistical Inferences , Unbiased Estimators , Efficiency of Estimators , CramÃ©râ€“Rao Lower Bound and Efficiency of Estimators , Consistency and Sufficiency of Estimators, Method of Moments, Method of Maximum Likelihood and Rao Blackwell Theorem.

Module 5: **(08 hrs.)**

Testing of hypothesis-II

Basic Concepts of Confidence Interval Estimation ,Confidence Interval for Mean in One Sample with Known Variance ,Confidence Interval for Mean and Variance, Basics of Tests of Hypothesis and Decision Rules ,Test Procedures for One Sample Test for Mean with Known Variance ,One Sample Test for Mean with Unknown Variance ,Two Sample Test for Mean with Known and Unknown Variances ,Test of Hypothesis for Variance in One and Two Samples.

Course Outcome:

CO1: To explain R software and apply basic concepts of probability in engineering problems.
CO2: To explain the basic concepts of statistics and apply in engineering fields.
CO3: To explain and apply the concepts of probability distribution in evaluation of engineering problems.
CO4: Apply the various test of significance to structure engineering decision-making problems.
CO5: To identify and solve problems regarding probability and hypothesis.

List of Text Books / Reference Books:

1. Introduction to Statistics and Data Analysis With Exercises, Solutions and Applications in R Authors: Heumann, Christian, Schomaker, Michael, Shalabh, Publisher" Springer 2016
2. Applied Statistics and Probability for Engineers, Douglas C. Montgomery, George C. Runger, 2018, Wiley (Low price edition available)
3. Introduction to. Mathematical. Statistics. Robert V. Hogg. Allen T. Craig,, Low price Indian edition by Pearson Education
4. Probability and Statistics for Engineers. Richard A. Johnson, Irwin Miller, John Freund
5. Mathematical Statistics with Applications. Irwin Miller, Marylees Miller, Pearson Education
6. The R Software-Fundamentals of Programming and Statistical Analysis -Pierre Lafaye de Micheaux, Rémy Drouilhet, Benoit Liquet, Springer 2013
7. A Beginner's Guide to R (Use R) By Alain F. Zuur, Elena N. Ieno, Erik H.W.G. Meesters, Springer 2009.
8. **NPTEL Course Link: https://nptel.ac.in/courses/111/104/111104146/**

List of Experiments:

1. Installation and Working with R.
2. Installation and Working with R Studio.
3. Calculations with R as a Calculator, Calculations with Data Vectors.
4. Built-in Commands and Bivariate Plots.
5. Logical Operators and Selection of Sample.
6. Computation of Probability using R
7. Data Based Moments and Variance in R Software.
8. Discrete Uniform Distribution in R.
9. Binomial Distribution in R.
10. Poisson distribution in R.
11. Geometric Distribution in R.
12. Normal Distribution in R.
13. Chi square Distribution in R
14. Bivariate Probability Distribution in R Software.
15. Covariance and Correlation in R.

HOCSDS-DML601	Data Analysis & Business Intelligence	2L: 1T: 2P (5 hrs.)	Credits: 04

Course Objective:

Expecting to build a solid business analytics foundation, this course has been designed to impart knowledge of machine learning and statistical methods for data analysis. The course shall also provide sufficient knowledge of python programming language for machine learning algorithms and python/ R programming for statistical methods. A brief introduction to neural networks and deep learning will also be covered.

Rationale

Data analytics and business intelligence are of great importance in today's world. Data analysis is required to understand organizational problems and to explore data while business intelligence helps companies to make better decisions by showing current and historical data within their business context. To make the organization run smoother and more efficiently, analysts use BI to provide performance and competitor benchmarks.

Course Contents

Module 1: Foundation of Data Analytics & Python Programming **(08 hrs.)**

Foundation of Data Analytics: - Introduction ,Evolution , Concept and Scopes , Data , Big Data, Metrics and Data classification, Data Reliability & Validity, Problem Solving with Analytics, Different phases of Analytics in the business and Data science domain, Descriptive Analytics, Predictive Analytics and Prescriptive Analytics , Different Applications of Analytics in Business, Text Analytics and Web Analytics, Skills for Business Analytics , Concepts of Data Science, Basic skills required for understanding Data Science. Concepts of packages/libraries – important packages like NumPy, SciPy, scikit-learn, Pandas, Matplotlib, seaborn, etc., installing and loading packages, reading and writing data from/to different formats, simple plotting, functions, list comprehensions, database connectivity, Playing with Date Format.

