BIG DATA FOR INDUSTRY 4.0: CHALLENGES AND APPLICATIONS

CRC Press
Taylor & Francis Group

Handbook of Artificial Intelligence for Smart City Development

Management Systems and Technology Challenges

Edited by
Sandhya Makkar
Gobinath Ravindran
Ripon Kumar Chakrabortty
Arindam Pal

Handbook of Artificial Intelligence for Smart City Development

This handbook explores smart cities of the future, provides an understanding of their development and management systems, and discusses the technology challenges. It also discusses why and how humankind can benefit from them.

Handbook of Artificial Intelligence for Smart City Development: Management Systems and Technology Challenges covers the whole journey of the development of smart cities, from the foundations to the usage of technologies, including the challenges and issues that policymakers and technologists may face during the concept, development, management, and implementation stage. The book also covers sustainable strategies and safety measures and offers real-life cases of advancements in manufacturing approaches for smart cities. The book includes upcoming AI technologies such as big data analytics, block chain, machine learning, fault diagnostics, and a lot more.

This handbook is intended to appeal to readers from industry and research. The book is also meant for academicians and students from across many disciplines. Moreover, this book is a medium for consultants, government agencies, and policymakers to grapple with topics and perspectives outside their lane.

Big Data for Industry 4.0: Challenges and Applications

Series Editors: Sandhya Makkar, K. Martin Sagayam, and Rohail Hassan

Industry 4.0 or the Fourth Industrial Revolution refers to interconnectivity, automation, and real-time data exchange between machines and processes. There is a tremendous growth in big data from Internet of Things (IoT) and information services, which drives the industry to develop new models and distributed tools to handle big data. Cutting-edge digital technologies are being harnessed to optimize and automate production, including upstream supply chain processes, warehouse management systems, automated guided vehicles, and drones. The ultimate goal of Industry 4.0 is to drive manufacturing or services in a progressive way to be faster, effective, and efficient, which can only be achieved by embedding modern day technology in machines, components, and parts that will transmit real-time data to networked IT systems. These, in turn, apply advanced soft computing paradigms such as machine learning algorithms to run the process automatically without any manual operations.

The new book series will provide readers with an overview of the state-of-the-art in the field of Industry 4.0 and related research advancements. The respective books will identify and discuss new dimensions of both risk factors and success factors, along with performance metrics that can be employed in future research work. The series will also discuss a number of real-time issues, problems, and applications with corresponding solutions and suggestions—sharing new theoretical findings, tools, and techniques for Industry 4.0, and covering both theoretical and application-oriented approaches. The book series will offer a valuable asset for newcomers to the field and practicing professionals alike. The focus is to collate the recent advances in the field, so that undergraduate and postgraduate students, researchers, academicians, and industry people can easily understand the implications and applications of the field.

Big Data for Entrepreneurship and Sustainable Development
Edited by Mohammed el Amine Abdelli, Wissem Ajili Ben Youssef, Ugur Ozgoker, and Imen Ben Slimene

Entrepreneurship and Big Data
The Digital Revolution
Edited by Meghna Chhabra, Rohail Hassan, and Amjad Shamim

Microgrids
Design, Challenges, and Prospects
Edited by Ghous Bakhsh, Biswa Ranjan Acharya, Ranjit Singh Sarban Singh, and Fatma Newagy

Handbook of Artificial Intelligence for Smart City Development
Management Systems and Technology Challenges
Edited by Sandhya Makkar, Gobinath Ravindran, Ripon Kumar Chakrabortty, Arindam Pal

For more information on this series, please visit: https://www.routledge.com/Big-Data-for-Industry-4.0-Challenges-and-Applications/book-series/CRCBDICA

Handbook of Artificial Intelligence for Smart City Development

Management Systems and Technology Challenges

Edited by
Sandhya Makkar, Gobinath Ravindran,
Ripon Kumar Chakrabortty, and Arindam Pal

