REVISED UG SYLLABUS UNDER CBCS
(Implemented from Academic Year 2020-21)

PROGRAMME: FOUR YEAR B.Sc. (Hons)

Domain Subject: **B.Sc -Data Science**
Skill Enhancement Courses (SECs) for Semester V, from 2022-23 (Syllabus/Curriculum)
Pair Options of SECs for Semester–V
(To choose one pair from the five alternate pairs of SECs)

Univ. Code	Courses 6&7	Name of Course	Th. Hrs. / Week	IE Marks	EE Marks	Credits	Prac. Hrs./ Wk	Marks	Credits
	6A	**DATA ANALYTICS WITH TABLEAU**	**3**	**25**	**75**	**3**	**3**	**50**	**2**
	7A	**AI CONCEPTS AND TECHNIQUES WITH PYTHON**	**3**	**25**	**75**	**3**	**3**	**50**	**2**
OR									
	6B	**SUPERVISED ML WITH PYTHON**	**3**	**25**	**75**	**3**	**3**	**50**	**2**
	7B	**UNSUPERVISED ML WITH PYTHON**	**3**	**25**	**75**	**3**	**3**	**50**	**2**
OR									
	6C	**NLP WITH PYTHON**	**3**	**25**	**75**	**3**	**3**	**50**	**2**
	7C	**DEEP LEARNING NEURAL NETWORKS WITH PYTHON**	**3**	**25**	**75**	**3**	**3**	**50**	**2**

Note-1: For Semester–V, for the domain subject DATA SCIENCE, any one of the three pairs of SECs shall be chosen as courses 6 and 7, i.e., 6A & 7A or 6B & 7B or 6C & 7C. The pair shall not be broken (ABC allotment is random, not on any priority basis).

Note-2: *One of the main objectives of Skill Enhancement Courses (SEC) is to inculcate field skills related to the domain subject in students. The syllabus of SEC will be partially skill oriented. Hence, teachers shall also impart practical training to students on the field skills embedded in the syllabus citing related real field situations.*

VTH SEMESTER
SKILL ENHANCEMENT COURSES

SKILL ENHANCEMENT COURSE-I

6A. DATA ANALYTICS WITH TABLEAU	--- DATA ANALYTICS WITH TABLEAU LAB
7A. AI CONCEPTS WITH PYTHON	--- AI CONCEPTS AND TECHNIQUES WITH PYTHON LAB

SKILL ENHANCEMENT COURSE-II

6B.SUPERVISED ML WITH PYTHON	-- SUPERVISED ML WITH PYTHON LAB
7B.UNSUPERVISED ML WITH PYTHON	-- UNSUPERVISED ML WITH PYTHON LAB

SKILL ENHANCEMENT COURSE-III

6C. NLP WITH PYTHON	-- NLP WITH PYTHON LAB
7C. DEEP LEARNING NEURAL NETWORKS WITH PYTHON	-- DEEP LEARNING NEURAL NETWORKS WITH PYTHON LAB

Semester-wise Revised Syllabus under CBCS, 2020-21
Four Year B.Sc. (Hons) - Semester – V (from 2022-23)
Subject: **B.Sc -Data Science**
Course-6A: **Data Analytics with Tableau**
(Skill Enhancement Course (Elective), 5 credits, Max Marks: 100 + 50

Learning Outcomes

Students at the successful completion of the course will be able to:

1. Understand Big Data and its usage
2. Identify various Data Quality and Preprocessing methods
3. Learn different Clustering techniques and Frequent Pattern Mining
4. Understand Regression, Classification and additional Predictive Methods

Syllabus: ***(Total Hours: 90 including Teaching, Lab and internal exams, etc.)***

UNIT I

Introduction to Data Analytics: Big Data and Data Science, Big Data Architectures, A Short Taxonomy of Data Analytics, Examples of Data Use, History on Methodologies for Data Analytics.

