U.S.N.										

B.M.S. College of Engineering, Bengaluru-560019

Autonomous Institute Affiliated to VTU

July 2023 Semester End Main Examinations

Programme: B.E. **Semester: VI**
Branch: ES – Cluster Elective **Duration: 3 hrs.**
Course Code: 19EC6CE1ML **Max Marks: 100**
Course: Machine Learning **Date: 19.07.2023**

Instructions: 1. Answer any FIVE full questions, choosing one full question from each unit.
2. Missing data, if any, may be suitably assumed.

Important Note: Completing your answers, compulsorily draw diagonal cross lines on the remaining blank pages. Revealing of identification, appeal to evaluator will be treated as malpractice.

		UNIT - I	CO	PO	Marks
1	a)	Define Machine Learning. How machine learning is different from rule based programming? Identify the type of machine learning to be applied to the following use cases: • Product price prediction • An intelligent agent navigates the vehicle • Prediction on a semiconductor chip faulty or not	*CO 1*	*PO 1*	**10**
	b)	Discuss on the utility of 'Numpy' and 'Pandas' libraries in building ML models.	*CO 1*	*PO 1*	**6**
	c)	• lst = [1,2,[3,4],[5,[6,7,['#NLP ']],8,9],10,11] • d={'k1':[1,2,3,{'voiceassistant':['Siri','Cortana',{'Alexa':[1, 2,3,'$hello']}]}]} Write python expression to extract "NLP" from the list and "Alexa" from the dictionary.	*CO 1*	*PO 1*	**4**
		UNIT - II			
2	a)	Develop the pseudocode for kNN algorithm and discuss it's pros and cons.	*CO 3*	*PO 3*	**8**
	b)	Discuss on the decision tree splitting criteria gini and entropy.	*CO 1*	*PO 1*	**6**
	c)	Analyze the utility of cross validation technique and random forest in feature selection.	*CO 2*	*PO 2*	**6**
		OR			
3	a)	Justify the effectiveness of ensembling: Bagging and Boosting in model performance improvement.	*CO 2*	*PO 2*	**8**
	b)	**Predicted** (Actual) |	1	0 | | 1 | 655 | 8 | | 0 | 12 | 125 | Given the confusion matrix from the result of a classifier, evaluate the model performance based on precision, recall, specificity and F1-score.	*CO 1*	*PO 1*	**8**

<table>
<tr><td></td><td></td><td>c)</td><td>Differentiate between Standard scaler and MinMax scaler.</td><td>CO 2</td><td>PO 2</td><td>4</td></tr>
<tr><td></td><td></td><td></td><td>UNIT - III</td><td></td><td></td><td></td></tr>
<tr><td></td><td>4</td><td>a)</td><td>Mention the mathematical model of Naive Baye’s classifier. Given in the data table two features X_1 & X_2 and ‘y’ as the class variable. If a new data point has X_1=2.6 and X_2= 70, applying Gaussian Naive Baye’s, predict it’s class.
<table>
<tr><td>y</td><td colspan="5">Class-A</td><td colspan="5">Class-B</td></tr>
<tr><td>X1</td><td>4.1</td><td>3.4</td><td>2.9</td><td>2.8</td><td>1.6</td><td>2.5</td><td>2.0</td><td>1.7</td><td>1.4</td><td>1.2</td></tr>
<tr><td>X2</td><td>78</td><td>70</td><td>62</td><td>66</td><td>58</td><td>68</td><td>64</td><td>62</td><td>70</td><td>72</td></tr>
</table></td><td>CO 3</td><td>PO 3</td><td>10</td></tr>
<tr><td></td><td></td><td>b)</td><td>Identify the pandas functions & sklearn APIs to : create dataframe, split features & target, split train & test set, create Logistic Regression model instance, train the model on trainset.</td><td>CO 3</td><td>PO 3</td><td>10</td></tr>
<tr><td></td><td></td><td></td><td>UNIT - IV</td><td></td><td></td><td></td></tr>
<tr><td></td><td>5</td><td>a)</td><td>Estimate the R2-score of linear regression with the illustration of residue error (MSE). Give justification on the error you get after one iteration of training. Training starts with θ_0=1 and θ_1=0.5. Refer the data table given below.
<table>
<tr><td>(x_i)</td><td>(y_i)</td></tr>
<tr><td>1</td><td>3</td></tr>
<tr><td>2</td><td>5</td></tr>
<tr><td>3</td><td>7</td></tr>
<tr><td>4</td><td>9</td></tr>
</table></td><td>CO 2</td><td>PO 2</td><td>8</td></tr>
<tr><td></td><td></td><td>b)</td><td>Write python code to create a numpy array as: ([[0, 1],[2, 3],[4, 5]]) and generate polynomial features with:
i)degree=2 and interaction_only=True
ii)degree=2 and interaction_only= False</td><td>CO 3</td><td>PO 3</td><td>6</td></tr>
<tr><td></td><td></td><td>c)</td><td>Interpret the relationship between IQR, skewness and outliers in a feature distribution.</td><td>CO 2</td><td>PO 2</td><td>6</td></tr>
<tr><td></td><td></td><td></td><td>OR</td><td></td><td></td><td></td></tr>
<tr><td></td><td>6</td><td>a)</td><td>Develop the algorithm flow of Gradient Descent for simple linear regression with the illustration of cost function.</td><td>CO 3</td><td>PO 3</td><td>10</td></tr>
<tr><td></td><td></td><td>b)</td><td>Interpret the evaluation metrics used for regression models.</td><td>CO 2</td><td>PO 2</td><td>10</td></tr>
<tr><td></td><td></td><td></td><td>UNIT - V</td><td></td><td></td><td></td></tr>
<tr><td></td><td>7</td><td>a)</td><td>Define Deep Learning and mention few of it’s applications. Illustrate a single layer perceptron model.</td><td>CO 1</td><td>PO 1</td><td>10</td></tr>
<tr><td></td><td></td><td>b)</td><td>Differentiate between agglomerative and divisive clustering. Given 4 sample data records with 2 features as: A(2,3), B(6,1), C(1,2) and D(3,0) and initialized 2 centroids as: C1(4,2) and C2(2,1). Form clusters with one iteration of training using Euclidean distance metric. Show the clusters in a 6X6 data space.</td><td>CO 3</td><td>PO 3</td><td>10</td></tr>
</table>

****