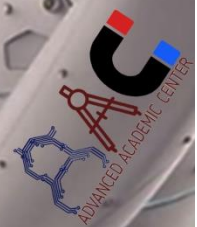




MACHINE LEARNING IN A NUTSHELL

K.HARSHINI
P.REETHIKA RAO



MACHINE LEARNING IN A NUTSHELL

-K.HARSHINI

-P.REETHIKA RAO

FORWORD

Each chapter in the book is self-contained. There is considerable flexibility in the choice of topics. This book includes material on application to Machine Learning, which is useful to solve real-time problems. The algorithms presented in this book are described in sufficient detail to enable students to use them in their programs. This book is easy to follow and fills a void none other does, and it contains algorithms, codes, their mathematical explanation, and visual representations. This book is suitable for self-study and is accessible to students in a variety of technical disciplines.



- Dr.Jandhyala Narayana Murthy

Director, GRIET

‘MACHINE LEARNING IN A NUTSHELL’ is a book that includes real-world applications for all foundation models. It also includes unique and distinctive chapters on Machine learning, Artificial Neural Networks, Convolutional Neural Networks. It also integrates well-documented python codes and their explanations which turn’s as an asset to the beginners. The book gives insight into what Machine learning involves and how it can impact the way you can weaponize data to gain incredible outputs.



- Dr.J.Praveen

Principal, GRIET

PREFACE

Machine Learning is slanting far and wide into almost everything. While we shop, search, work, and travel, our decisions, and our choices are learned by a machine to provide us with finer experiences. Machine Learning in A Nut Shell furnishes the reader with the entire pre-requisites essential to apply ML on a real-life complication. The implications of ML on industries, professions, and workforce are considered miraculous. This book comprises of clear explanations of core algorithms, their respective python codes, and the code explanations, along with the visual examples to make it convenient and engaging to follow for absolute beginners.

Starting from the importance of data in machine learning the book provides all the minute details of algorithms, their corresponding codes, and mathematics. The book contains some simple yet potential ideas which will seize the reader's mind and keep them absorbed. The book includes basic block diagrams and graphs that help in easy interpretation.

The whole book is bestowed on the best interest of the reader. It helps one understand machine learning algorithms simply and completely and assists the person in solving a challenging problem to make a better living.

ABOUT THE AUTHORS

I am K.Harshini, a student at Gokaraju Rangaraju Institute of Engineering and Technology. I am in my third year of communication engineering. I am a core committee member of the Advanced Academic center. Previously, I worked on building a vending machine. I am a voracious reader, a keen observer, and an intense thinker. My love for art is immense. I wrote this book along with Reethika to make Machine Learning fun, interesting, and easily understandable to everyone. The book consists of clear explanations of all machine learning algorithms, related python codes, and pictures showing the results. The book is a perfect guide for those who are starting machine learning.





I am P.Reethika Rao, a student at Gokaraju Rangaraju Institute of Engineering and Technology. I am in my third year of Electrical and Electronics Engineering. I am a student entrepreneur holding an event planning company, i love writing, photography and I'm an avid speaker! I have some good leadership qualities. I'm currently looking forward to develop my technical skills. I'm also passionate about ML and AI. This book makes Machine Learning fun and easy. My partner and I used techniques that would help beginners understand ML clearly.

ACKNOWLEDGEMENTS

It has been a wonderful experience for both of us to work on this book. We take this opportunity to thank all those who have been directly or indirectly related to this book. First and foremost, we would like to express our deep and sincere gratitude to our research supervisor Dr.Rama Murthi Suri, Dean of AAC Advanced Academic Center, for encouraging us to write this book and helping us out with the mathematics for the algorithms.

We are grateful to Y.Chethan Reddy for his valuable assistance and guidance throughout the book. We thank B.Charith Chandra Sai for his estimable contribution to the book. We have benefited from a short course on Machine Learning taught by Kirill Eremenko. We would like to thank reviewers for reviewing and providing insightful suggestions during the development of the script. We thank our parents and family for trusting and supporting us throughout the whole journey.

