Chhattisgarh Swami Vivekananda Technical University, Bhilai (C.G.)

Branch: B Tech Honours (Data Science) **Semester: V**

Subject: Pattern Recognition and Machine Learning **Subject Code: C128571(022)**

Total Theory Periods: 40 **Total Tutorial Periods: 10**

No. of Class tests to: 2 (Minimum) **No. of Assignments to be submitted: One per Unit**

ESE Duration: Three Hours, Maximum Marks in ESE: 100 **Minimum Marks in ESE: 35**

UNIT-I: Probability Theory and Statistics

Probability theory, Bayesian probabilities, Probability densities, Skewness and Kurtosis, Expectations and Covariances, Multivariate Gaussian Distribution, Combination of Random Variables; Moment Estimation, Estimating the mean and variance, Law of Large Numbers; Measure Theory, Multiple Dimensional Spaces, Metric Space, Normed Vector Space, Dot Product Space, Pre-Hilbert and Hilbert; Decision Theory, Minimizing the misclassification rate, Minimizing the expected loss, Inference and Decision, Loss functions for regression.

UNIT-II: Parameter Estimation

Maximum Likelihood Estimation, Maximum A-Posteriori (MAP) Estimation, Maximum Entropy Estimation, Minimum Relative Entropy Estimation, Maximum Mutual Information Estimation (MMIE); Model Selection, Akaike Information Criterion (AIC)Bayesian Information Criterion (BIC).

UNIT-III: Parameter Classification

Linear Models for Classification, Discriminant Functions, Two classes, Multiple classes, Least squares for classification, Fisher's linear discriminant, Relation to least squares, Fisher's discriminant for multiple classes, The perceptron algorithm; Probabilistic Generative Models, Continuous inputs, Maximum likelihood solution, Discrete features, Exponential family; Probabilistic Discriminative Models.

UNIT-IV: Clustering and Learning

Learning Algorithms, Risk Minimization, Empirical Risk Minimization, Capacity and Bounds on Risk, Structural Risk Minimization; Decision and Regression Trees, Vector Quantization (VQ); Basic Clustering Techniques, Standard k-Means (Lloyd) Algorithm, Generalized Clustering, Merging, Modifications to the k-Means Algorithm, k-Means Wrappers, Rough k-Means, Fuzzy k-Means, k-Harmonic Means Algorithm, Hybrid Clustering Algorithms; Estimation using Incomplete Data, Expectation Maximization (EM); Semi-Supervised Learning.

UNIT-V: Kernel Methods and Support Vector Machines

The Two-Class Problem, Dual Representation, Soft Margin Classification; Origins of Kernel methods, Kernel Mapping, The Kernel Trick; Constructing Kernels, Support Vector Machines: Formulation and Computation; Radial Basis Function Networks; Positive Semi-Definite Kernels, Linear Kernel, Polynomial Kernel, Gaussian Radial Basis Function (GRBF) Kernel, Cosine Kernel, Fisher Kernel, GLDS Kernel, GMM-UBM Mean Interval (GUMI) Kernel.

TEXT BOOKS:

1. Kevin P. Murphy, "Machine Learning: A Probabilistic Perspective", MIT Press, 2012.
2. Stephen Marsland, "Machine Learning –An Algorithmic Perspective", CRC Press, 2009.
3. SaikatDutt, Subramanian Chandramouli, Amit Kumar Das, "Machine Learning", Pearson Education, 2018.
4. Christopher Bishop, "Pattern Recognition and Machine Learning" Springer, 2011.