Descriptive Statistics: Scale Types, Descriptive Univariate Analysis, Descriptive Bivariate Analysis.

UNIT II

Descriptive Multivariate Analysis: Multivariate Frequencies, Multivariate Data Visualization, Multivariate Statistics, Infographics and Word Clouds.

Data Quality and Preprocessing: Data Quality, converting to a Different Scale Type, Converting to a Different Scale, Data Transformation, Dimensionality Reduction.

UNIT III

Clustering: Distance Measures, Clustering Validation, Clustering Techniques.

Frequent Pattern Mining: Frequent Itemsets, Association Rules, Behind Support and Confidence, Other Types of Pattern.

UNIT IV

Regression: Predictive Performance Estimation, Finding the Parameters of the Model, Technique and Model Selection.

Classification: Binary Classification, Predictive Performance Measures for Classification, Distance-based Learning Algorithms, Probabilistic Classification Algorithms.

UNIT V

Additional Predictive Methods: Search-based Algorithms, Optimization-based Algorithms.

Advanced Predictive Topics: Ensemble Learning, Algorithm Bias, Non-binary Classification Tasks, Advanced Data Preparation Techniques for Prediction.

III Text Books:

1. “A General Introduction to Data Analytics” by João Mendes Moreira, André C. P. L. F. de Carvalho, TomášHorváth, 2019 Edition, Wiley Publications.

2. “Data Analytics: Principles, Tools and Practices” by Dr. Gaurav Aroraa, ChitraLele, Dr. Munish Jindal, 2022 Edition, pbp publications

3. “Data Analytics” by Anil Maheshwari, First Edition, McGraw Hill Education

IV 6A (L): DATA ANALYTICS WITH TABLEAU LAB

OBJECTIVES:

To implement Map Reduce programs for processing big data

To realize storage of big data using H base, Mongo DB

To analyze big data using linear models

To Analyse big data using machine learning techniques such as SVM / Decision tree classification and clustering

LIST OF EXPERIMENTS

Hadoop
1. Install, configure and run Hadoop and HDFS
2. Implement word count / frequency programs using MapReduce
3. Implement an MR program that processes a weather dataset

R
4. Implement Linear and logistic Regression
5. Implement SVM / Decision tree classification techniques
6. Implement clustering techniques
7. Visualize data using any plotting framework
8. Implement an application that stores big data in Hbase / MongoDB / Pig using Hadoop / R.

III Text Books:

1. “A General Introduction to Data Analytics” by João Mendes Moreira, André C. P. L. F. de Carvalho, TomášHorváth, 2019 Edition, Wiley Publications.

2. “Data Analytics: Principles, Tools and Practices” by Dr. Gaurav Aroraa, ChitraLele, Dr. Munish Jindal, 2022 Edition, pbp publications

3. “Data Analytics” by Anil Maheshwari, First Edition, McGraw Hill Education

MODEL QUESTION PAPER (Sem-end. Exam)

B. Sc DEGREE EXAMINATION

SEMESTER –V

Course 6A: Data Analytics with Tableau

Time:3Hrs **Max.marks:75**

Section – A

(Answer any five of the following) **5x5=25M**

1. Write about natural taxonomy that exists in data analytics.
2. What are the multivariate frequencies?
3. Write about Clustering Validation.
4. Explain about simple linear regression model.
5. Write about Random Forests.
6. Write about Two Quantitative Attributes with an example.
7. Write about missing values in the data set.
8. Explain about Eclat

SECTION-B **5X10=50M**

9. Explain about The CRISP-DM Methodology.

(OR)

Explain about Univariate Data Visualization.

10. Explain about Multivariate Data Visualization.

(OR)

Explain about Converting data in a scale to another scale of the same type.

11. Finite about Distance Measures for Non-conventional Attributes.

(OR)

Explain about Apriori – a Join-based Method.

