YESHWANTRAO CHAVAN COLLEGE OF ENGINEERING
(An Autonomous Institution affiliated to R T M Nagpur University Nagpur)
Accredited by NAAC (1st Cycle) with 'A' Grade (Score 3.25 on 4 Point Scale)

Wanadongri, Hingna Road, Nagpur-441110

Department of Information Technology
(Honors in DSML)

B.E. Honors in Data Sciences and Machine Learning
SoE & Syllabus 2021-22

B.E Honors in Data Sciences and Machine Learning

Information Brochure for BE Honors Program

1. Title of Program: **B.E. Honors in Data Sciences and Machine Learning**
2. Type of Program: **Honors**
3. Department offering the program: **Information Technology**
4. Industry / Association / Collaboration: **No**
5. Department eligible to opt for the program: **Information Technology**
6. **General information about courses in program:**

With increasing developments in Artificial Intelligence (AI), Internet of Things (IoT) and other smart technologies, data science and machine learning jobs are gaining higher exposure and demand in the technology market. The modules designed for BE Honors in Data Science and Machine Learning, covers the modules such as Python programming, statistics for Data Science, Natural Language Processing, Deep Learning, and Computer Vision. In this honors program you have five theory courses and three lab courses spanned across four semesters (fifth to 8th Semesters). With this specialization in your CV you can apply for highly paid hot jobs in industries and Research and Development Organizations. These roles include Data Scientists, Data Analysts, Machine Learning Engineers, Business Analytics Developers, Machine Learning Scientists, Statisticians, etc.

7. **Advanced knowledge or research orientation of Program:**

Apart from the basic knowledge, students will be given some exposure to the advanced knowledge, research problems and current research issues in the following areas of the program:

- Machine learning
- Data Science
- Computer Vision
- Deep Learning
- Natural Language Processing

Both theoretical lessons and laboratory experiments shall be used to provide the glimpse of research problems being solved in these areas.

8. Employability potential of program:

India is the second-highest country to recruit employees in the field of data science and Machine learning, etc. with 50,000 positions available – second only to the United States. The demand for data experts is equally competitive, whether you look at the big companies, the e-commerce industry, or even start-ups.

Few top recruiters in India in the area of Data Sciences and Machine Learning are:

IBM, Amazon, Siemens, Fractal Analytics, Mu Sigma, LinkedIn, Flipkart, Accenture, Deloitte, Citrix, Myntra, Capgemini.

The companies across various sectors are using data analysing tools to draw meaningful insights for their growth. In the coming future, the demand for data science professionals will continue to grow. Data science and Machine Learning is still evolving and has abundant opportunities to grow over the next decade or so.

9. Departmental Steering committee:

SN	Name of the Faculty Member	Post	Designation	e-mail ID	Contact Number
1.	Dr. R. C. Dharmik	HOD, IT & Chairman	Asstt. Prof.	raj_dharmik@yahoo.com	9158003335
2.	Prof. S.S.Chavhan	Member	Asst.prof	sschavhan@ycce.edu	8888832405
3.	Prof. S.W. Shende	Member	Asso. Prof.	shailendra.shende@gmail.com	9766698600
4.	Prof. A.D. Gaikwad	Member	Asst.prof	amolgaikwad.ag@gmail.com	9970743434

10. Program Coordinator:

SN	Name of the Faculty Member	Post	Designation	e-mail ID	Contact Number
	Prof. S.S.Chavhan	Program Coordinator	Asst.prof	sschavhan@ycce.edu	8888832405

Scheme of Examinations
Honors in Data Sciences and Machine Learning

SN	Sem	Sub. Code	Course Name	T/P	L	P	Hrs	Credits	MSEs	TA	ESE	ESE-Hr
1	V	ITH101	Statistics for Data Science	**T**	3	0	3	3	30	30	40	3
2	V	ITH102	Introduction to Machine Learning using Python	**P**	0	2	2	1		60	40	
3	V	ITH103	Introduction to Machine Learning using Python-Lab	**T**	3	0	3	3	30	30	40	3
4	VI	ITH111	Basics of Natural Language Processing	**T**	3	0	3	3	30	30	40	**3**
5	VI	ITH112	Basics of Natural Language Processing-Lab	**P**	0	2	2	1		60	40	
6	VI	ITH113	Introduction to shallow and Deep learning	**T**	3	0	3	3	30	30	40	**3**
7	VI	ITH114	Introduction to Shallow and Deep learning-Lab	**P**	0	2	2	1		60	40	
8	VII	ITH121	Computer Vision Essentials	**T**	3	0	3	3	30	30	40	**3**
Total					**15**	**6**	**21**	**18**				

