DATA SCIENCE

UNIT 1- Introduction to Data Science

1.1 Overview of Data Science

a) Define data with examples.
b) Define Information with examples.
c) Concept of Data Science
d) Usage of Data Science
e) Main drivers of data science
f) Distinct capabilities of Data Science (Baldassarre, 2016)
g) Definition of Data Mining
h) Venn Diagram for Data Science
i) Concept and aim of Business Intelligence
j) Definition of Common terms in Data Science with examples– Artificial Intelligence, Machine learning, Training and testing sets, Data variables, Types of errors
k) Types of Machine learning
l) Definition and objective of Dimensionality reduction.
m) Explanation of sensitivity and specificity
n) Applications of Data Science with examples

1.2 Data Science Activities

a) Explanation of three dimensions of Data Science with illustration

1.3 Sources of Data

a) Explanation of different sources of data
b) Explanation with examples of Data types- Quantitative and Qualitative
c) Subtypes of Quantitative data with examples
d) Types of Data Shapes with examples
e) Explanation of 5Vs of Data with examples

1.4 Descriptive Statistics

a) Definition of Value, Standard deviation, Mean, median, mode with examples
b) Probability Theory
c) Definition of Probability with example
d) Definition and examples of Random event, mutually exclusive events, mutually independent events.
e) Definition and examples of Conditional Probability

f) Definition, examples and explanation of Probability density function
g) Explanation of Probability Distribution
h) Explanation, examples of Normal Distribution with graph
i) Explanation, examples of Binomial Distribution with graph
j) Explanation, examples, equation of Poisson Distribution with graph
k) Explanation of Bayesian Statistics
l) Explanation of Bayes Theorem with equation
m) Bayesian Statistical method
n) Example of Bayesian Statistics- Helmenstine's (2017) drug test analysis

Learning Objectives

- Explain the meaning of data science.
- Define the basic terms used in data science.
- Understand the applications of data science
- Identify and explain the typical sources of data.
- Understand the types and shapes of data.
- Explain probability distributions and Bayesian statistics

UNIT 2- Use Cases and Performance Evaluation

2.1 Data Science Use Cases (DSUCs)

a) Identifying DSUCs and their Value propositions
b) Identification of an Organizations's Use Cases Illustration
c) Value Propositions of applying data science tools – Three figures
d) Learning the Data Set and Prediction Model
e) Making Predictions and Decisions
f) Machine Learning Canvas (Dorard, 2017)

2.2 Performance Evaluation

a) Model-Centric Evaluation: Performance Metrics
b) Classification model evaluation metrics
c) Definition and equation of Accuracy
d) Confusion Matrix figure
e) Receiver Operator Characteristic Curve
f) Regression Model Evaluation Metrics

g) Definition, Equation and examples of Absolute Error, Relative Error, Mean absolute percentage error, Square error, Mean square error, Mean absolute error, Root mean square error
h) Business-Centric Evaluation: The Role of KPIs
i) Characteristics of effective KPIs
j) Examples of KPIs
k) Cognitive Bias with figure
l) Relevant cognitive biases
m) Explanation of Common Cognitive and Motivational Biases
n) Explanation of De-biasing techniques

Learning Objectives

- Understand the importance of a use case for business.
- Learn to identify use cases.
- Describe the steps to develop a predictive model for a specific use case.
- Determine metrics to evaluate the performance of a predictive model.
- Identify different cognitive biases which influence decision making process
- Understand and explain the role of KPIs in business centric evaluations.

UNIT 3- Data Preprocessing

3.1 Transmission of Data

a) Data Transmission Methods
b) Serial and Parallel Digital Data transmission
c) Synchronous and Asynchronous transmission

3.2 Data Quality, Cleansing, and Transformation

a) Definition, examples and explanation of Outliers, true outliers, fake outliers
b) Explanation with examples of methods to resolve missing values and outliers
c) Duplicate records
d) Redundant and Irrelevant variables and how to deal with them
e) Definition, example, equation and explanation of Correlation Coefficient
f) Detailed explanation of data transformation methods

3.3 Data Visualization

a) What is data visualization? (Runkler, 2012)
b) Definition, examples, graphs and explanation of data visualization types (Histogram, Scatter Plots, Geomaps, Area Charts, Bar Charts, Pie charts, Combo charts, Bubble Charts, Heat Maps)

Learning Objectives

- Identify and explain different data transmission methods and techniques.
- Apply methods to handle missing values and outliers in a dataset.
- Apply correlation analysis to a dataset.
- Learn the application of data transformation methods
- Learn the application of data visualization tools.