12. Explain about Predictive Performance Measures for Regression.

(OR)

Explain about binary classification.

13. Explain about back propagation in MLP.

(OR)

Explain about Algorithm Bias.

Semester-wise Revised Syllabus under CBCS, 2020-21
Four Year B.Sc. (Hons) - Semester – V (from 2022-23)
Subject: **B.Sc -Data Science**
Course-7A: **AI Concepts and Techniques with Python**
(Skill Enhancement Course (Elective), 5 credits, Max Marks: 100 + 50

Objectives of Course (AI Concepts and Techniques with Python):
This course provides an introduction to the fundamentals of artificial intelligence. Demonstrates fundamental understanding of the history of artificial intelligence (AI) and its foundations. Apply basic principles of AI in solutions that require problem solving, inference, perception, knowledge representation, and learning. Demonstrates awareness and a fundamental understanding of various applications of AI techniques in intelligent Agents.

Learning outcomes of Course:
1. List the objectives and functions of modern Artificial Intelligence.
2. Categorize an AI problem based on its characteristics and its constraints.
3. Understand and implement search algorithms.
4. Learn how to analyze the complexity of a given problem and come with suitable optimizations.
5. Demonstrate practical experience by implementing and experimenting with the learnt algorithms.

Syllabus: (Total Hours: 90 including Teaching, Lab and internal exams, etc.)

UNIT- 1
Problems and Search: What is Artificial Intelligence, The AI Problems, and Underlying Assumption, what is an AI Technique.
Problems, Problems Spaces, and Search: Defining the problem as a state space search, production systems, problems characteristics, issues in the design of search programs.
UNIT- II
Heuristic Search Techniques: Generate-and-test, Hill Climbing, Best-First Search, Problem Reduction, Constraint Satisfaction, Means-Ends Analysis
UNIT- III
Knowledge Representation Issues: Representations and Mapping, Approaches to Knowledge Representation, The frame problem. Using Predicate Logic: Representing simple facts in logic, Representing Isa relationships, predicates, Resolution
UNIT- IV
Representing Knowledge using Rules: Procedural Vs Declarative knowledge, Logic Programming, Forward Vs Backward Reasoning, Matching, Control Knowledge
UNIT- V
Symbolic Reasoning under Uncertainty: Introduction to Non-monotonic Reasoning, Logics for Non-monotonic Reasoning, Implementation issues, Augmenting a Problem solver, implementation: DFS, BFS.
Statistical Reasoning: Probability and Bayes Theorem, Certainty Factors and Rule-Based Systems, Bayesian Networks, Dempster-Shafer Theory.

III Textbooks:

Artificial Intelligence, Second Edition, Elaine Rich, Kevin Knight, Tata McGraw-Hill Edition.

References:

Russell, S., & Norvig, P. Artificial intelligence: a modern approach. Third Edition. Pearson new international edition. 2014.

IV Details of Lab/Practical/Experiments/Tutorials syllabus:

7A (L): AI Concepts and Techniques with Python Lab

1. Write a Program to Implement Breadth First Search using Python.
2. Write a Program to Implement Depth First Search using Python.
3. Write a Program to Implement Tic-Tac-Toe game using Python.
4. Write a Program to implement 8-Puzzle problem using Python.
5. Write a Program to Implement Water-Jug problem using Python.
6. Write a Program to Implement Travelling Salesman problem using Python.
7. Write a Program to Implement Towers of Hanoi problem using Python.
8. Write a Program to implement 8-Queens problem using Python.

MODEL QUESTION PAPER (Sem-end. Exam)

B. Sc DEGREE EXAMINATION

SEMESTER –V

Course 7A: AI Concepts and Techniques with Python

Time:3Hrs **Max.marks:75**

SECTION-A

(Answer any five of the following) **5x5=25M**

1) What is AI Technique?
2) Define State space search
3) Explain Generate and test
4) What is heuristic search technique?
5) What is resolution?
6) Explain Uncertainty implementation issues
7) Explain Bayes Theorem
8) Define Dempster-Shafer Theory.