MSEs* = Three MSEs of 15 Marks each will conducted and marks of better 2 of these 3 MSEs will be considered for Continuous Assessment

TA ** = for Theory : 20 marks on lecture quizzes, 8 marks on assignments, 2 marks on class performance
TA = for Practical : MSPA will be 15 marks each**

V Semester

ITH101	Statistics for Data Science			L= 3	T = 0	P = 0	Credits = 3
Evaluation Scheme *Best Two out of three MSE's would be considered	MSE-I*	MSE-II*	MSE-III*	TA	ESE	Total	ESE Duration
	15	15	15	30	40	100	3 Hrs
Prerequisites	Basic knowledge of Computer Programming, MATHEMATICS						

Course Objective

1. Understand a basic concepts of Statistics and its types and data analysis using statistics
2. Understand the concept of the estimators and various distributions
3. Understand the foundations for hypothesis testing.
4. Understand correlation and regression analysis.

Course Outcome

Students will be able to

CO1- Demonstrate the ability to apply fundamental concepts in exploratory data analysis and differentiate between descriptive and inferential statistics

CO2 – apply estimation and different statistical distributions

CO3-. Perform test of Hypothesis

CO4- Compute and interpret the results of Regression and Correlation Analysis

UNIT I :

Introduction to Statistics, Terminologies in Statistics, Categories in Statistics, introduction to python data science tool and environment , Introduction to Numpy, Pandas, data preprocessing / wrangling using python .

UNIT II :

Descriptive Statistics: Understanding Descriptive Analysis, fundamentals of descriptive statistics, measures of central tendency, asymmetry and variability, data visualization using python .

UNIT III :

Inferential Statistics: Understanding Inferential Analysis, distribution, the standard normal distribution, central limit theorem, standard error.

UNIT IV :

Estimators and estimates : Working with estimators and estimates, Confidence intervals - an invaluable tool for decision making, Calculating confidence intervals within a population with a known variance , Student's T distribution, Calculating confidence intervals within a population with an unknown variance,

Margin of error.
UNIT V: Hypothesis Testing: null and alternative hypotheses, establishing a rejection region and a significance level, Rejection region and significance level, Type I error vs Type II error
UNIT VI: Simple linear regression and verifying assumptions used in linear regression, Multivariate linear regression, model assessment, assessing importance of different variables, subset selection

Text Books:

	Title	Edition	Author	Publisher
1	Practical statistics for data scientist 50 essential concepts -		Peter Bruce and Andrew Bruce ,	O'reilly publication
2	Statistics for management ,	7^{th} edition ,	Levin Rubin	pearson publication.

VI Semester

ITH102	Introduction to Machine Learning using Python			L= 3	T = 0	P = 0	Credits = 3
Evaluation Scheme *Best Two out of three MSE's would be considered	MSE-I*	MSE-II*	MSE-III*	TA	ESE	Total	ESE Duration
	15	15	15	30	40	100	3 Hrs
Prerequisites	Basic knowledge of Computer Programming, Algorithms and Data Structures, Knowledge of probability theory, linear algebra and calculus						
Important Note	Students opting for BE Honours program in Data Science and Machine Learning are not allowed to take Machine Learning (Professional Elective)						

Course Objectives:

1. To introduce basic concepts of machine learning and explain the relative strengths and weaknesses of different machine learning Methods.
2. To understand the concept and different aspects of supervised learning
3. To understand the concepts and different aspects of unsupervised learning
4. To learn to apply supervised and unsupervised learning algorithms and to evaluate their performance

Course Outcomes:

After attending the course Students will be able to

1: Understand various models of supervised and unsupervised learning
2: Analyse a problem and identify appropriate machine learning paradigm to solve it using Python.
3: Apply supervised /unsupervised learning on given dataset and design the model to meet the desired outcomes
4: To evaluate the performance of supervised /unsupervised ML algorithms using Python

UNIT I :
Introduction to fundamental concepts of Machine Learning and its applications, Python essential libraries and tools: Jupyter Notebook, NumPy, SciPy, matplotlib, pandas, scikit-learn, Classification and Regression, Generalization, Overfitting, and Underfitting, Relation of Model Complexity to Dataset Size, Some Sample Datasets

