**GUJARAT TECHNOLOGICAL UNIVERSITY (GTU)**

**Competency-focused Outcome-based Green Curriculum-2021 (COGC-2021)**
Semester -V

**Course Title: Foundation of AI and ML**
(Course Code: 4351601)

| Diploma programme in which this course is offered | Semester in which offered |
|---|---|
| Information Technology | 5th semester |

## 1. RATIONALE

Artificial Intelligence (AI) and Machine Learning (ML) technologies have immense growth potential and are expected to shape the future of various fields. Understanding AI and ML gives learners a competitive edge in today's era. This course provides the foundation concepts of AI and ML. It covers topics such as Fundamentals of AI, Types of machine learning, Artificial Neural Networks (ANN) and Natural Language Processing (NLP). ANN and NLP is an integral part of the broader field of Artificial Intelligence (AI). This helps students to develop basic models of neural network to solve real world problems.

## 2. COMPETENCY

The purpose of this course is to help the student to attain the following industry identified competency through various teaching-learning experiences:

- Apply Artificial Intelligence and Machine Learning concepts to solve real world problems.

## 3. COURSE OUTCOMES (COs)

The practical exercises, the underpinning knowledge, and the relevant soft skills associated with this competency are to be developed in the student to display the following COs:

The student will develop underpinning knowledge, adequate programming skills of competency for implementing various applications using python programming language to attain the following course outcomes.

a) Understand fundamental principles of Artificial Intelligence.
b) Compare types of machine learning.
c) Build a simple Neural Network model to solve real world problem.
d) Apply data preprocessing on text/paragraph using NLTK library.
e) Demonstrate word embedding techniques to develop real world NLP applications.

## 4. TEACHING AND EXAMINATION SCHEME

| Teaching Scheme (In Hours) | | | Total Credits (L+T/2+P/2) | Examination Scheme | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | Theory Marks | | Practical Marks | | Total Marks |
| L | T | P | C | CA | ESE | CA | ESE | |
| 3 | - | 2 | 4 | 30 | 70 | 25 | 25 | 150 |

*(*): Out of 30 marks under the theory CA, 10 marks are for assessment of the micro-project to facilitate integration of COs and the remaining 20 marks is the average of 2 tests to be taken during the semester for the assessing the attainment of the cognitive domain UOs required for the attainment of the COs.*
***Legends: L**-Lecture; **T** – Tutorial/Teacher Guided Theory Practice; **P** -Practical; **C** – Credit, **CA** - Continuous Assessment; **ESE** -End Semester Examination.*

## 5. SUGGESTED PRACTICAL EXERCISES

The following practical outcomes (PrOs) are the subcomponents of the COs. These PrOs need to be attained to achieve the COs.

| Sr. No. | Practical Outcomes (PrOs) | Unit No. | Approx. Hrs. required |
|---|---|---|---|
| 1 | Use Cancer Dataset for detecting whether the cancer cells in data are benign or malignant. The data contains 2 types of cancers: 1. benign cancer (B) and 2. malignant cancer (M). Use relevant ML techniques for the same. | II | 04 |
| 2 | Implement following activation functions using python to build simple neural network.<br>a. ReLU<br>b. Sigmoid<br>c. Tanh | III | 06 |
| 3 | Implement following feed forward neural network using python programming:<br>a. Single layer feed forward neural network.<br>b. Multi-layer feed forward neural network. | III | 04 |
| 4 | Perform following data preprocessing on text/paragraph using NLTK library:<br>a. Write a Python program to tokenize words, sentence wise.<br>b. Write a python program that accepts the list of tokenized word and stems it into root word.<br>c. Write a program in python to identify the part of speech for each word in the text.<br>d. Write a Python NLTK program to remove stop words from a given text.<br>e. Write a python program for identifying and correcting misspelled words in a given text, such as an essay or a letter. | IV | 06 |
| 5 | Implement following Word embedding techniques in NLP.<br>a. TFIDF- Term Frequency Inverse document Frequency<br>b. BOW (Bag of Words)<br>c. Word2Vec | V | 04 |
| 6 | Implement Pre Trained word Embedding: GloVe Technique in NLP. | V | 04 |
| | **Total** | | **28** |

***Note***

*i. More **Practical Exercises** can be designed and offered by the respective course teacher to develop the industry relevant skills/outcomes to match the COs. The above table is only a suggestive list.*