REFERENCE BOOKS :

1. Andreas C. Muller, “Introduction to Machine Learning with Python: A Guide for Data Scientists”, O'Reilly, 2016.
2. Sebastian Raschka, “Python Machine Learning”, Packt Publishing, 2015.
3. Hastie, Tibshirani, Friedman, “The Elements of Statistical Learning: Data Mining, Inference, andPrediction**”**,2nd Edition, Springer, 2017.
4. EthemAlpaydin, “Introduction to Machine Learning”, 2nd Revised edition, MIT Press,2010.
5. Pattern Recognition and Machine Learning, Christopher M. Bishop, Springer.
6. Machine Learning with SVM and other Kernel methods, K.P. SomanR.Loganathan, V.Ajay, PHI Learning Private Limited.\
7. Machine Learning, Tom Mitchell, McGraw Hill, 1997.
8. Fundamentals of Speaker Recognition, HomayoonBeigi, Springer.

E BOOKS:

1. https://www.ibm.com/downloads/cas/GB8ZMQZ3

MOOC:

1. https://www.edx.org/course/machine-learning-fundamentals-2
2. https://www.coursera.org/learn/machine-learning

Chhattisgarh Swami Vivekananda Technical University, Bhilai (C.G.)

Branch: B Tech Honours (Data Science) **Semester: V**
Subject: Intelligent Data Analysis **Subject Code: C128572(022)**
Total Theory Periods: 40 **Total Tutorial Periods: 10**
No. of Class tests to: 2 (Minimum) **No. of Assignments to be submitted: One per Unit**
ESE Duration: Three Hours, Maximum Marks in ESE: 100 **Minimum Marks in ESE: 35**

UNIT-I:

Introduction: Data Mining and Analysis, Challenges, Types of Data and Patterns, Application, Understanding the Data - Measuring the Central Tendency: Mean, Median, and Mode, Measuring the Dispersion of Data: Range, Quartiles, Variance, Standard Deviation, Data Visualization Techniques.

Data Pre-Processing: Data Objects and Attribute Types, Measuring Data Similarity and Dissimilarity, Why Pre-process the Data, Data Cleaning, Data Integration, Data Reduction Data Transformation and Data Discretization.

UNIT-II:

Classification: Basic Concepts, Decision Trees, Perceptrons, Neural Networks, Rule-based Classifier, Nearest-Neighbor Classifier , Bayesian Classifier, Principles of Modelling: Train, Test, And Cross-Validation.

UNIT-III:

Association Analysis: Frequent Itemset Mining - Apriori Algorithm, Rule generation, FP-Growth Algorithm, Compact Representation of Frequent Itemsets – Maximal Frequent Itemset and Closed Frequent Itemsets.

UNIT-IV:

Clustering: Cluster Analysis, Partitioning Methods -k-Means and k-Medoids, Hierarchical Methods - Agglomerative versus Divisive Hierarchical Clustering Distance Measures in Algorithmic Methods, BIRCH, Density-Based Method - DBSCAN, Grid-Based Methods STING, Evaluation of Clustering.

UNIT-V:

Anomaly Detection: Types of Anomalies, Challenges of Anomaly Detection, Type of Anomaly Detection Methods, Statistical Methods, Proximity-based Methods, and Clustering-based Methods, Classification-based Methods.

Text Books:

1. Pang-Ning Tan, Michael Steinbach, and Vipin Kumar, "Introduction to data mining" Pearson Education India, 2016.
2. Jiawei Han, Micheline Kamber and Jian Pei, "Data Mining: Concepts and Techniques, 3rd ed., Morgan Kaufmann Publishers, 2011
3. Christopher M. Bishop, "Pattern Recognition and Machine Learning", Springer Science+ Business Media, LLC, 2006.

Chhattisgarh Swami Vivekananda Technical University, Bhilai (C.G.)

Branch: B Tech Honours (Data Science) **Semester: V**
Subject: Cryptography and Network Security **Subject Code: C127573(022)**
Total Theory Periods: 40 **Total Tutorial Periods: 10**
No. of Class tests to: 2 (Minimum) **No. of Assignments to be submitted: One per Unit**
ESE Duration: Three Hours, Maximum Marks in ESE: 100 **Minimum Marks in ESE: 35**

UNIT–I: CLASSICAL ENCRYPTION TECHNIQUES
Symmetric cipher model, Steganography, Limitations of Perfect Secrecy, Shannon's Theorem.

