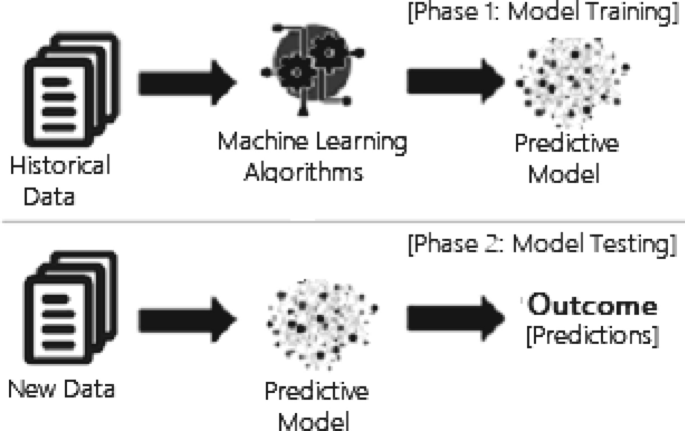
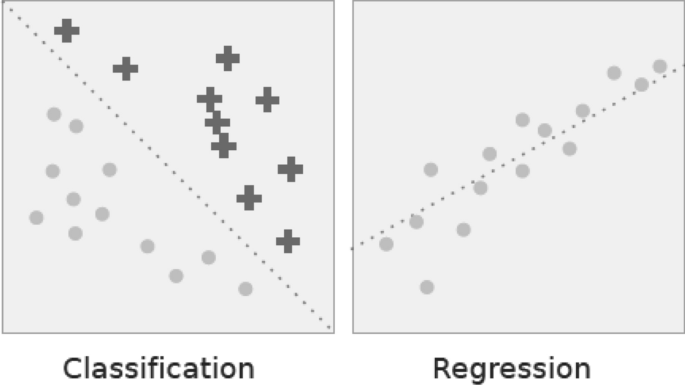
Machine Learning Tasks and Algorithms

In this section, we discuss various machine learning algorithms that include classification analysis, regression analysis, data clustering, association rule learning, feature engineering for dimensionality reduction, as well as deep learning methods. A general structure of a machine learning-based predictive model has been shown in where the model is trained from historical data in phase 1 and the outcome is generated in phase 2 for the new test data.

[](https://link.springer.com/article/10.1007/s42979-021-00592-x/figures/3)

A general structure of a machine learning based predictive model considering both the training and testing phase

[](https://link.springer.com/article/10.1007/s42979-021-00592-x/figures/6)

Classification vs. regression. In classification the dotted line represents a linear boundary that separates the two classes; in regression, the dotted line models the linear relationship between the two variables

* *Cybersecurity and threat intelligence:* Cybersecurity is one of the most essential areas of Industry 4.0. [[114](https://link.springer.com/article/10.1007/s42979-021-00592-x#ref-CR114)], which is typically the practice of protecting networks, systems, hardware, and data from digital attacks [[114](https://link.springer.com/article/10.1007/s42979-021-00592-x#ref-CR114)]. Machine learning has become a crucial cybersecurity technology that constantly learns by analyzing data to identify patterns, better detect malware in encrypted traffic, find insider threats, predict where bad neighborhoods are online, keep people safe while browsing, or secure data in the cloud by uncovering suspicious activity. For instance, clustering techniques can be used to identify cyber-anomalies, policy violations, etc. To detect various types of cyber-attacks or intrusions machine learning classification models by taking into account the impact of security features are useful [[97](https://link.springer.com/article/10.1007/s42979-021-00592-x#ref-CR97)]. Various deep learning-based security models can also be used on the large scale of security datasets [[96](https://link.springer.com/article/10.1007/s42979-021-00592-x#ref-CR96), [129](https://link.springer.com/article/10.1007/s42979-021-00592-x#ref-CR129)]. Moreover, security policy rules generated by association rule learning techniques can play a significant role to build a rule-based security system [[105](https://link.springer.com/article/10.1007/s42979-021-00592-x#ref-CR105)]. Thus, we can say that various learning techniques discussed in Sect. [Machine Learning Tasks and Algorithms](https://link.springer.com/article/10.1007/s42979-021-00592-x#Sec5), can enable cybersecurity professionals to be more proactive inefficiently preventing threats and cyber-attacks.
* *Internet of things (IoT) and smart cities:* Internet of Things (IoT) is another essential area of Industry 4.0. [[114](https://link.springer.com/article/10.1007/s42979-021-00592-x#ref-CR114)], which turns everyday objects into smart objects by allowing them to transmit data and automate tasks without the need for human interaction. IoT is, therefore, considered to be the big frontier that can enhance almost all activities in our lives, such as smart governance, smart home, education, communication, transportation, retail, agriculture, health care, business, and many more [[70](https://link.springer.com/article/10.1007/s42979-021-00592-x#ref-CR70)]. Smart city is one of IoT’s core fields of application, using technologies to enhance city services and residents’ living experiences [[132](https://link.springer.com/article/10.1007/s42979-021-00592-x#ref-CR132), [135](https://link.springer.com/article/10.1007/s42979-021-00592-x#ref-CR135)]. As machine learning utilizes experience to recognize trends and create models that help predict future behavior and events, it has become a crucial technology for IoT applications [[103](https://link.springer.com/article/10.1007/s42979-021-00592-x#ref-CR103)]. For example, to predict traffic in smart cities, parking availability prediction, estimate the total usage of energy of the citizens for a particular period, make context-aware and timely decisions for the people, etc. are some tasks that can be solved using machine learning techniques according to the current needs of the people.

import pandas as pd

import numpy as np arrayDataset=pd.read\_csv('mainSimulationAccessTraces.csv')  
x=Dataset.iloc[:,:-2].values  
y=Dataset.iloc[:,12].values

from sklearn.impute import SimpleImputer  
imputer=SimpleImputer(missing\_values=np.nan,strategy='constant',verbose=0)  
imputer=imputer.fit(x[:,[8]])  
x[:,[8]]=imputer.transform(x[:,[8]])imputer1=SimpleImputer(missing\_values=np.nan,strategy='mean',verbose=0)  
imputer1=imputer1.fit(x[:,[10]])  
x[:,[10]]=imputer1.transform(x[:,[10]])from sklearn.preprocessing import LabelEncoder  
labelencoder\_X = LabelEncoder()  
for i in range(0,10):  
 x[:,i] = labelencoder\_X.fit\_transform(x[:,i])  
x=np.array(x,dtype=np.float)y=labelencoder\_X.fit\_transform(y)

from sklearn.preprocessing import StandardScaler  
 sc = StandardScaler()  
 x\_train = sc.fit\_transform(x\_train)  
 x\_test = sc.transform(x\_test)