*ii. The following are some **sample** 'Process' and 'Product' related skills (more may be added/deleted depending on the course) that occur in the above listed **Practical Exercises** of this course required which are embedded in the COs and ultimately the competency..*

| S. No. | Sample Performance Indicators for the PrOs | Weightage in % |
|---|---|---|
| 1 | Applying accurate and effective preprocessing techniques to the dataset. | 25 |
| 2 | Correct implementation of AI models using suitable libraries or frameworks. | 20 |
| 3 | Fulfillment of specific requirements for the given task and appropriate handling of domain-specific challenges. | 25 |
| 4 | Use of relevant evaluation metrics for the given task. | 10 |
| 5 | Applying best practices and well structured documentation. | 20 |
| | **Total** | **100** |

## 6. MAJOR EQUIPMENT/ INSTRUMENTS REQUIRED

This major equipment with broad specifications for the PrOs is a guide to procure them by the administrators to usher in uniformity of practical in all institutions across the state.

| S. No. | Equipment Name with Broad Specifications | PrO. No. |
|---|---|---|
| 1 | Computer system with operating system: Windows 7 or higher Ver., macOS, and Linux, with 4GB or higher RAM, Python versions: 2.7.X, 3.6.X | All |
| 2 | Python IDEs and Code Editors<br>Open Source : Anaconda Navigator | |

## 7. AFFECTIVE DOMAIN OUTCOMES

The following ***sample*** Affective Domain Outcomes (ADOs) are embedded in many of the above-mentioned COs and PrOs. More could be added to fulfill the development of this competency.

a) Work as a data scientist.
b) Follow ethical practices.

The ADOs are best developed through the laboratory/field based exercises. Moreover, the level of achievement of the ADOs according to Krathwohl's 'Affective Domain Taxonomy' should gradually increase as planned below:

i. 'Valuing Level' in 1st year
ii. 'Organization Level' in 2nd year.
iii. 'Characterization Level' in 3rd year.

## 8. UNDERPINNING THEORY

Only the major Underpinning Theory is formulated as higher-level UOs of *Revised Bloom's taxonomy* in order development of the COs and competency is not missed out by the students and teachers. If required, more such higher-level UOs could be included by the course teacher to focus on the attainment of COs and competency.

| Unit | Unit Outcomes (UOs)<br>(4 to 6 UOs at Application and above level) | Topics and Sub-topics |
|---|---|---|
| **Unit – I Fundamentals of Artificial Intelligence (AI)** | 1a. Describe basic concept of Artificial Intelligence<br>1b. Discuss areas of Artificial Intelligence<br>1c. Identify various applications of Artificial Intelligence | 1.1 Artificial Intelligence Introduction, Future of Artificial Intelligence, Definition of AI, History and Types of AI, AI ethics and limitations<br>1.2 Artificial Intelligence Areas Expert Systems, Natural Language Processing, Neural Networks, Robotics, Fuzzy Logic Systems<br>1.3 AI applications in various industries (healthcare, finance, manufacturing, etc.) |
| **Unit – II Foundation of Machine learning (ML)** | 2a. Understand well posed learning problem in machine learning.<br>2b. Describe different types of Machine learning.<br>2c. Explain basic concepts of Reinforcement learning.<br>2d. Compare types of Machine learning. | 2.1 What is Machine Learning? Well-Posed learning Problem<br>2.2 Types of Machine Learning: Supervised, Unsupervised, Reinforcement<br>2.3 Reinforcement Learning:<br>-Terms, Key features<br>-Approaches to implement reinforcement learning: Value based, Policy Based, Model based,<br>-Elements of reinforcement learning- Policy, reward signal, value function, Model;<br>- Types of reinforcement learning: Positive and Negative Comparison<br>2.4 Comparison between Supervised, unsupervised and reinforcement learning |
| **Unit– III Foundation of Neural Networks** | 3a. Understand foundation of Neural networks.<br>3b. Implement activation functions to build neural network.<br>3c. Implement types of ANN.<br>3d. Describe weight learning process in ANN. | 3.1 Introduction to Neural Networks: Understanding the Biological Neuron, Exploring the Artificial Neuron<br>3.2 Types of Activation Functions: ReLU, Sigmoid, Hyperbolic Tangent Function<br>3.3 Architectures of Neural Network: |