Module 2: Probability & Statistics **(08 hrs.)**

Descriptive Analytics: Describing and summarizing data sets, measures of central tendency, dispersion, skewness, kurtosis, Correlation. Probability: Measures of probability, conditional probability, independent event, Bayes' theorem, random variable, discrete (binomial, Poisson, geometric, hypergeometric, negative binomial) and continuous (uniform, exponential, normal, gamma). Expectation and variance, markov inequality, chebyshev's inequality, central limit theorem.

Inferential Statistics: Sampling & Confidence Interval, Inference & Significance. Estimation and Hypothesis Testing, Goodness of fit, Test of Independence, Permutations and Randomization Test, t- test/z-test (one sample, independent, paired), ANOVA, chi-square.

Module 3: Data Munging with Python **(8 hours)**

Relevance in industry, Statistical learning vs machine learning, types and phases of analytics. Data pre-processing and cleaning: data manipulation steps (sorting, filtering, duplicates, merging, appending, subsetting, derived variables, data type conversions, renaming, formatting, etc.), normalizing data, sampling, missing value treatment, outliers. Exploratory data analysis: Data visualization using matplotlib, seaborn libraries, creating graphs (bar/line/pie/boxplot/histogram, etc.), summarizing data, descriptive statistics, univariate analysis (distribution of data), bivariate analysis (cross tabs, distributions and relationships, graphical analysis).

Module 4: Optimization in Analytics **(8 hours)**

Introduction to Operations Research (OR), Linear Programming Problems (LPP), Geometry of Linear programming, Sensitivity and Post-optimal analysis, Duality and its economic interpretation. Network models and project planning, Non-linear Programming – KKT conditions, Dynamic Programming.

Module 5: Introduction to SQL and Business Intelligence **(8 hours)**

Learning SQL query structure with examples, Data management and query system OLTP and OLAP and Their data models, Data warehousing, ETL and data integration Dashboard creation is using Tableau, Concepts of Business intelligence (BI), the relevance of BI in application to analytics industry and different domains.

Course Outcome:

1. Students will be able to apply principles of statistics, python programming, machine learning, probability and decision making to contexts in data analysis.
2. Moreover, they should design tested and effective advanced analytics models for decision making and communicate effectively in a variety of modes and contexts.

List of Text Books / Reference Books:

1. Rimvydas Skyrius**, Business** Intelligence: A Comprehensive Approach to Information Needs, Technologies and Culture, 2021

2. Tom Fawcett, Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking, 2013.

3. Rick Sherman, Business Intelligence Guidebook: From Data Integration to Analytics, 2014.

4. Swain Scheps, Business Intelligence for Dummies, 2008**.**

5. Carlo Vercellis, Business Intelligence: Data Mining and Optimization for Decision Making, 2006.

6. Anil Maheshwari, Data Analytics, 2017.

List of Experiments:

1. Study about different tools: Weka, XLMiner, Hadoop.
2. Write a note on KDD process.
3. Write a program to implement apriori algorithm in C language.
4. Demonstration of Association rule process on dataset contactlenses .arff using apriori algorithm.
5. Demonstration of classification rule process on dataset student .arff using j48 algorithm.
6. Introduction to the classification of Mining techniques.
7. Introduction to Decision Tree and Neural Networks.
8. Detailed study of any one BI tool: Oracle BI
9. Apply, if Analysis for data visualization. Design and generate necessary reports based on the data warehouse data.
10. Write a program to implement Data Analysis using Time Series Analysis.

HOCSDS-DML701	Deep Learning	2L: 1T: 2P (5 hrs.)	Credits:04

Course Objective:

This course aims to provide basic knowledge about the deep learning and computational challenges of building stable representations for high-dimensional data, such as images, text.

Module 1: **(08 hrs.)**

Introduction to Deep Learning, Machine learning Vs Deep learning, Feature Descriptor , Bayesian Learning , Discriminate Function , Linear Classifier.

Module 2: **(08 hrs.)**

Support Vector Machine, Linear Machine, Multiclass Support Vector Machine, Optimization, Optimization Techniques in Machine Learning, Nonlinear Functions.