UNIT II :
Supervised Machine Learning Algorithms: k-Nearest Neighbors, Linear Models, Naive Bayes Classifiers, Decision Trees, Logistic Regression, applications

UNIT III :
Support Vector Machines (SVM), Uncertainty Estimates from Classifiers, The Decision Function, Predicting Probabilities, Uncertainty in Multiclass Classification, applications

UNIT IV :
Challenges in Unsupervised Learning, Preprocessing and Scaling, Different Kinds of Preprocessing, Applying Data Transformations, Dimensionality Reduction, Feature Extraction, and Manifold Learning, Principal Component Analysis (PCA), applications

UNIT V: Clustering: k-Means Clustering, Agglomerative Clustering, DBSCAN, Comparing and Evaluating Clustering Algorithms, application
UNIT VI: Model Evaluation and Improvement: Cross-Validation, Cross-Validation in scikit-learn, Benefits of Cross-Validation, Stratified k-Fold Cross-Validation and Other Strategies, Evaluation Metrics and Scoring: Metrics for Binary Classification, Metrics for Multiclass Classification, Regression Metrics, Using Evaluation Metrics in Model Selection

Text Books:

	Title	Edition	Author	Publisher
1	Introduction to Machine Learning with Python: A Guide for Data Scientists		Andreas C. Müller, Sarah Guido	O'REILLY Publication

Reference Book:

	Title	Edition	Author	Publisher
1	Introduction to Machine Learning, , By	Third Edition	Ethem Alpaydin,	PHI

VI Semester

ITH103	Introduction to Machine Learning using Python- Lab			L= 0	T = 0	P = 1	Credits = 1
Evaluation Scheme	MSPA-I	MSPA-II	MSPA-III	MSPA-IV	ESE	Total	ESE Duration
	15	15	15	15	40	100	3 Hrs
Prerequisites							

Course Objective

1. To solve the real world problems using Machine learning with Python language as an implementation tool

Course Outcome

At the end of the course students will be able to

1. Implement the real-world problems using Python as an implementation tool

Experiments based on real world applications of Machine Learning from various domains implemented using Python, using various Python libraries and Tools.

Starting with a well-formed problem statement, each Application should be properly designed and implemented using the following steps:

1. Data collection
2. Data exploration and preparation
3. Model training
4. Model evaluation
5. Model improvement

Note: Number of Applications to be implemented will be decided by the course instructor.

VI Semester

ITH111	Basics of Natural Language Processing			L= 3	T = 0	P = 0	Credits = 3
Evaluation Scheme *Best Two out of three MSE's would be considered	MSE-I*	MSE-II*	MSE-III*	TA	ESE	Total	ESE Duration
	15	15	15	30	40	100	3 Hrs
Prerequisites							

Course Objective

- Learn students the leading trends and systems in natural language processing.
- Make them understand the concepts of morphology, syntax, semantics and pragmatics of the language and that they are able to give the appropriate examples that will illustrate the above mentioned concepts.
- Learn to recognize the significance of pragmatics for natural language understanding.
- Enable students to be capable to describe the application based on natural language processing and to show the points of syntactic, semantic and pragmatic processing.

Course Outcome

Students will be able to

- Understand approaches to syntax and semantics in NLP.
- Understand approaches to discourse, generation, dialogue and summarization within NLP.
- Understand current methods for statistical approaches to machine translation.
- Understand machine learning techniques used in NLP, including hidden Markov models and probabilistic context-free grammars, clustering and unsupervised methods, log-linear and discriminative models, and the EM algorithm as applied within NLP

UNIT I :

Introduction: What is Natural Language Processing, Brief history of the NLP, Applications of NLP, Challenges for NLP, stages of NLP ,Two approaches to NLP.

UNIT II :

Sequence labelling and noisy channel, argmax computation, Noisy channel application to NLP, Probabilistic parsing

UNIT III :

Preprocessing and language models: Introduction to word tokenization. sentence segmentation, stemming, word normalization.

Language Models: The role of language models. Simple N-gram models. Estimating parameters and smoothing

UNIT IV :

Part Of Speech Tagging and Sequence Labeling: Fundamental principles, challenges, accuracy measurement, Word categories. Hidden Markov Models.

UNIT V: Word sense Disambiguation: overlap based, Supervised,semi supervised and unsupervised methods. word net and other corpus for NLP. Resource constraints WSD.
UNIT VI: NLP and Information retrieval :IR basic ,IR model ,How NLP has used IR and ranked information retrieval.

