



SIMATS ENGINEERING

SAVEETHA INSTITUTE OF MEDICAL AND TECHNICAL SCIENCES

CHENNAI-602105



Engineer to Excel

| | |
|------------------|--------------|
| Register Number: | Name: |
| Course Code: | Course Name: |

SET – 1

- Mr Arun wants to start his own mobile phone company and he wants to wage an uphill battle with big smartphone brands like Samsung and Apple. But he doesn't know how to estimate the price of a mobile that can cover both marketing and manufacturing costs. So in this task, you don't have to predict the actual prices of the mobiles but you have to predict the price range of the mobiles. ”
 - Read the Mobile price dataset using the Pandas module
 - print the 1st five rows.
 - Basic statistical computations on the data set or distribution of data
 - the columns and their data types
 - Detects null values in the dataset. If there is any null values replaced it with mode value
 - Explore the data set using heatmap
 - Split the data in to test and train
 - Fit in to the model Naive Bayes Classifier
 - Predict the model
 - Find the accuracy of the model
- Implement a Python program for the most specific hypothesis using Find-S algorithm for the following given dataset and show the output:

| Example | Sky | Air Temp | Humidity | Wind | Water | Forecast | Enjoy Sport |
|---------|-------|----------|----------|--------|-------|----------|-------------|
| 1 | Sunny | Warm | Normal | Strong | Warm | Same | Yes |
| 2 | Sunny | Warm | High | Strong | Warm | Same | Yes |
| 3 | Rainy | Cold | High | Strong | Warm | Change | No |
| 4 | Sunny | Warm | High | Strong | Cool | Change | Yes |

- Develop a Python code for implementing Linear regression and show its performance
- Develop a Python code for implementing the KNN algorithm with an example.

| AIM ALGORITHM (20) | & PROGRAM (40) | DEBUGGING(10) | OUTPUT (20) | GITHUB UPLOAD (10) | TOTAL (100) |
|--------------------|----------------|---------------|-------------|--------------------|-------------|
| | | | | | |



SIMATS ENGINEERING

SAVEETHA INSTITUTE OF MEDICAL AND TECHNICAL SCIENCES

CHENNAI-602105



Engineer to Excel

| | |
|-------------------------|---------------------|
| Register Number: | Name: |
| Course Code: | Course Name: |

SET – 2

1. How is the Perceptron algorithm applied to the Iris flower classification problem?
Anna is a botanist who is studying the Iris genus. She has collected data on the sepal length, sepal width, petal length, and petal width of various Iris flowers and wants to classify the flowers into their respective species based on their physical characteristics. Anna decides to use the Perceptron algorithm for this task.
2. For a given set of training data examples stored in a .CSV file, implement and demonstrate the Candidate-Elimination algorithm to output a description of the set of all hypotheses consistent with the training examples.

| Example | Sky | Air Temp | Humidity | Wind | Water | Forecast | Enjoy Sport |
|---------|-------|----------|----------|--------|-------|----------|-------------|
| 1 | Sunny | Warm | Normal | Strong | Warm | Same | Yes |
| 2 | Sunny | Warm | High | Strong | Warm | Same | Yes |
| 3 | Rainy | Cold | High | Strong | Warm | Change | No |
| 4 | Sunny | Warm | High | Strong | Cool | Change | Yes |

3. Develop a Python code for implementing Logistic regression and show its performance
4. Develop a Python code for implementing the Expectation Maximization algorithm with an example

| AIM ALGORITHM & (20) | PROGRAM (40) | DEBUGGING(10) | OUTPUT (20) | GITHUB UPLOAD (10) | TOTAL (100) |
|----------------------|--------------|---------------|-------------|--------------------|-------------|
| | | | | | |



SIMATS ENGINEERING

SAVEETHA INSTITUTE OF MEDICAL AND TECHNICAL SCIENCES

CHENNAI-602105



Engineer to Excel

| | |
|-------------------------|---------------------|
| Register Number: | Name: |
| Course Code: | Course Name: |