SECTION-B

5X10=50M

9) a) Define Artificial Intelligence. Applications and characteristics of AI.

(or)

b) Explain the state space representation of Water – Jug problem.

10) a) Define Heuristic search? What are the advantages of Heuristic search?

(or)

b) Describe the Hill climbing.

11) a) What is predicate logic? Explain the predicate logic representation with reference to suitable example.

(or)

b) Describe the approaches to Knowledge Representation and explain the Issues in Knowledge Representation

12) a) Explain Procedural Vs Declarative knowledge

(or)

b) Explain the Issues in Knowledge Representation. Write notes on control knowledge.

13) a) Show how to implement Non-monotonic reasoning using JTMS in medical diagnosis. Consider rules such as "If you have a runny nose, assume you have a cold unless it is Allergy season."

(or)

b) Explain logics for Non-monotonic reasoning and discuss the implementation issues.

Semester-wise Revised Syllabus under CBCS, 2020-21
Four Year B.Sc. (Hons) - Semester – V (from 2022-23)
Subject: **B.Sc -Data Science**
Course-6B: **Supervised ML with Python**
(Skill Enhancement Course (Elective), 5 credits, Max Marks: 100 + 50

Objectives of Course:
The purpose of this course is to serve as an introduction to Supervised machine learning with Python. We will explore several classifications, regression algorithms and see how they can help us perform a variety of Supervised machine learning tasks.

Learning outcomes of Course:
- Able to understand introduction to machine learning concepts.
- Able to Loading datasets, build models and model persistence.
- Understand Feature extraction from data sets.
- Able to do Regression & Classification.
- Able to compare SVM with other classifiers.

Syllabus: ***(Total Hours: 90 including Teaching, Lab and internal exams, etc.)***

UNIT- 1
Machine Learning Basics: What is machine learning? Key terminology, Key tasks of machine learning, How to choose right algorithm, steps in developing a machine learning, why python? Getting started with Numpy library
Classifying with k-Nearest Neighbors: The k-Nearest Neighbors classification algorithm, Parsing and importing data from a text file, Creating scatter plots with Matplotlib, Normalizing numeric values
UNIT- II
Splitting datasets one feature at a time-Decision trees: Introducing decision trees, measuring consistency in a dataset, using recursion to construct a decision tree, plotting trees in Matplotlib
UNIT- III
Classifying with probability theory-Naïve Bayes: Using probability distributions for classification, learning the naïve Bayes classifier, Parsing data from RSS feeds, using naïve Bayes to reveal regional attitudes
UNIT- IV
Logistic regression: Classification with logistic regression and the sigmoid function, Using optimization to find the best regression coefficients, the gradient descent optimization algorithm, Dealing with missing values in the our data
UNIT- V
Support vector machines: Introducing support vector machines, using the SMO algorithm for optimization, using kernels to “transform” data, Comparing support vector machines with other classifiers

III Textbooks:
Machine learning in action, Peter Harrington by Manning publications

IV Lab Experiments
6B (L): Supervised ML with Python LaB

1. Implement and demonstrate the FIND-S algorithm for finding the most specific hypothesis based
on a given set of training data samples. Read the training data from a .CSV file.
2. For a given set of training data examples stored in a .CSV file, implement and demonstrate the
Candidate-Elimination algorithm to output a description of the set of all hypotheses consistent with
the training examples.
3. Write a program to demonstrate the working of the decision tree based ID3 algorithm.
4. Write a program to implement the naïve Bayesian classifier for a sample training data set stored as
a CSV file.
5. Assuming a set of documents that need to be classified, use the naïve Bayesian Classifier model to
perform this task. Built-in Java classes/API can be used to write the program. Calculate the accuracy, precision, and recall for your data set.
6. Write a program to construct a Bayesian network considering medical data. Use this model to
demonstrate the diagnosis of heart patients using standard Heart Disease Data Set. You can use
Java/Python ML library classes/API.
7. Write a program to implement k-Nearest Neighbour algorithm to classify the iris data set. Print
both correct and wrong predictions. Java/Python ML library classes can be used for this problem.
8. Implement the non-parametric Locally Weighted Regression algorithm in order to fit data points.
Select appropriate data set for your experiment and draw graphs.