UNIT 4- Processing of Data

4.1 Stages of Data Processing

a) Definition of data and information with examples.
b) Definition, example and explanation of Data Processing
c) Applications of data processing
d) Benefits of data processing
e) Detailed explanation of stages of data processing with figure, examples
f) Detailed explanation of Steps in Data Analysis

4.2 Methods and Types of Data Processing

a) Detailed explanation of methods of data processing- manual, mechanical and electronic with examples
b) Types of Electronic Data Processing

4.3 Output Formats of Processed Data

a) Processed data format criteria
b) Processed data format forms
c) Explanation of Software specific data formats (XLS, CSV, XML, JSON, Protobuf, Apache Parquet, SQL)

Learning Objectives

- Explain the concepts of data, information and data processing
- Describe the stages and cycle of data processing.

- Explain different methods and types of data processing.
- Identify various output forms and file formats for processes data.

UNIT 5- Selected Mathematical Techniques

5.1 Principal Component Analysis

a) Two modelling approaches for prediction- Regression and Classification
b) Aim of Regression and Classification
c) Explanation of Principle Component Analysis
d) Aim and Example of Principle Component Analysis
e) PCA Algorithm

5.2 Cluster Analysis

a) Definition of Clustering
b) Clustering Approaches
c) Explanation, Examples and Algorithm of - K-Means
d) Explanation, Examples and Algorithm of Hierarchical Clustering (Top-down and Bottom-up)

5.3 Linear Regression

a) Objective of Regression model
b) Definition and example of Linear Regression (Runker, 2012)
c) Explanation, example and equation of Linear Regression Model
d) Example, Algorithm and equation of Simple Linear Regression model
e) Example, Algorithm and equation of Multiple Linear Regression model

5.4 Time-Series Forecasting

a) Explanation of Forecasting model
b) Examples of time series data
c) Analysis of time series data
d) Autoregressive Method
e) Concept of Stationary with examples
f) Explanation and equation of Autoregressive model
g) Explanation and equation of Moving Average model
h) Explanation, examples and equation of Autocorrelation, Partial Autocorrelation, Autoregressive Integrated Moving Average (ARIMA) Model
i) Algorithm of ARIMA
j) Seasonal Autoregressive Integrated Model (SARIMA)

5.5 Transformation Approaches

a) Definition, objective and explanation of dataset transformation

b) Logarithm Transformation
c) Power Law Transformation
d) Reciprocal Transformation
e) Radial Transformation
f) Discrete Fourier Transform

Learning Objectives

- Learn to apply principal component analysis to data.
- Learn to perform cluster analysis on a dataset.
- Describe the linear regression model and compute its coefficients.
- Describe the important features of time-series data.
- Explain popular models for forecasting future values in time-series data.
- Identify common approached for dataset transformation.

UNIT 6- Selected Artificial Intelligence Techniques

6.1 Support Vector Machines

a) Explanation of Support Vector Machines with graphics
b) Kernel Trick

6.2 Artificial Neural Networks

a) Purpose of Artificial Neural Network
b) Artificial Neural Network Architecture
c) Definition and explanation of Activation Function
d) Typical Activation Functions
e) Feed forward networks
f) Back Propagation Algorithm
g) Forward Pass Phase
h) Gradient descent
i) Backward pass phase
j) Recurrent Networks and Memory Cells
k) Reinforcement Learning
l) Comparison of learning types: supervised, unsupervised, reinforcement
m) Markov decision process

6.3 Further Approaches

a) Genetic Algorithm
b) Fuzzy Logic
c) Naïve Bayes Classification

Learning Objectives

- Understand data classification by support vector machines.
- Explain feedforward neural network structure.
- Understand the back propagation algorithm in neural networks.
- Learn to develop an artificial neural networks prediction model.
- Understand recurrent networks and reinforcement learning.
- Explain genetic algorithms, fuzzy logic, and Naïve Bayes classification.