SET – 3

- Jack is a car enthusiast and wants to buy a new car. He wants to find the best deal and decides to use machine learning to predict the prices of different car models. Jack collects data on various features such as the make, model, year, engine size, and number of doors, as well as the sale price of each car. He splits the data into a training set and a test set and trains a linear regression model on the training data. Car Price Prediction with Machine Learning
 - Read the dataset using the Pandas module
 - print the 1st five rows.
 - Basic statistical computations on the data set or distribution of data
 - the columns and their data types
 - Detects null values in the dataset. If there is any null values replaced it with mode value
 - Explore the data set using heatmap
 - Split the data in to test and train
 - Fit in to the model Naive Bayes Classifier
 - Predict the model
 - Find the accuracy of the model
- Implement a Python program for the most specific hypothesis using Find-S algorithm for the following given dataset and show the output:

| Origin | Manufacturer | Color | Decade | Type | Example Type |
|--------|--------------|-------|--------|---------|--------------|
| Japan | Honda | Blue | 1980 | Economy | Positive |
| Japan | Toyota | Green | 1970 | Sports | Negative |
| Japan | Toyota | Blue | 1990 | Economy | Positive |
| USA | Chrysler | Red | 1980 | Economy | Negative |
| Japan | Honda | White | 1980 | Economy | Positive |

- Develop a Python code for implementing Polynomial regression and show its performance
- Develop a Python code for implementing the KNN algorithm with an example

| AIM & ALGORITHM (20) | PROGRAM (40) | DEBUGGING (10) | OUTPUT (20) | GITHUB UPLOAD (10) | TOTAL (100) |
|----------------------|--------------|----------------|-------------|--------------------|-------------|
| | | | | | |



SIMATS ENGINEERING

SAVEETHA INSTITUTE OF MEDICAL AND TECHNICAL SCIENCES

CHENNAI-602105



| | |
|-------------------------|---------------------|
| Register Number: | Name: |
| Course Code: | Course Name: |

SET – 4

- John is a young professional who wants to buy his first home. He knows that his credit score is an important factor in determining whether he will be approved for a loan, so he decides to check it. He goes to a financial website that offers a free credit score prediction service based on machine learning algorithms.
 - print the 1st five rows.
 - Basic statistical computations on the data set or distribution of data
 - the columns and their data types
 - Detects null values in the dataset. If there is any null values replaced it with mode value
 - Explore the data set using `ps.box(Credit Scores Based on Occupation)`
 - Split the data in to test and train
 - Fit in to the model Naive Bayes Classifier
 - Predict the model
- For a given set of training data examples stored in a .CSV file, implement and demonstrate the Candidate-Elimination algorithm to output a description of the set of all hypotheses consistent with the training examples.

| Origin | Manufacturer | Color | Decade | Type | Example Type |
|--------|--------------|-------|--------|---------|--------------|
| Japan | Honda | Blue | 1980 | Economy | Positive |
| Japan | Toyota | Green | 1970 | Sports | Negative |
| Japan | Toyota | Blue | 1990 | Economy | Positive |
| USA | Chrysler | Red | 1980 | Economy | Negative |
| Japan | Honda | White | 1980 | Economy | Positive |

- Develop a Python code for implementing Linear and Polynomial regression and show its performance
- Develop a Python code for implementing the Expectation Maximization algorithm with and example

| AIM ALGORITHM (20) | & PROGRAM (40) | DEBUGGI NG(10) | OUTPUT (20) | GITHUB UPLOAD (10) | TOTAL (100) |
|--------------------------|----------------------|-------------------|-------------|--------------------------|----------------|
| | | | | | |



SIMATS ENGINEERING

SAVEETHA INSTITUTE OF MEDICAL AND TECHNICAL SCIENCES

CHENNAI-602105



Engineer to Excel

| | |
|-------------------------|---------------------|
| Register Number: | Name: |
| Course Code: | Course Name: |

SET – 5

- Sarah is a botanist who studies different species of plants. She is particularly interested in the Iris genus and has collected data on the sepal length, sepal width, petal length, and petal width of various Iris flowers. She wants to use this data to classify the flowers into their respective species based on their physical characteristics. Sarah decides to use a machine learning algorithm for this task and trains a model on her collected data. The algorithm uses the sepal and petal measurements as input features and predicts the species of the flower based on these features. One day, Sarah is out in the field collecting new samples of Iris flowers. She measures the sepal and petal characteristics of each flower and inputs this information into the trained model. The model then predicts the species of each flower based on its physical characteristics.
 - Read the IRIS.csv Data set using the Pandas module
 - Plot the data using a scatter plot "sepal_width" versus "sepal_length" and color species.
 - Split the data
 - Fit the data to the model
 - Predict the model with new test data [5, 3, 1, .3]
- Implement a Python program for the most specific hypothesis using Find-S algorithm for the following given dataset and show the output:

| Example | Shape | Size | Color | Surface | Thickness | Target Concept |
|---------|----------|-------|-------|-----------|-----------|----------------|
| 1 | Circular | Large | Light | Smooth | Thick | Malignant (+) |
| 2 | Circular | Large | Light | Irregular | Thick | Malignant (+) |
| 3 | Oval | Large | Dark | Smooth | Thin | Benign (-) |
| 4 | Oval | Large | Light | Irregular | Thick | Malignant (+) |