MODEL QUESTION PAPER (Sem-end. Exam)

B. Sc DEGREE EXAMINATION

SEMESTER –V

Course 6B: Supervised ML with Python

Time:3Hrs **Max.marks:75**

SECTION-A

(Answer any five of the following) **5x5=25M**

1. What is Machine Learning? Discuss its key terminology.
2. How to Normalize numeric values.
3. What is Decision tree? Explain.
4. Discuss how to reveal regional attributes.
5. Explain Logistic Regression
6. Define Support Vector Machine.
7. How to deal with missing values.
8. List some of the Numpy library functions**.**

SECTION-B

Answer all the questions. Each question carries 10 marks **5X10=50M**

9. A) Discuss the steps in developing Machine Learning.

(OR)

B) Discuss k-Nearest Neighbours classification algorithm.

10. A) How to construct a decision tree.

(OR)

B) What are the steps for plotting trees in Matplotlib.

11. A) What is Classification ? Discuss naïve Bayes classifier.

(OR)

B) What is Parsing? How to Parse data from RSS feeds.

12. A) Discuss classification with logistic regression and the sigmoid function.

(OR)

B) Discuss gradient descent optimization algorithm.

13. A) Comparing support vector machines with other classifiers.

(OR)

B) Discuss SMO algorithm for optimization

Semester-wise Revised Syllabus under CBCS, 2020-21
Four Year B.Sc. (Hons) - Semester – V (from 2022-23)
Subject: **B.Sc -Data Science**
Course-7B: **Unsupervised ML with Python**
(Skill Enhancement Course (Electi ve), 5 credits, Max Marks: 100 + 50

I Aim and objectives of Course (Unsupervised ML with Python):
Unsupervised Machine Learning involves finding patterns in datasets. The core of this course involves study of Clustering, feature extraction and optimization algorithms. The purpose of this course is to serve as an introduction to machine learning with Python.

Learning outcomes of Course:
- Able to do Clustering, feature extraction and optimization.
- Students will be able to understand and implement in Python algorithms of Unsupervised

Machine Learning and apply them to real-world datasets.

II Syllabus: (Total Hours: 90 including Teaching, Lab and internal exams, etc.)

UNIT- 1
Unsupervised Learning: Clustering: k-means clustering algorithm, Improving cluster performance with post processing, Bisecting k-means, Example: clustering points on a map
UNIT- II
Association analysis : Apriori algorithm: Association analysis, The Apriori principle, Finding frequent item sets with the Apriori algorithm, Mining association rules from frequent item sets, uncovering patterns in congressional voting
UNIT- III
Finding frequent item sets: FP-growth –FP trees, Build FP-tree, mining frequent from an FP-tree, finding co-occurring words in a Twitter feed, mining a click stream from a news site.
UNIT- IV
Principal component analysis: Dimensionality reduction techniques, using PCA to reduce the dimensionality of semiconductor manufacturing data
UNIT- V
Singular value decomposition: Applications of the SVD, Matrix factorization, SVD in Python, Collaborative filtering–based recommendation engines, a restaurant dish recommendation engine

III Text Books:

Machine learning in action, Peter Harrington by Manning publications

IV Lab Experiments

7B (L): Unsupervised ML with Python

1. Implementation of K-Means Clustering
2. Implement the bisecting k-means clustering algorithm
3. Implement Apriori algorithm
4. Implement Association rule-generation functions
5. Implement FP-tree creation
6. Write a function to find all paths ending with a given item.
7. Implement Code to access the Twitter Python library
8. Implement the PCA algorithm
9. Write a program to find Rating estimation by using the SVD
10. Implement Image-compression functions using SVD

