School of Computer Science Engineering and Technology

Course-B.Tech.	**Type**- Specialization Core II
Course Code- CSET228	**Course Name**- Data Mining and Predictive Modelling (Lab)
Year- 2024	**Semester**- Even
Date- 26/02/2024	**Batch**- 2022-2025

CO-Mapping

Q(s)	CO1	CO2	CO3
Q1	√	√	
Q2	√	√	

Objectives

1. Students will be able to gain understanding of Support Vector Machine Classifier.
2. Students will be able to implement decision tree.

Ques. 1

Imagine you are a data scientist tasked with developing a handwriting recognition system for a postal service company. The goal is to automate the sorting process by accurately identifying handwritten digits on envelopes. You decide to use SVMs for this task due to their effectiveness in classification problems. Follow the following steps.

1. **Dataset:** Download the dataset from the link.
   https://archive.ics.uci.edu/ml/datasets/Optical+Recognition+of+Handwritten+Digits
   Or use load_digits() to load the dataset.
2. Check the shape of data.
3. Display the first 10 images using matplotlib.
4. Extract the Independent and Dependent Variable (**Hint:** Indep Var=digits.data, Dep Var= digits.target)
5. Split the dataset into 80% for training and the rest 20% for testing. (sklearn.model_selection.train_test_split function)
6. Build an SVM model using Sklearn with default parameters.
7. Predict the target values in the testing set.
8. Apply classification metrics and visualize the results as graphs.
9. Playing with SVM: Change the following parameters of the SVM and analyze their performance for training and testing using evaluation measures.
   a) Kernel: {'linear', 'poly', 'rbf', 'sigmoid', 'precomputed'}
   b) degreeint, default=3
   c) gamma {'scale', 'auto'} or float, default='scale'
   d) random_stateint, RandomState instance or None, default=None
   e) Cfloat, default=1.0

Ques. 2

Use the given dataset Carseats to classify the high sales. Condition: sales>8 then yes good sale otherwise no.

1. In this approach first read the dataset using pandas.
2. Define target variable in Y.
3. use Classification Decision Tree to classify the high sales.
4. You need to required EDA and preprocessing on this dataset.

Dataset link - http://tinyurl.com/carsalesdatasetbu