**MCA230301 – Machine Learning**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Subject Total Credit** | **Teaching scheme (per week)** | | | **Examination scheme** | | | | |
| **Internal [60%]** | | **External [40%]** | | **Total** |
| **Theory** | **Tutorial** | **Practical** | **Theory** | **Practical** | **Theory** | **Practical** |
| 5 | 2 | 1 | 4 | 60 | 60 | 40 | 40 | 200 |

**Prerequisites:**

* Knowledge of statistics, and core fundamentals of computer science to create automated systems

**Course Objectives:**

* To provide a broad introduction to machine learning and its applications
* To understand the different types of machine learning algorithms
* To Work with neural networks

**Course Contents:**

|  |  |  |
| --- | --- | --- |
| **Units** | **Contents** | **Weightage** |
| **Unit – 1** | **Introduction** **to Machine Learning**   * What is Machine Learning * Types of Machine Learning * [Applications of Machine Learning](https://learning.oreilly.com/library/view/machine-learning/9789389588132/xhtml/chapter001.xhtml#ch1_7) * [State-of-The-Art Languages/Tools in Machine Learning](https://learning.oreilly.com/library/view/machine-learning/9789389588132/xhtml/chapter001.xhtml#ch1_8) * [Issues in Machine Learning](https://learning.oreilly.com/library/view/machine-learning/9789389588132/xhtml/chapter001.xhtml#ch1_9)   **Preparing to Model**   * [Machine Learning Activities](https://learning.oreilly.com/library/view/machine-learning/9789389588132/xhtml/chapter002.xhtml#ch2_2) * [Basic Types of Data in Machine Learning](https://learning.oreilly.com/library/view/machine-learning/9789389588132/xhtml/chapter002.xhtml#ch2_3) * [Exploring the Structure of Data](https://learning.oreilly.com/library/view/machine-learning/9789389588132/xhtml/chapter002.xhtml#ch2_4) * [Data Quality and Remediation](https://learning.oreilly.com/library/view/machine-learning/9789389588132/xhtml/chapter002.xhtml#ch2_5) * [Data Pre-Processing](https://learning.oreilly.com/library/view/machine-learning/9789389588132/xhtml/chapter002.xhtml#ch2_6) | **20%** |
| **Unit – 2** | **Modelling and Evaluation**   * [Selecting a Model](https://learning.oreilly.com/library/view/machine-learning/9789389588132/xhtml/chapter003.xhtml#ch3_2) * [Training a Model (for Supervised Learning)](https://learning.oreilly.com/library/view/machine-learning/9789389588132/xhtml/chapter003.xhtml#ch3_3) * [Model Representation and Interpretability](https://learning.oreilly.com/library/view/machine-learning/9789389588132/xhtml/chapter003.xhtml#ch3_4) * [Evaluating Performance of a Model](https://learning.oreilly.com/library/view/machine-learning/9789389588132/xhtml/chapter003.xhtml#ch3_5) * [Improving Performance of a Model](https://learning.oreilly.com/library/view/machine-learning/9789389588132/xhtml/chapter003.xhtml#ch3_6)   **Basics of Feature Engineering**   * Introduction * Feature Transformation * Feature Subset selection   **Bayesian Concept Learning**   * Bayes Theorem * Bayes Theorem and Concept Learning | **20%** |