Text Books:

Machine learning in action, Peter Harrington by Manning publications

MODEL QUESTION PAPER (Sem-end. Exam)

B. Sc DEGREE EXAMINATION

SEMESTER –V

Course 7B: Unsupervised ML with Python

Time:3Hrs **Max.marks:75**

Section – A

Answer any 5 Questions. Each Question Carries 5 marks **5 X 5 = 25**

1. What is Unsupervised Learning.
2. Define Clustering.
3. What is Associative analysis.
4. How to mine a click stream from a news site.
5. Explain mining frequent from an FP-tree
6. What are Dimensionality reduction techniques.
7. List Applications of the SVD.
8. Explain Matrix factorization.

Section – B

Answer all the questions. Each question carries 10 marks **5 X 10 = 50**

9. a) Discuss k-means clustering algorithm.
   (OR)
   b) How to improve cluster performance with post processing.

10. a) Explain Apriori algorithm along with its principles.
    (OR)
    b) Discuss Mining association rules from frequent item sets.

11. a) Define Finding frequent item sets: FP-growth –FP trees, Build FP-tree
    (OR)
    b) List out steps to find co-occurring words in a Twitter feed

12. a) Discuss Principal component analysis to reduce dimensionality.
    (OR)
    b) How PCA is used to reduce the dimensionality of semiconductor manufacturing data

13. a) Discuss how Singular value decomposition(SVD) is implemented in Python.
    (OR)
    b) Discuss Collaborative filtering–based recommendation engines.

Four Year B.Sc. (Hons) - Semester – V (from 2022-23)

Subject: **B.Sc -Data Science**

Course-6C: **NLP with Python**

(Skill Enhancement Course (Elective), 5 credits, Max Marks: 100 + 50

Objectives of Course:

This course introduces the fundamental concepts and techniques of natural language processing (NLP). Students will gain an in-depth understanding of the computational properties of natural languages and the commonly used algorithms for processing linguistic information. The course examines NLP models and algorithms using both the traditional symbolic and the more recent statistical approaches.

Learning outcomes of Course:

- Able to describe the fundamental concepts and techniques of natural language processing.
- Ability to distinguish among the various techniques, taking into account the assumptions, strengths, and weaknesses of each.
- Use appropriate descriptions, visualizations, and statistics to communicate the problems and their solutions.
- Analyze large volume text data generated from a range of real-world applications.
- Understanding semantics and pragmatics of English language for processing
- Writing programs in Python to carry out natural language processing

II. Syllabus: ***(Total Hours: 90 including Teaching, Lab and internal exams, etc.)***

UNIT I

Natural Language Processing: What is NLP? NLP and linguistics -Syntax and semantics, Pragmatics and context, Two views of NLP, Tasks and super tasks. Linguistic tools- Sentence delimiters and tokenizers, Stemmers and taggers, Noun phrase and name recognizers, Parsers and grammars.

UNIT II

Document Retrieval: Information retrieval, Indexing technology Query processing: Boolean search, Ranked retrieval, Probabilistic retrieval, Language modeling Evaluating search engines: Evaluation studies Evaluation Metrics Relevance Judgments Total system evaluation

Attempts to enhance search performance: Table of contents Query expansion and thesauri, Query expansion from relevance information

UNIT III

Information extraction: The Message Understanding Conferences, Regular expressions Finite automata in FASTUS: Finite State Machines and regular languages, Finite State Machines as parsers Pushdown automata and context-free grammars: Analyzing case reports

Context free grammars Parsing with a pushdown automaton, Coping with incompleteness and ambiguity

UNIT IV

Text categorization: Overview of categorization tasks and methods , Handcrafted rule based methods Inductive learning for text classification : Naïve Bayes classifiers , Linear classifiers, Decision trees and decision lists Nearest Neighbor algorithms Combining classifiers : Data fusion, Boosting, Using multiple classifiers

UNIT V

Text mining: What is text mining? Reference and coreference, Named entity recognition, The coreference task, Automatic summarization: Summarization tasks, Constructing summaries from document fragments, Multi-document summarization (MDS) Testing of automatic summarization programs: Evaluation problems in summarization research, Building a corpus for training and testing.