- Develop a Python code for implementing Logistic regression and show its performance
- Develop a Python code for implementing the Naive Bayes algorithm with an example

| AIM ALGORITHM (20) | & PROGRAM (40) | DEBUGGI NG(10) | OUTPUT (20) | GITHUB UPLOAD (10) | TOTAL (100) |
|--------------------------|----------------------|-------------------|-------------|--------------------------|----------------|
| | | | | | |



SIMATS ENGINEERING

SAVEETHA INSTITUTE OF MEDICAL AND TECHNICAL SCIENCES

CHENNAI-602105



Engineer to Excel

| | |
|-------------------------|---------------------|
| Register Number: | Name: |
| Course Code: | Course Name: |

SET – 6

- Mark and his family are planning to move to a new city and are in the market for a new home. They have been searching online for homes in their desired area and have found several properties that meet their requirements. However, they are not sure about the prices of these homes and want to get a rough estimate before making an offer. How will you help Mark to buy a new house.
 - Read the house Data set using the Pandas module
 - print the 1st five rows.
 - Basic statistical computations on the data set or distribution of data
 - Print the columns and their data types
 - Detects null values in the dataset. If there is any null values replaced it with mode value
 - Explore the data set using heatmap
 - Split the data in to test and train
 - Predict the price of a house
- For a given set of training data examples stored in a .CSV file, implement and demonstrate the Candidate-Elimination algorithm to output a description of the set of all hypotheses consistent with the training examples.

| Example | Shape | Size | Color | Surface | Thickness | Target Concept |
|---------|----------|-------|-------|-----------|-----------|----------------|
| 1 | Circular | Large | Light | Smooth | Thick | Malignant (+) |
| 2 | Circular | Large | Light | Irregular | Thick | Malignant (+) |
| 3 | Oval | Large | Dark | Smooth | Thin | Benign (-) |
| 4 | Oval | Large | Light | Irregular | Thick | Malignant (+) |

- Develop a Python code for implementing Linear regression and show its performance
- Develop a Python code for implementing the EM algorithm with an example

| AIM & ALGORITHM (20) | PROGRAM (40) | DEBUGGING (10) | OUTPUT (20) | GITHUB UPLOAD (10) | TOTAL (100) |
|----------------------|--------------|----------------|-------------|--------------------|-------------|
| | | | | | |



SIMATS ENGINEERING
SAVEETHA INSTITUTE OF MEDICAL AND TECHNICAL SCIENCES
CHENNAI-602105



| | |
|-------------------------|---------------------|
| Register Number: | Name: |
| Course Code: | Course Name: |

SET – 7

1. Can the breast cancer classification problem be solved using Naive Bayes classification
 - a) print the 1st five rows.
 - b) Basic statistical computations on the data set or distribution of data
 - c) the columns and their data types
 - d) Detects null values in the dataset. If there is any null values replaced it with mode value
 - e) Split the data in to test and train
 - f) evaluate the performance of the model by evaluation metrics such as confusion matrix.
2. Implement a Python program for the most specific hypothesis using Find-S algorithm for the following given dataset and show the output:

| Size | Color | Shape | Class |
|-------|-------|----------|-------|
| Big | Red | Circle | No |
| Small | Red | Triangle | No |
| Small | Red | Circle | Yes |
| Big | Blue | Circle | No |
| Small | Blue | Circle | Yes |

3. Develop a Python code for implementing Polynomial regression and show its performance
4. Develop a Python code for implementing the KNN algorithm with an example.

| AIM & ALGORITHM (20) | PROGRAM (40) | DEBUGGING(10) | OUTPUT (20) | GITHUB UPLOAD (10) | TOTAL (100) |
|----------------------|--------------|---------------|-------------|--------------------|-------------|
| | | | | | |



SIMATS ENGINEERING

SAVEETHA INSTITUTE OF MEDICAL AND TECHNICAL SCIENCES

CHENNAI-602105



Engineer to Excel

| | |
|-------------------------|---------------------|
| Register Number: | Name: |
| Course Code: | Course Name: |