| **Unit – 3** | [**Supervised Learning: Classification**](https://learning.oreilly.com/library/view/machine-learning/9789389588132/xhtml/chapter007.xhtml)   * [Example of Supervised Learning](https://learning.oreilly.com/library/view/machine-learning/9789389588132/xhtml/chapter007.xhtml#ch7_2) * [Classification Model](https://learning.oreilly.com/library/view/machine-learning/9789389588132/xhtml/chapter007.xhtml#ch7_3) * [Classification Learning Steps](https://learning.oreilly.com/library/view/machine-learning/9789389588132/xhtml/chapter007.xhtml#ch7_4) * [Common Classification Algorithms](https://learning.oreilly.com/library/view/machine-learning/9789389588132/xhtml/chapter007.xhtml#ch7_5)   [**Supervised Learning: Regression**](https://learning.oreilly.com/library/view/machine-learning/9789389588132/xhtml/chapter007.xhtml)   * [Example of Regression](https://learning.oreilly.com/library/view/machine-learning/9789389588132/xhtml/chapter008.xhtml#ch8_2) * [Common Regression Algorithms](https://learning.oreilly.com/library/view/machine-learning/9789389588132/xhtml/chapter008.xhtml#ch8_3) | **20%** |
| **Unit – 4** | [**Unsupervised Learning**](https://learning.oreilly.com/library/view/machine-learning/9789389588132/xhtml/chapter009.xhtml)   * [Unsupervised vs Supervised Learning](https://learning.oreilly.com/library/view/machine-learning/9789389588132/xhtml/chapter009.xhtml#ch9_2) * [Application of Unsupervised Learning](https://learning.oreilly.com/library/view/machine-learning/9789389588132/xhtml/chapter009.xhtml#ch9_3) * [Clustering](https://learning.oreilly.com/library/view/machine-learning/9789389588132/xhtml/chapter009.xhtml#ch9_4) * [Finding Pattern using Association Rule](https://learning.oreilly.com/library/view/machine-learning/9789389588132/xhtml/chapter009.xhtml#ch9_5)   **Other Types of Learning**   * Representation Learning * Association Rule Learning Algorithm * Ensemble Learning Algorithm | **20%** |
| **Unit – 5** | **Basics of Neural Network**   * [Understanding the Biological Neuron](https://learning.oreilly.com/library/view/machine-learning/9789389588132/xhtml/chapter010.xhtml#ch10_2) * [Exploring the Artificial Neuron](https://learning.oreilly.com/library/view/machine-learning/9789389588132/xhtml/chapter010.xhtml#ch10_3) * [Types of Activation Functions](https://learning.oreilly.com/library/view/machine-learning/9789389588132/xhtml/chapter010.xhtml#ch10_4) * [Early Implementations of ANN](https://learning.oreilly.com/library/view/machine-learning/9789389588132/xhtml/chapter010.xhtml#ch10_5) * [Architectures of Neural Network](https://learning.oreilly.com/library/view/machine-learning/9789389588132/xhtml/chapter010.xhtml#ch10_6) * [Learning Process in ANN](https://learning.oreilly.com/library/view/machine-learning/9789389588132/xhtml/chapter010.xhtml#ch10_7) * [Backpropagation](https://learning.oreilly.com/library/view/machine-learning/9789389588132/xhtml/chapter010.xhtml#ch10_8) * [Deep Learning](https://learning.oreilly.com/library/view/machine-learning/9789389588132/xhtml/chapter010.xhtml#ch10_9) | **20%** |