| | | |
|---|---|---|
| | | Single-layer feed forward network, Multi-layer feed forward ANNs, Recurrent network<br>3.4 Learning Process in ANN:<br>Number of layers, Direction ofsignal flow, Number of nodes in layers, Weight of interconnection between<br>3.5 Back Propagation |
| **Unit– IV Foundation of Natural Language processing (NLP)** | 4a. Understand basics of NLP.<br>4b. Discuss components of NLP.<br>4c. Explain phases of NLP.<br>4d. Apply data preprocessing techniques on text/ paragraph/ document. | 4.1 Introduction To NLP:<br>History of NLP, Advantages of NLP, Disadvantages of NLP<br>4.2 Components of NLP : Natural Language Understanding (NLU) ,Natural Language Generation (NLG)<br>4.3 Phases of NLP:<br>Lexical Analysis, Syntactic Analysis, Semantic Analysis, Discourse Integration, Pragmatic Analysis<br>4.4 Data Preprocessing Using NLTK: Tokenization, Frequency Distribution of Words, Filtering Stop Words, Stemming, Lemmatization, Parts Of Speech (POS)Tagging, Name Entity Recognition, WordNet<br>4.5 Types of Ambiguities in NLP |
| **Unit– V Word Embedding in Natural Language Processing** | 5a. Understand and Apply Word Embedding techniques.<br>5b. Compare word embedding techniques.<br>5c. Identify applications of NLP. | 5.1 Word Embedding techniques:<br>Term Frequency – Inverse Document Frequency (TFIDF), Bag Of Words (BoW), Word2Vec<br>5.2 Challenges with TF-IDF and BoW, Pre Trained word Embedding: GloVe : Global Vector for word representation<br>5.3 Applications of NLP : Question Answering, Spam Detection, Sentiment analysis, Machine Translation, Spelling Correction, Chatbot (ChatGPT) |

***Note**: The UOs need to be formulated at the 'Application Level' and above of Revised Bloom's Taxonomy' to accelerate the attainment of the COs and the competency.*

## 9. SUGGESTED SPECIFICATION TABLE FOR QUESTION PAPER DESIGN

| Unit No. | Unit Title | Teaching Hours | Distribution of Theory Marks | | | |
|---|---|---|---|---|---|---|
| | | | R Level | U Level | A | Total Marks |
| I | Fundamentals of Artificial Intelligence | 06 | 4 | 6 | - | 10 |
| II | Foundation of Machine learning | 09 | 4 | 8 | - | 12 |
| III | Foundation of Neural Networks | 09 | 4 | 6 | 6 | 16 |
| IV | Foundation of Natural Language processing | 09 | 4 | 6 | 6 | 16 |
| V | Word Embedding in Natural Language processing | 09 | 4 | 6 | 6 | 16 |
| **Total** | | **42** | **20** | **32** | **18** | **70** |

***Legends:*** *R=Remember, U=Understand, A=Apply and above (Revised Bloom's taxonomy)*
***Note****: This specification table provides general guidelines to assist students for their learning and to teachers to teach and question paper designers/setters to formulate test items/questions assess the attainment of the UOs. The actual distribution of marks at different taxonomy levels (of R, U and A) in the question paper may vary slightly from the above table.*

## 10. SUGGESTED STUDENT ACTIVITIES

Other than the classroom and laboratory learning, following are the suggested student-related ***co-curricular*** activities which can be undertaken to accelerate the attainment of the various outcomes in this course: Students should conduct following activities in group and prepare reports of about 5 pages for each activity, also collect/record physical evidences for their (student's) portfolio which will be useful for their placement interviews:

a) Explore different data repositories and register for ML based competitions on platforms like kaggle.
b) Undertake micro-projects in teams.
c) Give a seminar on any relevant topics.
d) Collect data from snapchat/ facebook and do sentiment analysis using NLP.

## 11. SUGGESTED SPECIAL INSTRUCTIONAL STRATEGIES (if any)

These are sample strategies, which the teacher can use to accelerate the attainment of the various outcomes in this course:

a) Massive open online courses (***MOOCs***) may be used to teach various topics/subtopics.
b) Guide student(s) in undertaking micro-projects.
c) ***'L' in section No. 4*** means different types of teaching methods that are to be employed by teachers to develop the outcomes.
d) About ***20% of the topics/sub-topics*** which are relatively simpler or descriptive in nature is to be given to the students for ***self-learning***, but to be assessed using different assessment methods.
e) With respect to ***section No.11***, teachers need to ensure to create opportunities and provisions for ***co-curricular activities***.
f) Guide students for open source python editors.

## 12. SUGGESTED MICRO-PROJECTS

***Only one micro-project*** is planned to be undertaken by a student that needs to be assigned to him/her in the beginning of the semester. In the first four semesters, the micro-project are group-based. However, in the fifth and sixth semesters, it should be preferably be ***individually*** undertaken to build up the skill and confidence in every student to become problem solver so that s/he contributes to the projects of the industry. In special situations where groups have to be formed for micro-projects, the number of students in the group should ***not exceed three.***

The micro-project could be industry application based, internet-based, workshop-based, laboratory-based or field-based. Each micro-project should encompass two or more COs which are in fact, an integration of PrOs, UOs and ADOs. Each student will have to maintain a dated work diary consisting of individual contributions in the project work and give a seminar presentation of it before submission. The total duration of the micro-project should not be less than ***16 (sixteen) student engagement hours*** during the course. The student ought to submit a micro-project by the end of the semester to develop the industry oriented COs.