UNIT–II: NUMBER THEORY
Prime numbers and factoring, modular arithmetic, computations in finite fields, Discrete logarithms.

UNIT–III: PRIVATE KEY ENCRYPTION SCHEMES AND BLOCK CIPHERS
Pseudorandom Functions and Permutations, Private-Key Encryption Schemes, DES – The Data Encryption Standard, Attacks on DES, Single-Round DES, Two-Round DES, Three-Round DES, Best Known Attacks on Increasing the Key size for DES, Modes of Operation.

UNIT–IV: PUBLIC KEY ASYMMETRIC CRYPTOGRAPHY
Public-Key Problems and Mathematical Background, Diffie-Hellman Key Agreement, El-Gamal Encryption Scheme, RSA Encryption, Security of RSA, Hybrid Encryption , Attacks on RSA, Private and Public-Key, , Timing Attacks, Elliptic Curve Cryptography.

UNIT–V: HASH FUNTIONS
Definition and Properties, Constructions of Collision-Resistant Hash Functions, Popular Uses of Collision-Resistant Hash Functions, Random Oracle Model. Hash algorithms: SHA-512. Message Authentication: Message Authentication Codes Definitions, Constructions of Secure Message Authenticate Codes, Practical Constructions of Message Authentication Codes.

UNIT–VI:DIGITAL SIGNATURE AND APPLICATIONS AND SECURITY AT TRANSPORT LAYER

Definitions, and Public-Key Infrastructure, Combining Encryption and Signatures – SignCryption, SSL Architecture, Protocols, SSL Message formats, Transport Layer Security, Kerberos Password Management.

TEXT BOOKS:

1. "Cryptography & Network Security" by William Stallings 4th Edition, 2006,Pearson Education Asia.
2. Kahate A, "Cryptography & Network Security", Tata McGraw Hill, 2004.

REFERENCE BOOKS:

1. "Applied Cryptology" by Schiner Bruce, John Wiley & Sons,2001.
2. "Introduction to Cryptography with Coding Theory" by Wade Trappe & Lawrence C Washington, New Jersey, Pearson Education,2006.
3. Charlie Kaufman, Radia Perlman and Mike Speciner, "Network Security: Private Communication in a Public World", Prentice Hall of India PrivateLimited.
4. Behrouz A. Forouzan, "Cryptography and Network Security", McGrawHill

Chhattisgarh Swami Vivekananda Technical University, Bhilai (C.G.)

Branch: B Tech Honours (Data Science) **Semester: V**
Subject: Natural Language Processing **Subject Code: C128574(022)**
Total Theory Periods: 40 **Total Tutorial Periods: 10**
No. of Class tests to: 2 (Minimum) **No. of Assignments to be submitted: One per Unit**
ESE Duration: Three Hours, Maximum Marks in ESE: 100 **Minimum Marks in ESE: 35**

UNIT-I: Introduction
Definition, History, Applications of NLP, Goals of NLP.

UNIT-II: Words and Phonetics
Regular expressions and Automata, Morphology and phonetics fundamentals, morphological diversity of Indian languages, morphology paradigms, finite state machine based morphology, Computational Phonology and Text-to-Speech, Probabilistic Models of Pronunciation and Spelling, N-grams, HMMs and Speech Recognition, Wordnet and linking.

UNIT-III: Parsing
Part-of Speech Tagging, theories of parsing, syntactic and statistical parsing, parsing algorithms, hybrid of rule based and probabilistic parsing, scope ambiguity and attachment ambiguity resolution, Tree banks.

UNIT-IV: Discourse and dialogue
discourse and dialogue analysis, anaphora resolution, named entity resolution, event anaphora, Information extraction and retrieval.

UNIT-V: Applications
sentiment analysis, text entailment, machine translation, automated speech recognition systems, question-answering based systems, shallow parsers.

TEXT BOOKS:
1. Dan Jurafsky and James H. Martin. Speech and Language Processing (3rd ed. draft), 2019.