Module 3: **(08 hrs.)**

Introduction to Deep Neural Network, Multilayer Perceptron, Back propagation Learning, Loss Function, Auto encoder, Auto encoder Vs. PCA , Auto encoder Training, Auto encoder Variants , Convolution, Cross Correlation.

Module 4: **(08 hrs.)**

CNN Architecture, MLP versus CNN, Popular CNN Architecture: LeNet, AlexNet,VGC16,Transfer Learning, Vanishing and Exploding Gradient, GoogleNet, ResNet, Optimisers: Momentum and Nesterov Accelerated Gradient(NAG), Adagrad, RMSprop, AdaDelta and Adam Optimiser

Module 5: **(08 hrs.)**

Normalization, Batch Normalization, Layer, Instance, Group Normalization, Training Trick, Regularization, Early Stopping , Face Recognition, Deconvolution Layer, Semantic Segmentation , Image Denoising, Variational Autoencoder , Generative Adversarial Network.

Course Outcome:

1. Understand the concept of Deep Learning.
2. Identify various learning model and optimization technique.
3. State the concepts of Deep Neural Network and Auto encoder.
4. Explain CNN Architecture.
5. Explain Normalization techniques.

List of Text Books / Reference Books:

1. Ian Goodfellow and Yoshua Bengio and Aaron Courville, “Deep Learning”, MIT Press, 2016.
2. Richard O. Duda, Peter E. Hart, David G. Stork, “Pattern Classification”, John Wiley & Sons Inc.
3. Andrew Glassner, “Deep Learning: A Visual Approach” Penguin Random House Publisher Services ,2021.
4. John D. Kelleher, “Deep Learning (The MIT Press Essential Knowledge series)
5. Francois Chollet, “ Deep Learning with Python ”.
6. **NPTEL Course Link: https://nptel.ac.in/courses/106/105/106105215/**

List of Experiments:

Apply and Analyze the Concepts of Deep Learning on following Data Set.

1. Image Data set
2. Text Data Set
3. Sound Data.
4. Biological Data Set.

HOCSDS-DML801	**Practical Machine Learning with TensorFlow**	**2L: 1T: 0P (3 Hrs.)**	**Credits:03**

Course Objective:

This course provides a foundation of machine learning using TensorFlow

Course Contents :(40 hrs)

Module 1 **(08 hrs.)**

Overview of Tensorflow & Machine Learning, Steps in Machine Learning Process, Loss Functions in Machine Learning, Gradient Descent, Gradient Descent Variations, Model Selection and Evaluation, Machine Learning Visualization.

Module 2: **(08 hrs.)**

Introduction to Tensors, Mathematical Foundations of Deep Learning - Building Data Pipelines for Tensorflow-Part, Text Processing with Tensorflow

Module 3: **(08 hrs.)**

Classify Images, Regression, Classify Structured Data, Text Classification, Under fitting and Overfitting, Save and Restore Models, CNNs, Transfer learning with pretrained CNNs, Transfer learning with TF hub, Image classification and visualization, Estimator API

Module 4: **(08 hrs.)**

Logistic Regression, Boosted Trees, Introduction to word embeddings, Recurrent Neural Networks ,Time Series Forecasting with RNNs, Text Generation with RNNs.,

Module 5: **(08 hrs.)**

TensorFlow Customization, Customizing tf.keras, TensorFlow Distributed Training

Course Outcome:

1. Understand the concept of Tensor Flow and Machine learning
2. State the concept of Text Processing with Tensorflow
3. Explain classification and visualization.
4. Explain RNN.
5. State the concept of Keras

List of Text / Reference Books:

1. Pramod Singh and Avish Manure, “Implement Machine Learning and Deep Learning Models with Python”, Apress 2019.
2. Rowel Atienza, “Advanced Deep Learning with TensorFlow 2 and Keras”, Packt Publishing Limited, 2020.
3. Josh Patterson and Adam Gibson, “Deep Learning: A Practitioner's Approach”, O'Reilly Media,2017.
4. Seth Weidman, “Deep Learning from Scratch: Building with Python from First Principles” O’Reilly Media, 2019.
5. **NPTEL Course Link: https://nptel.ac.in/courses/106/106/106106213/**