III Text Books:

1. Natural Language Processing for Online Applications, Text Retrieval Extraction & Categorization. Peter Jackson, Isabelle Moulinier, Thomson Legal & Regulatory

IV List of Experiments
6C(L): NLP with Python Lab

1. INSTALLATION
2. WORD TOKENIZER
3. SENTENCE TOKENIZER
4. PARAGRAPH TOKENIZER
5. PROBABILISTIC PARSING
6. PROBABILISTIC CONTEXT FREE GRAMMER
7. LEARNING GRAMMAR
8. CONDITIONAL FREQUENCY DISTRIBUTIONS
9. LEXICAL ANALYSER
10. WORDNET
11. CONTEXT FREE GRAMMAR
12. LARGE CONTEXT FREE GRAMMAR AND PARSING
13. NAMED ENTITY RECOGNITION

TEXT BOOKS:

1. Natural Language with Python, Steven Bird and Oreilly , First Edition.

MODEL QUESTION PAPER (Sem-end. Exam)

B. Sc DEGREE EXAMINATION

SEMESTER –V

Course 6C: NLP with Python

Time:3Hrs **Max.marks:75**

Section – A

Answer any 5 Questions. Each Question Carries 5 marks **5 X 5 = 25**

1. What is NLP? Explain its syntax and semantics.
2. Discuss two views of NLP.
3. Explain how information is retrieved.
4. Discuss Finite State Machines.
5. Discuss Parsing with Pushdown Automata.
6. What are Handcrafted rule based methods.
7. What is Text mining? Explain.
8. Discuss Multi-document summarization (MDS).

Section – B

Answer all the questions. Each question carries 10 marks **5 X 10 = 50**

9. a) Discuss Linguistic tools in detail.

(OR)

b) What are the existing Parsers and grammars in NLP? Explain.

10. a) Explain methods in Indexing Technology Query processing.

(OR)

b) Discuss in detail about Language modeling Evaluating search engines.

11 a) Finite State Machines as parsers Pushdown automata Discuss.

(OR)

b) What is Parsing? Explain Context free grammars Parsing with a pushdown automaton.

12. a) Discuss Text categorization tasks and methods.

(OR)

b) What is Naive Bayes algorithm? When we can use this algorithm in NLP?

13. a) Discuss the tasks involved in Automatic summarization.

(OR)

b) How Testing of automatic summarization programs done explain.

Semester-wise Revised Syllabus under CBCS, 2020-21
Four Year B.Sc. (Hons) - Semester – V (from 2022-23)
Subject: **B.Sc -Data Science**
Course-7C: **Deep Learning Neural Networks with Python**
(Skill Enhancement Course (Elective), 5 credits, Max Marks: 100 + 50

I Aim and Objectives of Course:

Deep learning has resurged with the availability of massive datasets and affordable computing, enabling new applications in computer vision and natural language processing. This course introduces convolutional, recurrent, and other neural network architectures for deep learning. Students design, implement, and train these models to solve real-world problems**.**

Learning outcomes of Course:

- Solve problems in linear algebra, probability, optimization, and machine learning.
- The advantages and disadvantages of deep learning neural network architectures and other approaches.
- Implement deep learning models in Python using the PyTorch library and train them with real-world datasets.
- Design convolution networks for handwriting and object classification from images or video.
- Design recurrent neural networks with attention mechanisms for natural language classification, generation, and translation.