REFERENCES BOOK:
1. Jurafsky, D. & J. Martin, "Speech and Language Processing: An Introduction to Natural Language Processing Computational Linguistics, and Speech Recognition" Prentice Hall, 2000.
2. Grosz, B.J., Sparck Jones, K. & Webber, B.L. (eds) "Readings in natural language processing", Los Altos, CA. Morgan Kaufmann, 1986.
3. Allen, J., "Natural Language Understanding", Redwood City, CA: 1994. Benjamin/Cummings.
4. Bharti, Akshar,ChaitanyaVineet, Sangal Rajeev, "Natural Language Processing", Prentice Hall.
5. Steven Bird, Ewan Klein, and Edward Loper, Natural Language Processing with Python, First Edition, O'reilly, 2009
6. Yoav Goldberg, University of Toronto, Neural Network Methods for Natural language Processing, Morgan & Claypool, 2017
7. Christopher D. Manning, and HinrichSchutze. Foundations of statistical natural language processing. First Edition, MIT press, 1999

E BOOKS:
1. https://www.cs.vassar.edu/~cs366/docs/Manning_Schuetze_StatisticalNLP.pdf
2. https://www.nltk.org/book/
3. https://www.nltk.org/genindex.html

MOOC:
1. https://www.coursera.org/learn/language-processing

Chhattisgarh Swami Vivekananda Technical University, Bhilai (C.G.)

Branch: B Tech Honours (Data Science) **Semester: V**
Subject: Advanced Computer Network **Subject Code: C127531(022)**
Total Theory Periods: 40 **Total Tutorial Periods: 10**
No. of Class tests to: 2 (Minimum) **No. of Assignments to be submitted: One per Unit**
ESE Duration: Three Hours, Maximum Marks in ESE: 100 **Minimum Marks in ESE: 35**

UNIT-I:
Computer Networks and the Internet: What is the Internet, The Network edge, The Network core, Access Networks and Physical media, ISPs and Internet Backbones, Delay and Loss in Packet-Switched Networks, History of Computer Networking and the Internet - Foundation of Networking Protocols: 5-layer TCP/IP Model, 7-Layer OSI Model, Internet Protocols and Addressing, Equal-Sized Packets Model: ATM - Networking Devices: Multiplexers, Modems and Internet Access Devices, Switching and Routing Devices, Router Structure.

UNIT-II:
The Link Layer and Local Area Networks: Link Layer: Introduction and Services, Error- Detection and Error-Correction techniques, Multiple Access Protocols, Link Layer Addressing, Ethernet, Interconnections: Hubs and Switches, PPP: The Point-to-Point Protocol, Link Virtualization, Routing and Internetworking: Network Layer Routing, Least-Cost-Path algorithms, Non-Least-Cost-Path algorithms, Intradomain Routing Protocols, Interdomain Routing Protocols, Congestion Control at Network Layer.

UNIT-III:
Logical Addressing: IPv4 Addresses, IPv6 Addresses - Internet Protocol: Internetworking, IPv4, IPv6, Transition from IPv4 to IPv6, Multicasting Techniques and Protocols: Basic Definitions and Techniques, Intradomain Multicast Protocols, Interdomain Multicast Protocols, Node-Level Multicast algorithms, Transport and End-to-End Protocols: Transport Layer, Transmission Control Protocol (TCP), User Datagram Protocol (UDP), Mobile Transport Protocols, TCP Congestion Control, Application Layer: Principles of Network Applications, The Web and HTTP, File Transfer: FTP, Electronic Mail in the Internet, Domain Name System (DNS), P2P File Sharing, Socket Programming with TCP and UDP, Building a Simple Web Server.

UNIT-IV:
Wireless Networks and Mobile IP: Infrastructure of Wireless Networks, Wireless LAN Technologies, IEEE 802.11 Wireless Standard, Cellular Networks, Mobile IP, Wireless Mesh Networks (WMNs) - Optical Networks and WDM Systems: Overview of Optical Networks, Basic Optical Networking Devices, Large-Scale Optical Switches, Optical Routers, Wavelength Allocation in Networks, Case Study: An All-Optical Switch.

