# GUJARAT TECHNOLOGICAL UNIVERSITY

**BACHELOR OF ENGINEERING SYLLABUS**

**Subject Code : 3155202**

**Subject Name: Machine Learning using Python**

| | |
|---|---|
| WEF Academic Year: | 2021-22 |
| Semester: | 5 |
| Category of the Course: | Professional Core |

**Prerequisite:** Basic programming knowledge (Python) and a good understanding of foundational mathematics concepts (linear algebra, calculus, and probability)

**Rationale:** By providing a comprehensive introduction to essential ML concepts, algorithms, and practical applications, the course aims to equip students with the foundational knowledge and hands-on experience required to tackle real-world data challenges.

**Course Scheme:**

| Teaching Scheme | | | Total Credits | Assessment Pattern and Marks | | | | Total Marks |
|---|---|---|---|---|---|---|---|---|
| | | | | Theory | | Practical | | |
| L | T | PR | C | ESE (E) | PA(M) | ESE (V) | PA (I) | |
| 4 | 0 | 2 | 4 | 70 | 30 | 30 | 20 | 150 |

**Course Contents:**

| Sr. No. | Course Content | No. of Hours | % of Weightage |
|---|---|---|---|
| 1 | Introduction to Machine Learning<br>Overview of machine learning and its applications, Types of machine learning techniques: supervised, unsupervised, and reinforcement learning | 4 | 10 |
| 2 | Basics of python for Machine Learning<br>Introduction to Python programming language, Python libraries like NumPy and pandas for data manipulation, Introduction to scikit-learn for machine learning implementations | 4 | 10 |
| 3 | Overview of Probability<br>Statistical tools in Machine Learning, Concepts of probability, Random variables, Discrete, distributions, Continuous distributions, Multiple random variables, Central limit theorem, Sampling distributions, Hypothesis testing, Monte Carlo Approximation | 4 | 10 |
| 4 | Supervised Learning Techniques and Algorithms<br>Linear regression and logistic regression, Support Vector Machines (SVM), Decision Trees and Random Forests, k-Nearest Neighbors (k-NN) | 8 | 20 |
| 5 | Unsupervised Learning Techniques and Algorithms<br>Challenges in Unsupervised Learning, Preprocessing and Scaling, Different Kinds of Preprocessing, Data transformations and Scaling, Dimensionality | 8 | 20 |

| | | | |
|---|---|---|---|
| | Reduction, Feature Extraction, Principal Component Analysis, Clustering, k-Means Clustering, Agglomerative Clustering, DBSCAN | | |
| 6 | Model Evaluation and Improvement<br>Cross-Validation, Cross-Validation in scikit-learn, Benefits of Cross-Validation, k-Fold Cross-Validation and Other Strategies, Overfitting, Evaluation Metrics, Metrics for Binary Classification, Metrics for Multiclass Classification | 6 | 20 |
| 7 | Natural Language Processing (NLP)<br>Basics of NLP and text preprocessing, Using NLP techniques for sentiment analysis and text classification, Generative models | 6 | 10 |

**Reference Book:**

1. Introduction to Machine Learning with Python, O'REILLY
2. Machine Learning with Python for Everyone, Mark Fenner, Pearson
3. Python for data science for dummies 2nd Edition, John Paul Mueller, Luca Massaron, Wiley
4. Machine Learning, Anuradha Srinivasaraghavan, Vincy Joseph, Wiley

**Course Outcomes:**

After Completion of the Course, Student will able to :

| No | Course Outcomes | RBT Level* |
|---|---|---|
| 01 | Understand applications and basic concepts of Machine Learning | UN |
| 02 | Learn various supervised ML techniques and algorithms and apply to solve real world problems | UN |
| 03 | Learn and apply various supervised ML techniques and algorithms to solve real world problems | UN, AP |
| 04 | Learn and apply various unsupervised ML techniques and algorithms to solve real world problems | UN, AP |
| 05 | Apply various techniques for model evaluation and improvement | AP |

*RM: Remember, UN: Understand, AP: Apply, AN: Analyze, EL: Evaluate, CR: Create

**Suggested Course Practical List: The list below is very limited. Faculty is expected to create new practical definitions.**

1. Use NumPy to create arrays and perform basic mathematical operations on them.
2. Utilize pandas to load and manipulate datasets, perform data cleaning, and extract useful insights.
3. Implement a basic machine learning model using scikit-learn to classify different types of flowers in the Iris dataset.
4. Simulate the rolling of dice and analyze the distribution of outcomes.
5. Calculate the probability of a random variable using different probability distributions, such as binomial and normal distributions.
6. Implement logistic regression to classify emails as spam or non-spam based on given features.
7. Use k-Means clustering to segment customers based on their purchase behavior.
8. Apply PCA for dimensionality reduction on a high-dimensional dataset and visualize the reduced features.
9. Perform k-Fold Cross-Validation on a regression model to evaluate its performance on different subsets of the data.
10. Preprocess text data by tokenizing, removing stop words, and converting text to lowercase.

**List of Laboratory/Learning Resources Required:**

Web Resources:
- www.anaconda.com
- www.python.org
- www.w3schools.com
- https://www.learnpython.org/

* * * * * * *