A suggestive list of micro-projects is given here. This has to match the competency and the COs. Similar micro-projects could be added by the concerned course teacher:

- **Project idea 1: Credit Scoring System:** Student can use neural networks to develop an intelligent credit scoring system for the banks. Loan defaulters and fraudulent entities are required to be managed to avoid financial losses. Neural networks can be exceptionally useful in developing smart alternatives to traditional credit scoring systems.
- **Project idea 2: Sentiment analysis:** Businesses use it to track client product feedback, making it popular. If most reviews are positive, the companies are on the right course. Moreover, if the majority of evaluations generated by this NLP Project are deficient, the corporation can make efforts to improve the product.
- **Project idea 3: Conversational Bots:** ChatBots**,** most tech companies today use Chatbots, which are conversational bots, to communicate with their consumers and handle issues. It is an excellent method for both customers and businesses to save time.

## 13. SUGGESTED LEARNING RESOURCES

| S.No. | Title of Book | Author | Publication with place, year and ISBN |
|---|---|---|---|
| 1 | Machine Learning_ Step-by-Step Guide To Implement Machine Learning Algorithms with Python. | Rudolph Russell | Rudolph Russell Publications |
| 2 | Machine Learning | Saikat Dull, S.Chjandramouli | Das, Pearson |
| 3 | Natural Language Processing with Python | Steven Bird, Ewan klein & Edward Loper | O'Reilly, 2009 |
| 4 | Speech and Language Processing | Daniel Jurafsky & James H. Martin. | *Second Edition, Prentice Hall, 2008.* |

## 14. SOFTWARE/LEARNING WEBSITES

a. https://www.geeksforgeeks.org/machine-learning/
b. https://www.geeksforgeeks.org/natural-language-processing-nlp-tutorial/
c. https://www.tutorialspoint.com/machine_learning_with_python/index.htm
d. https://onlinecourses.nptel.ac.in/noc19_cs56/preview
e. https://onlinecourses.nptel.ac.in/noc20_cs29/preview

## 15. PO-COMPETENCY-CO MAPPING

| Semester V | Foundation of AI and ML (Course Code: 4351603) | | | | | | |
|---|---|---|---|---|---|---|---|
| | POs and PSOs | | | | | | |
| Competency & Course Outcomes | PO 1 Basic & Discipline specific knowledge | PO 2 Problem Analysis | PO 3 Design/ development of solutions | PO 4 Engineering Tools, Experimentation &Testing | PO 5 Engineering practices for society, sustainability & environment | PO 6 Project Management | PO 7 Life-long learning |
| **Competency**<br>Applying Artificial Intelligence and Machine Learning concepts to solve real world problems. | | | | | | | |
| Course Outcomes<br>CO a) Understand fundamental principles of Artificial Intelligence. | 3 | - | - | - | - | - | - |
| CO b) Compare types of machine learning. | 3 | 2 | 2 | 2 | 2 | 2 | 2 |
| CO c) Build a simple Neural Network model to solve real world problem. | 3 | 2 | 3 | 3 | 3 | 3 | 2 |
| CO d) Apply data preprocessing on text/paragraph using NLTK library. | 3 | 2 | 2 | 3 | 2 | 2 | 2 |
| CO e) Demonstrate word embedding techniques to develop real world NLP applications. | 3 | 2 | 3 | 3 | 3 | 3 | 2 |

Legend: '**3'** for high, '**2**' for medium, **'1'** for low or **'-'** for the relevant correlation of each competency, CO, with PO/ PSO

## 16. COURSE CURRICULUM DEVELOPMENT COMMITTEE

**GTU Resource Persons**

| Sr. No | Name and Designation | Institute | Email |
|---|---|---|---|
| 1 | Ms. Rikita D. Parekh<br>Lecturer - IT | Government Polytechnic for Girls, Ahmedabad | rikita.nagar@gmail.com |
| 2 | Dr. Lataben J. Gadhavi<br>Lecturer - IT | Government Polytechnic, Gandhinagar | latagpg@gmail.com |
| 3 | Dr. Esan P. Panchal-<br>Lecturer - IT | Government Polytechnic, Gandhinagar | esan.gpg@gmail.com |