UNIT-V:
VPNs, Tunneling and Overlay Networks: Virtual Private Networks (VPNs), Multiprotocol Label Switching (MPLS), Overlay Networks, VoIP and Multimedia Networking: Overview of IP Telephony, VoIP Signaling Protocols, Real-Time Media Transport Protocols, Distributed Multimedia Networking, Stream Control Transmission Protocol - Mobile A-Hoc Networks: Overview of Wireless Ad-Hoc Networks, Routing in Ad-Hoc Networks, Routing Protocols for Ad-Hoc Networks, Wireless Sensor Networks: Sensor Networks and Protocol Structures, Communication Energy Model, Clustering Protocols, Routing Protocols.

TEXT BOOKS:
1. Computer Networking: A Top-Down Approach Featuring the Internet, *James F. Kurose, Keith W.Ross*, Third Edition, Pearson Education, 2007
2. Computer and Communication Networks, *Nader F. Mir,* Pearson Education, 2007.

REFERENCES:

1. Data Communications and Networking, *Behrouz A. Forouzan*, Fourth Edition, Tata McGraw Hill, 2007
2. Guide to Networking Essentials, *Greg Tomsho,Ed Tittel, David Johnson,*Fifth Edition, Thomson.
3. An Engineering Approach to Computer Networking *, S.Keshav,* Pearson Education.
4. Campus Network Design Fundamentals, *Diane Teare, Catherine Paquet,* Pearson Education (CISCO Press)
5. Computer Networks, *Andrew S. Tanenbaum*, Fourth Edition, Prentice Hall.
6. The Internet and Its Protocols, A. Farrel, Elsevier.

Chhattisgarh Swami Vivekananda Technical University, Bhilai (C.G.)

Branch: B Tech Honours (Data Science) **Semester: V**
Subject: Computational Complexity **Subject Code: C127532(022)**
Total Theory Periods: 40 **Total Tutorial Periods: 10**
No. of Class tests to: 2 (Minimum) **No. of Assignments to be submitted: One per Unit**
ESE Duration: Three Hours, Maximum Marks in ESE: 100 **Minimum Marks in ESE: 35**

UNIT – I:
Computational Complexity: Polynomial time and its justification, Nontrivial examples of polynomial-time algorithms, the concept of reduction (reducibility), Class P Class NP and NP- Completeness, The P versus NP problem and why it's hard

UNIT - II:
Algorithmic paradigms: Dynamic Programming, Longest common subsequence, matrix chain multiplication, knapsack problem, Greedy – 0-1 knapsack, fractional knapsack, scheduling problem, Huffman coding, MST, Branch-and-bound – travelling sales person problem, 0/1 knapsack problem, Divide and Conquer, Merge sort, binary search, quick sort.

UNIT - III:
Randomized Algorithms: Finger Printing, Pattern Matching, Graph Problems, Algebraic Methods, Probabilistic Primality Testing, De-Randomization Advanced Algorithms.

UNIT - IV:
Graph Algorithms: Shortest paths, Flow networks, Spanning Trees; Approximation algorithms, Randomized algorithms. Approximation algorithms: Polynomial Time Approximation Schemes.

UNIT - V:
Advanced Data Structures and applications: Decision Trees and Circuits, B-Trees, AVL Trees, Red and Black trees, Dictionaries and tries, Maps, Binomial Heaps, Fibonacci Heaps, Disjoint sets, Union by Rank and Path Compression

TEXT BOOKS:

1. T. Cormen, C. Leiserson, R. Rivest and C. Stein, Introduction to Algorithms, Third Edition, McGraw-Hill, 2009.
2. R. Motwani and P. Raghavan, Randomized Algorithms, Cambridge University Press, 1995.
3. J. J. McConnell, Analysis of Algorithms: An Active Learning Approach, Jones & Bartlett Publishers, 2001.
4. D. E. Knuth, Art of Computer Programming, Volume 3, Sorting and Searching, Second Edition, Addison-Wesley Professional, 1998.
5. S. Dasgupta, C. H. Papadimitriou and U. V. Vazirani, Algorithms, McGraw-Hill, 2008.

