

Task 4

1- Summary of chapter 1 from Hands-on Machine Learning:

The chapter dives into the world of machine learning (ML), a field of computer science that sits under the umbrella of artificial intelligence (AI). It discusses and defines machine learning and its use cases and applications.

It also explains the different types of machine learning systems. The first way of classifying machine learning systems is via labeling of data, there is supervised, unsupervised, and semi-supervised learning. There are also other types like self-supervised learning (which is partially unsupervised) and reinforcement learning.

This chapter also explains another way of classifying machine learning systems, and it's by the data processing, there is batch learning which throws a dataset into the machine learning model and then adds new data every once and a while, and there is online learning which aims to add data into the machine learning model as it arrives.

And the last way of classifying the machine learning models is by how they generalize, and there exists two main approaches to generalization: instance-based learning and model-based learning.

This book's chapter also delves into the main challenges in machine learning which include:

- 1- Insufficient Quantity of Training Data
- 2- Nonrepresentative Training Data
- 3- Poor-Quality Data
- 4- Irrelevant Features
- 5- Overfitting the Training Data
- 6- Underfitting the Training Data

Finally, the last topic for this chapter is testing and validation and some of the challenges that come with them.

Task 4

2- What is Machine Learning?

Machine Learning is building a system that relies on data, this system will automate tedious tasks that were previously done by real people. It does that by imitating how humans learn gradually improving its accuracy.

2- Machine Learning application.

Image identification software that uses computer vision to identify documents, classify them for their intended purpose and then extract the most relevant info for each document.

For this app to work we will need a huge dataset of labeled official papers for different documents depending on our use case.

After training our model we can deploy it online to make the processes that include official papers a breeze.