Text Books:				
	Title	**Edition**	**Author**	**Publisher**
1	Speech and Language Processing		Jurafsky and Martin	Prentice Hall
2				
Reference Book:				
	Title	**Edition**	**Author**	**Publisher**
1	Natural language understanding		James allen	Pearson
2	Learning Python		Lutz and Ascher	O'Reilly

VI Semester

ITH112	Basics of Natural Language Processing-Lab			L= 0	T = 0	P = 1	Credits = 1
Evaluation Scheme	MSPA-I	MSPA-II	MSPA-III	MSPA-IV	ESE	Total	ESE Duration
	15	15	15	15	40	100	3 Hrs
Prerequisites							

Course Objective

- Learn students the leading trends and systems in natural language processing.
- Make them understand the concepts of morphology, syntax, semantics and pragmatics of the language and that they are able to give the appropriate examples that will illustrate the above mentioned concepts.
- Learn them to recognize the significance of pragmatics for natural language understanding.
- Enable students to be capable to describe the application based on natural language processing and to show the points of syntactic, semantic and pragmatic processing.

Course Outcome

Students will be able to

- Understand approaches to syntax and semantics in NLP.
- Understand approaches to discourse, generation, dialogue and summarization within NLP.
- Understand current methods for statistical approaches to machine translation.
- Understand machine learning techniques used in NLP, including hidden Markov models and probabilistic context-free grammars, clustering and unsupervised methods, log-linear and discriminative models, and the EM algorithm as applied within NLP

(Experiment List)

1. Installing and understanding NLTK in python.
2. Processing of Raw text
3. Accessing text corpora and lexical resources.
4. Categorizing and Tagging the words.
5. Classification of text
6. Extracting information form text.
7. Practical based on WSD.

Text Books:

	Title	Edition	Author	Publisher
1	Speech and Language Processing		Jurafsky and Martin	Prentice Hall

Reference Book:

	Title	Edition	Author	Publisher
1	Natural language understanding		James allen	Pearson

		May 2021	1.00	Applicable for AY2021-22 Onwards
Chairperson	Dean (Acad. Matters)	Date of Release	Version	

2	Learning Python		Lutz and Ascher	O'Reilly

VI Semester

ITH113	**Introduction to shallow and Deep learning**			L= 3	T = 0	P = 0	Credits = 3
Evaluation Scheme *Best Two out of three MSE's would be considered	MSE-I*	MSE-II*	MSE-III*	TA	ESE	Total	ESE Duration
	15	15	15	30	40	100	3 Hrs

Course Objectives

1. To understand the theoretical foundations, algorithms and methodologies of Neural Network
2. To introduce the fundamental theory and concepts of machine learning
3. To provide a comprehensive foundation to artificial neural networks and their applications to pattern recognition.
4. To explore the learning paradigms of supervised and unsupervised shallow/deep neural networks
5. To impart adequate knowledge on deep learning frameworks and their applications to solving engineering problems

Course Outcome

Students will be able to

1. Familiarize with the mathematical foundations and working of neural networks as pattern classifier
2. Comprehend the neural networks as means for computational learning and to analyze the basic network architectures and algorithms for supervised and unsupervised learning
3. Gain knowledge about basic concepts of machine learning algorithms and identify machine learning techniques suitable for the given problem
4. Understand the differences between shallow neural networks and deep neural networks for supervised and unsupervised learning
5. Identify the deep feed forward, convolution and recurrent neural networks which are more appropriate for various types of learning tasks in various domains
6. Recognize the characteristics of deep learning models that are useful to solve real-world problems.
7. Identify and apply appropriate deep learning algorithms for analyzing the data for variety of problems

UNIT I :

An Introduction to Neural Networks,The Basic Architecture of Neural Networks, supervised and unsupervised learning, Neural Network learning rules, Neural Network with Backpropagation, Practical Issues in Neural Network Training, Common Neural Architectures

UNIT II :
Machine Learning with Shallow Neural Networks, Neural Architectures for Binary Classification Models, Neural Architectures for Multiclass Models, Matrix Factorization with Autoencoders