Chhattisgarh Swami Vivekananda Technical University, Bhilai (C.G.)

Branch: B Tech Honours (Data Science) **Semester: V**
Subject: Distributed Computing **Subject Code: C127533(022)**
Total Theory Periods: 40 **Total Tutorial Periods: 10**
No. of Class tests to: 2 (Minimum) **No. of Assignments to be submitted: One per Unit**
ESE Duration: Three Hours, Maximum Marks in ESE: 100 **Minimum Marks in ESE: 35**

UNIT-I: INTRODUCTION
Introduction: Definition-Relation to Computer System Components, Motivation, Message, Passing Systems versus Shared Memory Systems, Primitives for Distributed Communication, Synchronous versus Asynchronous Executions, Design Issues and Challenges; A Model of Distributed Computations: A Distributed Program, A Model of Distributed Executions, Models of Communication Networks, Global State of a Distributed System.

UNIT-II: LOGICAL TIME AND GLOBAL STATE
Logical Time: Physical Clock Synchronization: NTP, A Framework for a System of Logical Clocks, Scalar Time, Vector Time; Message Ordering and Group Communication: Message Ordering.
Paradigms, Asynchronous Execution with Synchronous Communication, Synchronous Program Order on Asynchronous System, Group Communication, Causal Order, Total Order; Global State and Snapshot Recording Algorithms: Introduction – System Model and Definitions, Snapshot. Algorithms for FIFO Channels.

UNIT-III: DISTRIBUTED MUTEX AND DEADLOCK
Distributed Mutual exclusion Algorithms: Introduction, Preliminaries, Lamport's algorithm, Ricart Agrawala's Algorithm, Token-Based Algorithms, Suzuki-Kasami's Broadcast Algorithm; Deadlock Detection in Distributed Systems: Introduction, System Model, Preliminaries Models of Deadlocks, Chandy-Misra-Haas Algorithm for the AND model and OR Model.

UNIT-IV: CONSENSUS AND RECOVERY
Consensus and Agreement Algorithms: Problem Definition, Overview of Results, Agreement in a Failure-Free System (Synchronous and Asynchronous), Agreement in Synchronous Systems with Failures; Check pointing and Rollback Recovery: Introduction, Background and Definitions, Issues in Failure Recovery, Checkpoint-based Recovery, Coordinated Check pointing Algorithm, Algorithm for Asynchronous Check pointing and Recovery

UNIT-V: CLOUD COMPUTING
Definition of Cloud Computing, Characteristics of Cloud, Cloud Deployment Models, Cloud Service Models, Driving Factors and Challenges of Cloud, Virtualization, Load Balancing, Scalability and Elasticity, Replication, Monitoring, Cloud Services and Platforms: Compute Services, Storage Services, Application Services

TEXT BOOKS:

1. Kshemkalyani Ajay D, Mukesh Singhal, "Distributed Computing: Principles, Algorithms and Systems", Cambridge Press, 2011.
2. Mukesh Singhal, Niranjan G Shivaratri, "Advanced Concepts in Operating systems", McGraw Hill Publishers, 1994.

REFERENCES:

1. George Coulouris, Jean Dollimore, Time Kindberg, “Distributed Systems Concepts and Design”, Fifth Edition, Pearson Education, 2012.
2. Pradeep L Sinha, “Distributed Operating Systems: Concepts and Design”, Prentice Hall of India, 2007.
3. Tanenbaum A S, Van Steen M, “Distributed Systems: Principles and Paradigms”, Pearson Education, 2007.
4. Liu M L, “Distributed Computing: Principles and Applications”, Pearson Education, 2004.
5. Nancy A Lynch, “Distributed Algorithms”, Morgan Kaufman Publishers, 2003.
6. Arshdeep Bagga, Vijay Madisetti, “ Cloud Computing: A Hands-On Approach”, Universities Press, 2014.