TABLE OF CONTENTS

1. ARTIFICIALINTELLIGENCE.....	6
2. DATA.....	8
3. DATAPREPROCESSING.....	9
3.1. LIBRARIES	
3.2. IMPORTING DATA SETS	
3.3. DIVIDING DATA SET INTO INDEPENDENT AND DEPENDENT VARIABLES	
3.4. MISSING DATA	
3.5. CATEGORICAL DATA	
3.6. FEATURE SCALING	
3.7. TYPES OF DISTANCES	
3.8. TYPES OF LEARNING IN MACHINE LEARNING	
4. REGRESSION.....	28
4.1. LINEAR REGRESSION	
4.1.1.1. SIMPLE LINEAR REGRESSION	
4.1.1.2. MULTIPLE LINEAR REGRESSION	
4.2. POLYNOMIAL REGRESSION	
4.3. SUPPORT VECTOR REGRESSION	
4.4. DECISION TREE REGRESSION	
5. CLASSIFICATION.....	74
5.1. LOGISTIC REGRESSION	
5.2. K-NEAREST NEIGHBOURS CLASSIFICATION	
5.3. SUPPORT VECTOR MACHINE	
5.4. KERNEL SVM	
5.5. NAÏVE BAYES CLASSIFICATION	
5.6. DECISION TREE CLASSIFICATION	
5.7. RANDOM FOREST CLASSIFICATION	
5.8. CONFUSION MATRIX	
6. CLUSTERING.....	39
6.1. K-MEANS CLUSTERING	
6.2. HIERARCHICAL CLUSTERING	
7. DEEPLARNING.....	153
7.1. AIRTIFICIAL NEURAL NETWORKS	
7.2. CONVOLUTION NEURAL NETWORKS	
8. DIMENSIONALITYREDUCTION.....	181
8.1. FEATURE SELECTION	
8.2. FEATURE EXTRACTION	
8.2.1.1. PRINCIPLE COMPONENT ANALYSIS	
8.2.1.2. LINEAR DISCRIMINANT ANALYSIS	

CHAPTER 1

ARTIFICIAL INTELLIGENCE

Artificial Intelligence(AI) refers to the demonstration of human intelligence in machines when they are programmed to think and perform human traits and actions. Artificial intelligence makes it possible for machines to learn from experience, adjust to new inputs, and perform human-like tasks. The term may also include actions performed by machines that exhibit traits associated with a human mind such as learning, problem-solving, and decision making.

When people refer to the term artificial intelligence, they often refer it to robots due to the fictional stories woven by big-budget films and novels. The fact is Artificial Intelligence is not just limited to robots, but there are many fields where AI finds its applications. As technology is progressing, previous standards that defined artificial intelligence became outdated. For example, machines that execute basic functions and tasks can no longer be considered to define artificial intelligence, as these simple functions are now taken for granted as a built-in computer function.

Further applications of artificial intelligence may also include computers that play chess and self-driving cars. As each action of the machine embodied with artificial intelligence will impact the result, each of these machines must weigh the consequences of any action they take. For example, consider Self-driving cars. In self-driving cars, the system must account for all the external consequences and compute itself to act in a way that prevents an accident.

From basic chat-bots to humanoid robots, which can have a country's citizenship, Artificial intelligence has crossed all the barriers of the resemblance of human intelligence in machines. AI extends its applications to most of the fields like medicine, business, agriculture, space exploration, etc and there is no doubt that AI will take control over the existing technologies. Clearly, Artificial Intelligence will be the future of the human era. Artificial intelligence has many branches, such as

Machine learning, deep learning, Neural networks, Natural language processing, etc.

Machine Learning is a branch of Artificial Intelligence that trains a machine, utilizing mathematical models for decision-making. The ideal characteristic of machine learning is its ability to justify and take actions accordingly to have the best chance of achieving the desired goal. Machine learning is continuously evolving to benefit many industries.