II Syllabus: (Total Hours: 90 including Teaching, Lab and internal exams, etc.)

UNIT I

Introduction to Deep Learning: Artificial intelligence, machine learning and deep learning, history of machine learning, Why deep learning? Why now?
The mathematical building blocks of neural networks: A first look at a neural network, Data representations for neural networks, The gears of neural networks: tensor operations, The engine of neural networks: gradient-based optimization.

UNIT II

Getting started with neural networks: Anatomy of a neural network, Introduction to Keras, Setting up a deep-learning workstation, Classifying movie reviews: a binary classification
Example, Classifying newswires: a multiclass classification example, Predicting house prices: a regression example.
Fundamentals of machine learning: Four branches of machine learning, Evaluating machine-learning models, Data preprocessing, feature engineering and feature learning, Overfitting and underfitting, The universal workflow of machine learning.

UNIT III

Deep learning for computer vision: Introduction to convnets, Training a convnet from scratch on a small dataset, Using a pretrained convent, Visualizing what convnets learn.

UNIT IV

Deep learning for text and sequences: Working with text data, Understanding recurrent neural networks, Advanced use of recurrent neural networks, Sequence processing with convnets.

UNIT V

Advanced deep-learning best practices: Going beyond the Sequential model: theKeras functional API, Inspecting and monitoring deep-learning models using Keras callbacks and TensorBoard, Getting the most out of your models.

III Text Books:

1. "Deep Learning with Python" by Francois Chollet, , 2018 Edition, Manning Publications.
2. "Deep Learning with Python" by Nikhil Ketkar, JojoMoolayil, Second Edition, Apress.
3. "Python Deep Learning" by Ivan Vasilev, Daniel Slatter, Second Edition, Packt Publications.

IV List of Deep Learning Programs

7C (L): DEEP LEARNING NEURAL NETWORKS WITH PYTHON LAB

1. How to train a network using Keras in Python
2. Write programs to demonstrate Tensor Operations
3. Classifying movie reviews: a binary classification example
4. Predicting house prices: a regression example
5. Demonstrate Convnets by the following tasks

i. Instantiating a Convnet

ii. Adding classifier on top of the Convnet

iii. Training the Convnet on MNIST images

6. Display curves of loss and accuracy during training
7. Word level one-hot encoding (Toy example)
8. Character level one-hot encoding (Toy example)
9. Using Keras for Word level one-hot encoding
10. Word level one-hot encoding with hashing trick

Text Books:

1. "Deep Learning with Python" by Francois Chollet, , 2018 Edition, Manning Publications.

MODEL QUESTION PAPER (Sem-end. Exam)

B. Sc DEGREE EXAMINATION

SEMESTER –V

Course 7C: Deep Learning Neural Networks with Python

Time:3Hrs **Max.marks:75**

Section – A

Answer any 5 Questions. Each Question Carries 5 marks **5 X 5 = 25**

1. What is Machine Learning?
2. Write about the relationship between network, layers, loss function and optimizer.
3. Explain max poling operation.
4. Explain about word-level one-hot encoding with example.
5. Write about multi input model.
6. What are the tensor operations?
7. Write about feature engineering for reading the time on a clock.
8. Write how a bidirectional RNN works.

Section – B

Answer all the questions. Each question carries 10 marks **5 X 10 = 50**

10. a) Explain how deep learning works in three figures.

(OR)

b) Explain about Data representations for neural networks.

11. a) Explain about binary classification example.

(OR)

b) Explain about Four branches of machine learning

12. a) Finite about Data preprocessing.

(OR)

b) Explain how to plot the results with an example.

13. a) Explain about LSTM and GRU layers.

(OR)

b) Explain about Combining CNNs and RNNs to process long sequences.

14. a)Explain about Directed acyclic graphs of layer

(OR)

b) Explain about TensorFlow visualization framework