Chhattisgarh Swami Vivekananda Technical University, Bhilai (C.G.)

Branch: B Tech Honours (Data Science) **Semester: V**

Subject: Pattern Recognition and Machine Learning Lab **Subject Code: C128591(022)**

Maximum Marks in ESE: 40

LIST OF EXPERIMENTS :

1. Installation of Python Libraries/ MATLAB tools for Machine Learning
2. Data pre-processing using Python Machine Learning libraries/ MATLAB.
3. Design a model to predict the housing price from Boston Dataset using Multivariate Linear Regression.
4. Build a classifier using Logistic Regression, k- Nearest Neighbor / Decision Tree to classify whether the given user will purchase a product or not from a social networking dataset.
5. Segment a customer dataset based on the buying behaviour of customers using K- means/Hierarchical clustering.
6. Dimensionality reduction of any CSV/image dataset using Principal Component Analysis.
7. Recognition of MNIST handwritten digits using Artificial Neural Network.
8. Build an email spam classifier using SVM.
9. Classify the given text segment as ‘Positive’ or ‘Negative’ statement using the Naive Bayes Classifier.
10. Predict future stock price of a company using Monte Carlo Simulation.

Chhattisgarh Swami Vivekananda Technical University, Bhilai (C.G.)

Branch: B Tech Honours (Data Science) **Semester: V**

Subject: Intelligent Data Analysis Lab **Subject Code: C128592(022)**

Maximum Marks in ESE: 40

LIST OF EXPERIMENTS:

Note – Datasets can be selected from UCI Machine Learning Repository and/or kaggle dataset repository for different experiments

1. Demonstration of different pre-processing techniques including missing value handling and data discretization on Adult dataset.
2. Demonstration of data reduction techniques including PCA and Histogram on Predict students' dropout and academic success dataset.
3. Demonstration of classification rules process on dataset of your choice using ID3 and J48 algorithm in Weka explorer.
4. Implement the classification rules process on dataset of your choice using Naive Baye's algorithm in Weka explorer.
5. Build a Neural Network model to predict whether tumor is malignant or benign for Breast Cancer Wisconsin (Diagnostic) dataset using Python.
6. Demonstration of clustering on dataset of your choice using simple K-means algorithm in weka explorer and python.
7. Demonstration of clustering on dataset of your choice using simple DBSCAN algorithm in weka explorer and python.
8. Demonstration of clustering on dataset of your choice using simple BIRCH algorithm in weka explorer and python.
9. Demonstration of association rule generation on Groceries dataset for Market Basket Analysis using Apriori algorithm in weka explorer.
10. Perform comparative analysis of Apriori and FP-Growth algorithms on Market Basket Analysis using Python.

Chhattisgarh Swami Vivekananda Technical University, Bhilai (C.G.)

Branch: B Tech Honours (Data Science) **Semester: V**

Subject: Natural Language Processing (Lab) **Subject Code: C128593(022)**

Maximum Marks in ESE: 40

List of Experiments:-

1. Convert the text into tokens
2. Find the word frequency
3. Demonstrate a bigram language model
4. Demonstrate a trigram language model
5. Generate regular expression for a given text
6. Perform Lemmatization
7. Perform Stemming
8. Identify parts-of Speech using Penn Treebank tag set.
9. Implement HMM for POS tagging
10. Build a Chunker
11. Find the synonym of a word using WordNet
12. Find the antonym of a word
13. Implement semantic role labeling to identify named entities
14. Resolve the ambiguity
15. Translate the text using First-order logic
16. Implement RNN for sequence labeling
17. Implement POS tagging using LSTM
18. Implement Named Entity Recognizer
19. Word sense disambiguation by LSTM/GRU
20. Develop a Movie review system
21. Create a chatbot for HITS**.**