**Practical List**

**All practical coding should be done in Python**

1. Read the .csv file and calculate the missing value of each column and handle the missing values with imputation techniques.
2. Implement the naïve Bayesian classifier for a sample training data set stored in a .CSV file. Compute the accuracy of the classifier, considering few test data sets.
3. Write a program to implement the k-Nearest Neighbour algorithm to classify the iris to predict correct and wrong predictions. Use Python ML library classes for the prediction.
4. Write a program to implement a Logistic Regression algorithm to classify the housing price data set to predict correct and wrong predictions. Use Python ML library classes for predicting the problem.
5. Implement and compare SVM, KNN and Logistic regression algorithm to classify the Android mobile purchase records data set for predicting both correct and wrong predictions.
6. Write a python code to implement K-nearest neighbor for the given dataset. Assume that value of K=2. Calculate the accuracy and predict the Shirt size for the single customer using the below given sample dataset.

|  |  |  |
| --- | --- | --- |
| Height (in cms) | Weight (in kgs) | T Shirt Size |
| 158 | 58 | M |
| 158 | 59 | M |
| 158 | 63 | M |
| 160 | 59 | M |
| 160 | 60 | M |
| 163 | 60 | M |
| 163 | 61 | M |
| 160 | 64 | L |
| 163 | 64 | L |
| 165 | 61 | L |
| 165 | 62 | L |
| 165 | 65 | L |
| 168 | 62 | L |
| 168 | 63 | L |
| 168 | 66 | L |
| 170 | 63 | L |
| 170 | 64 | L |
| 170 | 68 | L |

1. Write a python code to implement NaiveBayes for the given dataset. Calculate the accuracy and predict the class for individual temperature factor?

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Outlook** | **Temperature** | **Humidity** | **Wind** | **Play Tennis** |
| Sunny | Hot | High | Weak | No |
| Sunny | Hot | High | Strong | No |
| Overcast | Hot | High | Weak | Yes |
| Rain | Mild | High | Weak | Yes |
| Rain | Mild | Low | Strong | No |
| Overcast | Hot | High | Strong | No |
| Sunny | Hot | Low | Weak | Yes |
| Sunny | Hot | High | Weak | Yes |
| Overcast | Hot | High | Strong | Yes |
| Rain | Mild | High | Strong | No |

1. Write a python code to apply a classification algorithm to classify whether a person can buy a computer or not based on given test data :

**Test Data**

Age : Youth Income : Low Student : No Credit Rating : Fair Buy Computer - ??

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **ItemNo** | **Age** | **Income** | **Student** | **Credit** | **BuyComputer** |
| 1 | Youth | High | No | Fair | No |
| 2 | Youth | High | No | Excellent | No |
| 3 | Middle | High | No | Fair | Yes |
| 4 | Senior | Medium | No | Fair | Yes |
| 5 | Senior | Low | Yes | Fair | Yes |
| 6 | Middle | Low | Yes | Excellent | No |
| 7 | Senior | Low | Yes | Excellent | Yes |
| 8 | Youth | Medium | No | Fair | No |
| 9 | Youth | Low | Yes | Fair | Yes |
| 10 | Senior | Medium | Yes | Fair | Yes |
| 11 | Youth | Medium | Yes | Excellent | Yes |
| 12 | Middle | Medium | No | Excellent | Yes |
| 13 | Middle | High | Yes | Fair | Yes |
| 14 | Senior | Medium | Yes | Excellent | No |

1. Write a python code for the below mentioned dataset to perform:
   1. To calculate missing values of every feature
   2. Clean and pre-process the dataset by removing the missing values
   3. Impute the missing values with Mean, Median, Mode
   4. Encode Categorical Data
   5. Find the accuracy using Linear Regression Model
   6. Represent the confusion matrix
   7. Represent in a graph

|  |  |  |  |
| --- | --- | --- | --- |
| **Country** | **Age** | **Salary** | **Purchased** |
| India | 49 | 62000 | No |
| Pakistan | 32 | 38000 | Yes |
| Bhutan | 35 | 44000 | No |
| Bangladesh | 43 | 51000 | No |
| Nepal | 45 | NaN | Yes |
| Srilanka | 40 | 48000 | Yes |
| Burma | NaN | 42000 | No |
| China | 53 | 69000 | Yes |
| Afganistan | 55 | 73000 | No |

1. Implement supervised machine learning algorithm Random forest, Decision Tree algorithm in python on Pima Indians Diabetes dataset and obtain its accuracy level.