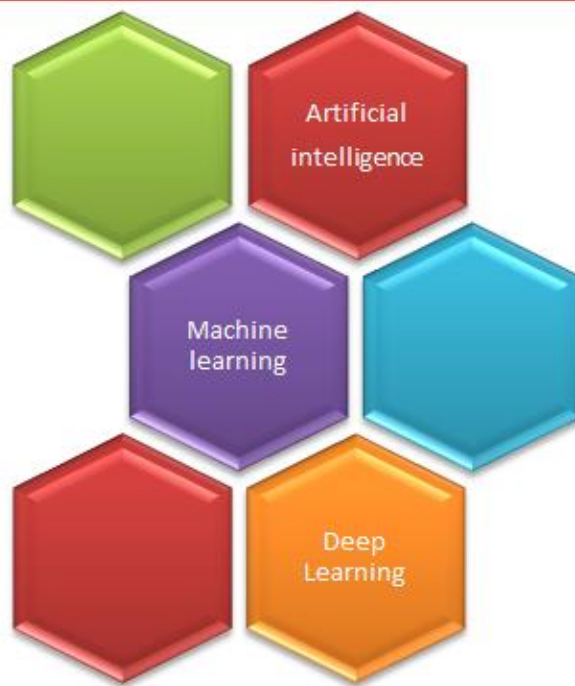


Fig 1.1. Artificial intelligence

CHAPTER 2

DATA

In today's world, everything is automated. All human actions are being replaced by machines. Machine learning plays an important role in the process of automation.

For example, consider, Facebook suggesting friends and Youtube recommending videos, based on the user's preference, this scenario is carried out by machine learning. The process behind this scenario is Facebook or Youtube collects the past data and feeds it into a machine, which will incorporate the data to recommend future preferences.

Machine Learning entirely depends on data. Without Data, it is impossible to train a Machine learning model. Data is the most essential aspect of Machine Learning as it enables a machine to learn and work properly. Many machine learning projects have failed due to a lack of good data set despite having the best algorithms.

Data is the collection of facts and figures that are gathered and formatted in a special way for analysis. Data is the most powerful and valuable tool any organization can have.

Data is raw, unorganized figures that need to be processed. The data which is processed and organized is used for the analysis of any real-time situations. Organizing the whole data creates a dataset that is further used for analysis. Data, when used to its maximum potential, will have a prodigious impact on an organization's long term success.

ML helps in analyzing and processing the data and gives predictions about our future. Machine learning (ML) focuses on analyzing and interpreting patterns and structures in data to enable learning, reasoning, and decision making outside of human interaction.

ML comprises of several algorithms. The data is analyzed by using these algorithms and can make data-driven outputs and recommendations by using the input data. The algorithm has the ability to make some corrections and incorporate that information to improve its further decision making.

ML has many factors to make it accessible, one of them is data handling. Machine Learning can be multidimensional. In Machine learning, the work is constantly done on data. Huge amount of data is taken for training and testing. This data handling and organizing make Machine learning an important key in automation.