UNIT III :
Teaching Deep Learners to Generalize, The Bias-Variance Trade-Of, Generalization Issues in Model Tuning and Evaluation, Penalty-Based Regularization, Ensemble Methods, Early Stopping, Unsupervised Pretraining, Continuation and Curriculum Learning, Parameter Sharing, Regularization in Unsupervised Applications Radial Basis Function Networks, Training an RBF Network, Variations and Special Cases of RBF Networks, Relationship with Kernel Methods

UNIT IV :
Restricted Boltzmann Machines, Hopfield Networks, The Boltzmann Machine, Restricted Boltzmann Machines, Applications of Restricted Boltzmann Machines

UNIT V:
Recurrent Neural Networks, The Architecture of Recurrent Neural Networks, The Challenges of Training Recurrent Networks, Applications of Recurrent Neural Networks

UNIT VI:
Convolutional Neural Networks, The Basic Structure of a Convolutional Network, Training a Convolutional Network, Applications of Convolutional Networks
Deep Reinforcement Learning, Stateless Algorithms, The Basic Framework of Reinforcement Learning, Policy Gradient Methods

Text Books:

	Title	Edition	Author	Publisher
1	Neural Network and Deep Learning		Charu C Agarwal	Springer

Reference Book:

	Title	Edition	Author	Publisher
1	Neural Networks and Learning Machines	Third Edition,	Simon Haykin	Pearson, Prentice Hall

VI Semester

ITH114	Introduction to Shallow and Deep learning-Lab			L= 0	T = 0	P = 1	Credits = 1
Evaluation Scheme	MSPA-I	MSPA-II	MSPA-III	MSPA-IV	ESE	Total	ESE Duration
	15	15	15	15	40	100	3 Hrs

Course Objective

1. Understand different methodologies to create application using deep nets
2. To design and develop an application using specific deep learning models
3. To provide the practical knowledge in handling and analysing real world applications.

Course Outcome

Students will be able to

1. Understand the foundations of neural networks, how to build neural networks and learn how to lead successful machine learning projects
2. Develop and train neural networks for classification, regression and clustering
3. Implement deep learning algorithm and solve real world problems

1. Implementation of Single layer Perceptron Learning Algorithm.
2. Implement unsupervised learning algorithm by tacking appropriate input dataset.
3. Implement character recognition neural network for recognizing English alphabets using Back Propagation training algorithm.
4. Implement a neural network for handwritten character recognition and classification.
5. Implement a neural network for Optical character recognition in python.
6. Implement RBF network for input dataset and perform testing and training.
7. Implement Hopfield Networks in python by tacking appropriate input dataset.
8. Implement Convolutional neural network for pattern classification.
9. Implement Convolutional neural network for object detection.
10. Implement Reinforcement Q-Learning in python with OpenAI Gym.

VII Semester

ITH121	Computer Vision Essentials using Open CV			L= 3	T = 0	P = 0	Credits = 3
Evaluation Scheme *Best Two out of three MSE's would be considered	MSE-I*	MSE-II*	MSE-III*	TA	ESE	Total	ESE Duration
	15	15	15	30	40	100	3 Hrs
Prerequisites	Knowledge of Computer programming using Python						

Course Objectives:

1. To understand the processing in the spatial and frequency domain
2. To understand and apply the knowledge of image and video processing to solve real world problems problems
3. To expose students to implementation using OpenCV

Course Outcomes:

Students will be able to

1. understand and explain image processing in the spatial and frequency domain
2. understand and apply image and video processing knowledge for solving real world problems
3. develop Computer Vision applications using OpenCV

UNIT I :

Human Visual System: Eye, Retina, and the vision in the brain, Review of digital images processing in Spatial domain

UNIT II :

Introduction to digital images processing in the frequency domain using Fourier Transform

UNIT III :

Introduction to Open CV, the Core Functionality (core module), Image Processing module, High level GUI and media module, Image input and output m

UNIT IV :

Video Input and Output module, Camera calibration and 3D reconstruction, 2D Features framework

UNIT V:

Video analysis (video module), Object Detection module

UNIT VI:

Image Stitching Module, GPU-Accelerated Computer Vision (CUDA module), developing CV applications

Text Books:				
	Title	**Edition**	**Author**	**Publisher**
1	Digital Image Processing	Third edition 2007	Rafael C. Gonzalez, Richard E. Woods	PH
2	Open CV Tutorials: https://docs.opencv.org/master/d9/df8/tutorial_root.html			

Reference Book:				
	Title	**Edition**	**Author**	**Publisher**
1	Computer Vision: Algorithms and Applications	Second	Richard Szeliski	Springer