SET – 8

- You are a data scientist at a retail company and your manager has asked you to create a model to predict future sales. The company has been collecting data on sales, and advertising expenditures, for the past 5 years. Your manager wants to use this information to forecast sales for the next quarter and make informed decisions about advertising and inventory. Your task is to build a predictive model that takes into account past sales data, and advertising expenditures, to forecast sales for the next quarter. You decide to use linear regression to build your model because it is a simple and interpretable method for predicting a continuous outcome.
 - print the 1st five rows.
 - Basic statistical computations on the data set or distribution of data
 - the columns and their data types
 - Explore the data using scatterplot
 - Detects null values in the dataset. If there is any null values replaced it with mode value
 - Split the data in to test and train
 - Predict the model
- For a given set of training data examples stored in a .CSV file, implement and demonstrate the Candidate-Elimination algorithm to output a description of the set of all hypotheses consistent with the training examples.

| Size | Color | Shape | Class |
|-------|-------|----------|-------|
| Big | Red | Circle | No |
| Small | Red | Triangle | No |
| Small | Red | Circle | Yes |
| Big | Blue | Circle | No |
| Small | Blue | Circle | Yes |

- Develop a Python code for implementing Logistic regression and show its performance
- Develop a Python code for implementing the Naive Bayes algorithm with an example.

| AIM ALGORITHM (20) | & PROGRAM (40) | DEBUGGI NG(10) | OUTPUT (20) | GITHUB UPLOAD (10) | TOTAL (100) |
|--------------------------|----------------------|-------------------|-------------|--------------------------|----------------|
| | | | | | |



SIMATS ENGINEERING

SAVEETHA INSTITUTE OF MEDICAL AND TECHNICAL SCIENCES

CHENNAI-602105



Engineer to Excel

| | |
|-------------------------|---------------------|
| Register Number: | Name: |
| Course Code: | Course Name: |

SET – 9

1. Julia is Botanist who is studying the Iris genus, She has collected the data of different the sepal length, sepal width, petal length, and petal width of various Iris flowers and wants to classify the flowers into their respective species based on their physical characteristics.
Julia decides to compare the performance of different machine learning algorithms for this task. She splits her data into a training set and a test set and trains several models, including Decision tree classifier, Logistic Regression, KNN classifier. Julia wants to the performance measures based on accuracy and speed of execution. Help her do the comparison of the classification algorithms
2. For a given set of training data examples stored in a .CSV file, implement and demonstrate the Candidate-Elimination algorithm to output a description of the set of all hypotheses consistent with the training examples.

| Example | Citations | Size | In Library | Price | Editions | Buy |
|---------|-----------|--------|------------|------------|----------|-----|
| 1 | Some | Small | No | Affordable | Few | No |
| 2 | Many | Big | No | Expensive | Many | Yes |
| 3 | Many | Medium | No | Expensive | Few | Yes |
| 4 | Many | Small | No | Affordable | Many | Yes |

3. Develop a Python code for implementing Polynomial regression and show its performance
4. Develop a Python code for implementing the KNN algorithm with an example.

| AIM & ALGORITHM (20) | PROGRAM (40) | DEBUGGING(10) | OUTPUT (20) | GITHUB UPLOAD (10) | TOTAL (100) |
|----------------------|--------------|---------------|-------------|--------------------|-------------|
| | | | | | |



SIMATS ENGINEERING

SAVEETHA INSTITUTE OF MEDICAL AND TECHNICAL SCIENCES

CHENNAI-602105



Engineer to Excel

| | |
|-------------------------|---------------------|
| Register Number: | Name: |
| Course Code: | Course Name: |

SET – 10

- How is the Perceptron algorithm applied to the Iris flower classification problem?
Rani is a botanist who is studying the Iris genus. She has collected data on the sepal length, sepal width, petal length, and petal width of various Iris flowers and wants to classify the flowers into their respective species based on their physical characteristics. Anna decides to use the Perceptron algorithm for this task
- Implement a Python program for the most specific hypothesis using Find-S algorithm for the following given dataset and show the output:

| Example | Citations | Size | In Library | Price | Editions | Buy |
|---------|-----------|--------|------------|------------|----------|-----|
| 1 | Some | Small | No | Affordable | Few | No |
| 2 | Many | Big | No | Expensive | Many | Yes |
| 3 | Many | Medium | No | Expensive | Few | Yes |
| 4 | Many | Small | No | Affordable | Many | Yes |

- Develop a Python code for implementing to compare Linear and Logistic regression and show its performance
- Develop a Python code for implementing the EM algorithm with an example.

| AIM ALGORITHM & (20) | PROGRAM (40) | DEBUGGING(10) | OUTPUT (20) | GITHUB UPLOAD (10) | TOTAL (100) |
|----------------------|--------------|---------------|-------------|--------------------|-------------|
| | | | | | |