CHAPTER 3

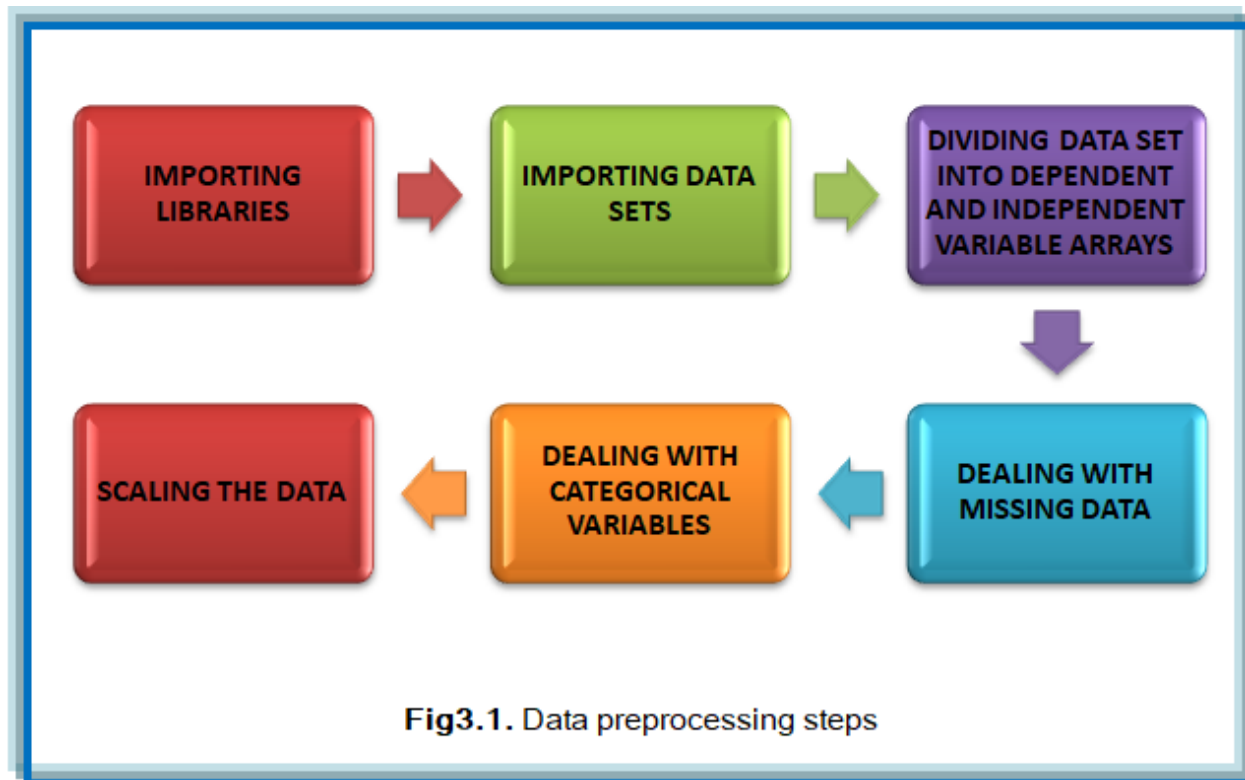
DATA PREPROCESSING

Applying machine learning algorithms to the data sets does not give the best outcomes. Processing the data before applying any machine learning algorithm to it is the most important aspect. The output of the machine learning model depends on how well the data is processed. Data processing has several steps, which are explained below in detail.

In the first step of dealing with data in Machine Learning, creating the data sets plays an important role. A data set is a structured collection of data. After the collection of data and organizing it, the data has to be arranged into a dataset for the creation of machine learning models. Datasets may include any number of figures and characters arranged in rows and columns.

In creating the best machine learning model, gathering the data isn't just enough, but classifying and labeling of data plays a major role. This makes the dataset accurate enough to give a realistic vision.

A data set consists of several variables; machine learning algorithms are applied to a data set to predict the value of variable using factors that affect it. The value being predicted is called the dependent variable, and the factors it depends on are called independent variables.



Most Machine learning algorithms possess two ways to precede, learning, and predicting. Therefore the data set is divided into:

1. Training Dataset
2. Testing Dataset

Training dataset:

The machine learning algorithm needs to be trained by feeding it data. Training data will help the algorithm or the machine to connect the patterns in the data to give the best output. This training data comprises the majority of the dataset.

Testing Dataset:

The testing dataset is used to evaluate the performance of the machine learning model which has already been trained. The reliability of the model is checked on the training dataset. The testing dataset is basically used for interfacing how well the model performs. The most important point that has to be noted here is training should never be done on the testing set.

A test dataset is independent of the training dataset, but it has the same probability distribution as the training set. If the machine learning model is trained with a training set and tested with the same set, the best results are obtained, but not the accurate one. Thus it is important to test the machine with the test set.

3.1. LIBRARIES:

When working with datasets in machine learning models, the algorithm can be carried out in Python, or R. Python is employed to apply machine learning algorithms to the data. Python offers concise and readable codes that make complex algorithms in machine learning easily accessible. Python is the combination of consistent syntax, shorter development time, and is flexible; this makes it well suitable for any kind of complex machine learning model. It is a comprehensive set of libraries that are the greatest asset for python.

Libraries are a set of well- structured and well-tested functions to empower the best quality in codes. A vast set of libraries makes it easier to perform functions on machine learning models.

The most important libraries in machine learning are:

1. NumPy
2. Pandas
3. Matplotlib
4. Scikit-Learn

NumPy:

Numpy is a very popular Python library used for fundamental scientific computations. It is vastly used for linear algebra, Fourier transform, and random number capabilities.

NumPy is popularly applicable for matrix and multidimensional array processing in machine learning models.

Pandas:

Pandas is the most popular machine learning library which is used for data analysis. It is also used for data manipulation. A lot of functionality and extensive documentation are the benefits of using the Pandas library. It is indirectly related to machine learning but has many inbuilt functions like grouping filtering and combining data.