Represent it with the bar graph

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Pregnancies** | **Glucose** | **BP** | **SkinThickness** | **Insulin** | **BMI** | **DiabetesPedigreeFunction** | **Age** | **Outcome** |
| 6 | 148 | 72 | 35 | 0 | 33.6 | 0.627 | 50 | 1 |
| 1 | 85 | 66 | 29 | 0 | 26.6 | 0.351 | 31 | 0 |
| 8 | 183 | 64 | 0 | 0 | 23.3 | 0.672 | 32 | 1 |
| 1 | 89 | 66 | 23 | 94 | 28.1 | 0.167 | 21 | 0 |
| 0 | 137 | 40 | 35 | 168 | 43.1 | 2.288 | 33 | 1 |
| 5 | 116 | 74 | 0 | 0 | 25.6 | 0.201 | 30 | 0 |
| 3 | 78 | 50 | 32 | 88 | 31 | 0.248 | 26 | 1 |
| 10 | 115 | 0 | 0 | 0 | 35.3 | 0.134 | 29 | 0 |
| 2 | 197 | 70 | 45 | 543 | 30.5 | 0.158 | 53 | 1 |
| 8 | 125 | 96 | 0 | 0 | 0 | 0.232 | 54 | 1 |
| 4 | 110 | 92 | 0 | 0 | 37.6 | 0.191 | 30 | 0 |
| 10 | 168 | 74 | 0 | 0 | 38 | 0.537 | 34 | 1 |
| 10 | 139 | 80 | 0 | 0 | 27.1 | 1.441 | 57 | 0 |
| 1 | 189 | 60 | 23 | 846 | 30.1 | 0.398 | 59 | 1 |
| 5 | 166 | 72 | 19 | 175 | 25.8 | 0.587 | 51 | 1 |
| 7 | 100 | 0 | 0 | 0 | 30 | 0.484 | 32 | 1 |
| 0 | 118 | 84 | 47 | 230 | 45.8 | 0.551 | 31 | 1 |
| 7 | 107 | 74 | 0 | 0 | 29.6 | 0.254 | 31 | 1 |
| 1 | 103 | 30 | 38 | 83 | 43.3 | 0.183 | 33 | 0 |
| 1 | 115 | 70 | 30 | 96 | 34.6 | 0.529 | 32 | 1 |
| 3 | 126 | 88 | 41 | 235 | 39.3 | 0.704 | 27 | 0 |

1. Write the python code for the following dataset which is mentioned below,
2. Import the libraries
3. Import the data-set
4. Check out the missing values
5. See the Categorical Values
6. Splitting the data-set into Training and Test Set
7. Feature Scaling

|  |  |  |  |
| --- | --- | --- | --- |
| **Country** | **Age** | **Salary** | **Purchased** |
| France | 44.00 | 44000 | Yes |
| Span | 44.00 | 50000 | No |
| German | 54.00 | 23000 | Yes |
| German | 34.00 | NaN | Yes |
| France | 44.00 | NaN | No |
| Span | 34.00 | 23667 | Yes |
| German | 23.00 | 67000 | Yes |
| Span | 45.00 | 45000 | No |
| France | 23.00 | 56000 | No |
| Span | 23.00 | 66000 | No |

1. Perform evaluation metrics for all algorithms
   1. AUC
   2. Precision
   3. Recall
   4. Specificity
   5. Sensitivity
   6. Mean absolute percentage error

**Learning Outcome:**

This Course Covers -

CO1: Learning the fundamentals of machine learning with its algorithms.

CO2: Understanding the types of machine learning algorithms

CO3: Neural network algorithm allows to learn about perceptron and model representation

CO4: These learning algorithms allows to build their career as Machine learning engineers, Data Scientist and Business Intelligent developers

**Text Book:**

1. Machine Learning, Saikat Dutt, Subramanian Chandramouli, Amit Kumar Das, Pearson Publications

**Reference Books & Links**

1. Tom M Mitchell, “Machine Learning”, McGraw Hill
2. Anuradha Srinivasaraghavan, Vincy Joseph, “Machine Learning”, Wiley India

**Chapter wise coverage**

Unit 1: Chapter No. 1, 2 of Text Book-1

Unit 2: Chapter No. 3, 4 and 6 of Text Book-1

Unit 3: Chapter No. 7, 8 of Text Book - 1

Unit 4: Chapter No. 9, 10 of Text Book – 1

Unit 5: Chapter No. 11 of Text Book - 1