Matplotlib:

Matplotlib is the machine learning library extensively used for Data visualization. It is used for the data visualizing and plotting the patterns and features. It is a 2D plotting library used for creating 2D graphs and plots.

Scikit-Learn:

Scikit-learn is one of the most prominent machine learning libraries used for supervised and unsupervised learning algorithms. It is used for classic machine learning algorithms and also for data mining and data analysis.

3.2. IMPORTING THE DATASETS:

While programming in python, it is important to set the working directory so that the code will execute on the data set. To set the working directory, first select the folder where the data set is present and save the python file in the same folder. After setting the working directory, import the data set on which the machine learning algorithm is to be applied.

3.3. DIVIDING DATA SET INTO INDEPENDENT AND DEPENDENT VARIABLE ARRAYS:

After importing the data set, dividing the data set into training and testing sets is important.

Data preprocessing in python:

Table 3.1

Hotel	Rating	Cost for 2 people for 1 day
Trident	4.3	15,530
ITC Kakatiya	4.1	14,340
Taj Falaknuma	4.7	30,000
Red Fox Hotel	3.5	3,389
Hotel sahara star	4.6	26,890
The Leela ambience	3.8	8,567
Itc Grand Chola	4.3	11,678
Novotel	4.2	10,457
ITC Maratha	3.8	8,987
The LaliT Ashok	3.9	9,876
Sheraton	3.2	6,587
Taj Krishna	4.2	13,678
Gopuram hotel	3.1	4,098
Sadguru hotel	3.2	6,567
Amish Hotel	2.8	3,456
Gyana hotel	3.7	6,657
Blisshotel	4.7	30,456
Shasam hotel	2.5	4,543
Dyana hotel	3.5	4,567

Table 3.1 gives information about hostel ratings and the cost for two people. To apply any machine learning algorithm on this data set, we should first preprocess it.

The code in python for data preprocessing is as follows:

```

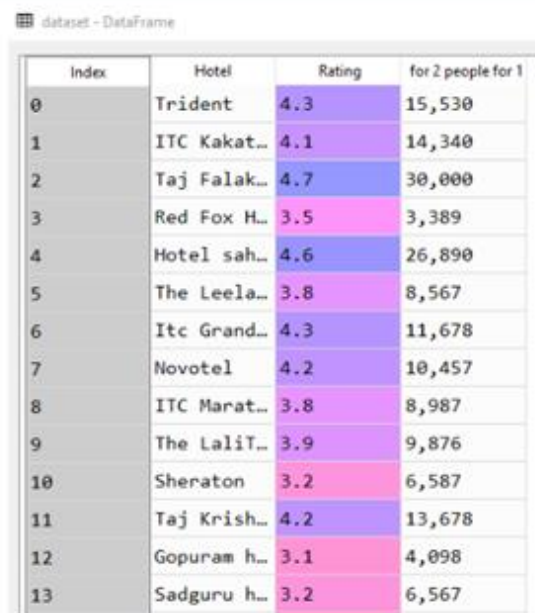
1 import numpy as np
2 import matplotlib.pyplot as plt
3 import pandas as pd
4
5 # Importing the dataset
6 dataset = pd.read_csv('dp.csv')
7 a = dataset.iloc[:, :-1].values
8 a=pd.DataFrame(a);
9 b = dataset.iloc[:, -1].values
10 b=pd.DataFrame(b);
11
12 # Splitting the dataset into the Training set and Test
13 from sklearn.model_selection import train_test_split
14 a_train, a_test, b_train, b_test = train_test_split(a, b, test_size = 0.1, random_state = 0)
15 |

```

Code 1. Data preprocessing in python

The first three lines of code import NumPy, matplotlib, and pandas libraries.

The sixth line imports the required data set.



Index	Hotel	Rating	for 2 people for 1
0	Trident	4.3	15,530
1	ITC Kakat...	4.1	14,340
2	Taj Falak...	4.7	30,000
3	Red Fox H...	3.5	3,389
4	Hotel sah...	4.6	26,890
5	The Leela...	3.8	8,567
6	Itc Grand...	4.3	11,678
7	Novotel	4.2	10,457
8	ITC Marat...	3.8	8,987
9	The LaliT...	3.9	9,876
10	Sheraton	3.2	6,587
11	Taj Krish...	4.2	13,678
12	Gopuram h...	3.1	4,098
13	Sadguru h...	3.2	6,567

IMPORTED DATA SET

The code lines from seventh to tenth; divide the imported data set into independent and dependent variable arrays.

a_train - DataFrame			b - DataFrame	
Index	0	1	Index	0
8	ITC Marat...	3.8	0	15,530
18	Dyana hot...	3.5	1	14,340
14	Amish Hot...	2.8	2	30,000
16	Blisshotel	4.7	3	3,389
6	Itc Grand...	4.3	4	26,890
4	Hotel sah...	4.6	5	8,567
2	Taj Falak...	4.7	6	11,678
5	The Leela...	3.8	7	10,457
13	Sadguru h...	3.2	8	8,987
9	The LaliT...	3.9	9	9,876
7	Novotel	4.2	10	6,587
17	Shasam ho...	2.5	11	13,678
11	Taj Krish...	4.2	12	4,098
3	Red Fox H...	3.5	13	6,567

INDEPENDENT VARIABLE ARRAY

DEPENDENT VARIABLE ARRAY

The thirteenth and fourteenth lines split the independent and dependent variable arrays into testing and training sets.

b_train - DataFrame	
Index	0
8	8,987
18	4,567
14	3,456
16	30,456
6	11,678
4	26,890
2	30,000
5	8,567
13	6,567
9	9,876
7	10,457
17	4,543
11	13,678
3	3,389

b_test - DataFrame	
Index	0
10	6,587
1	14,340

DEPENDENT VARIABLE ARRAY DIVIDED INTO TESTING AND TRAINING SETS

a_train - DataFrame		
Index	0	1
8	ITC Haratha	3.8
18	Oyana hotel	3.5
14	Amish hotel	2.8
16	Bliss hotel	4.7
6	ITC Grand Chola	4.3
4	Hotel Sahara star	4.6
2	Taj Falaknuma	4.7
5	The Leela Ambience	3.8
13	Sadguru hotel	3.2
9	The Lalit Ashok	3.9
7	Novotel	4.2
17	Shasana hotel	2.5
11	Taj Krishna	4.2
3	Red Fox hotel	3.5

a_test - DataFrame		
Index	0	1
10	Sheraton	3.2
1	ITC Kakatiya	4.1

INDEPENDENT VARIABLE ARRAY DIVIDED INTO TRAINING AND TESTING SET

Now available

Get full version @Amazon Kindle



All Machine Learning in a nutshell

Hello
Select your address

[Mobiles](#) [Best Sellers](#) [Today's Deals](#) [Computers](#) [New Releases](#) [Amazon Pay](#) [Books](#) [Gift Ideas](#) [Customer Service](#) [Sell](#) [Baby](#) [AmazonBasics](#) [Coupons](#)

[Kindle e-Readers](#) [Free Kindle Reading Apps](#) [Kindle eBooks](#) [Kindle Unlimited](#) [Prime Reading](#) [Deals on Kindle eBooks](#) [Best Sellers](#) [Indian language eBooks](#) [Kindle Exam Central](#) [Kindle eTextbooks](#) [Content and devices](#) [Ki](#)



[Back to results](#)



Machine Learning in a nutshell Kindle Edition

by [Harshini Kavuri](#) (Author), [P.Reethika Rao](#) (Author) | Format: Kindle Edition

[See all formats and editions](#)

Kindle Edition
₹ 0.00 kindleunlimited

This title and over 1 million more available with Kindle Unlimited
₹ 100.00 to buy

Machine Learning is slanting far and wide into almost everything. While we shop, search, work, and travel, our decisions, and our choices are learned by a machine to provide us with finer experiences. Machine Learning in a nutshell furnishes the reader with the entire pre-requisites essential to apply ML on a real-life complication. The implications of ML on industries, professions, and workforce are considered miraculous. This book comprises of clear explanations of core algorithms, their respective

[Read more](#)

Length: 264 pages
Page Flip: Enabled
Available on these devices

Word Wise: Enabled
Language: English

Enhanced Typesetting: Enabled