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

Designed cover image: Shutterstock

First edition published 2024
by CRC Press
6000 Broken Sound Parkway NW, Suite 300, Boca Raton, FL 33487-2742

and by CRC Press
4 Park Square, Milton Park, Abingdon, Oxon, OX14 4RN

CRC Press is an imprint of Taylor & Francis Group, LLC

Library of Congress Cataloging-in-Publication Data

Names: Makkar, Sandhya, editor. | Ravindran, Gobinath, editor. | Chakrabortty, Ripon K. (Ripon Kumar), editor. | Pal, Arindam, editor.
Title: Handbook of artificial intelligence for smart city development : management systems and technology challenges / edited by Sandhya Makkar, Gobinath Ravindran, Ripon Kumar Chakrabortty, Arindam Pal.
Description: First edition. | Boca Raton : CRC Press, 2024. | Includes bibliographical references and index.
Identifiers: LCCN 2023009098 (print) | LCCN 2023009099 (ebook) | ISBN 9781032125923 (hardback) | ISBN 9781032125947 (paperback) | ISBN 9781003225317 (ebook)
Subjects: LCSH: Smart cities. | Artificial intelligence.
Classification: LCC TD159.4 .H3634 2024 (print) | LCC TD159.4 (ebook) | DDC 307.1/416028563--dc23/eng/20230517
LC record available at https://lccn.loc.gov/2023009098
LC ebook record available at https://lccn.loc.gov/2023009099

ISBN: 978-1-032-12592-3 (hbk)
ISBN: 978-1-032-12594-7 (pbk)
ISBN: 978-1-003-22531-7 (ebk)

DOI: 10.1201/9781003225317

Typeset in Times
by KnowledgeWorks Global Ltd.

Dear Miraya

My loving daughter, you are an incredible child. You are my inspiration, and I am here because of you. I dedicate this book to you.

For my parents Mr. Sat Paul Bajaj and Mrs. Pramila Bajaj.

Words can hardly describe my thanks and appreciation to you. You have been my source of inspiration, support, and guidance. You have taught me to be unique, determined, to believe in myself, and to always persevere. I am truly thankful and honored to have you as my parents.

Sandhya Makkar

I wish to convey my heartfelt thanks to my wife Rajanimol PR and to my kids Natasha RG Nair, and Suryadev RG Nair for being with me always. Also, I wish to convey my thanks to my parents T. Ravindran and M. Gowri who are instrumental in my life.

Gobinath Ravindran

The book is dedicated to my parents and lovely family.

Ripon Chakrabortty

I dedicate this book to my father Bani Prasad Pal and mother Bani Pal for their love, care, patience, and support in bringing me up.

Arindam Pal

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Contents

List of Figures and Tables

Editors

Dr. Sandhya Makkar is as a senior lecturer of Data Analytics in the Department of Digital Analytics and Technologies at Sheffield Hallam University, United Kingdom. Before that, she worked as the area head for the Department of Operations and Analytics at Lal Bahadur Shastri Institute of Management, Delhi, India. She has more than a decade and half years of academic and research experience in optimization, machine learning, and big data analytics. She secured first class throughout her academic career. She has authored and coauthored numerous research articles that are published in international journals, books, and conference proceedings. She has also been appointed as a book series editor with CRC Press, Taylor & Francis Group. She is involved in innovations and patent development. She can be contacted at bajajsandhya@gmail.com or visit the following website for more details: https://sites.google.com/view/sandhyamakkar/home

Dr. Gobinath Ravindran is presently serving as professor and head of the Department, Civil Engineering of SR University, Telangana, India and also a member of Center for Construction Materials and Methods division. He completed his PhD in the field of Environmental Geotechnology and Disaster Management from the Center for Disaster Management and Mitigation, VIT, Vellore, his ME in Environmental Management from the College of Engineering, Guindy, Anna University, and his BE in Civil Engineering from Bharathiar University. He has handled several undergraduate courses in civil and environmental engineering, including project management, smart materials and structures, geosynthetics, engineering mechanics and is presently handling elective courses related to disaster management and intellectual property rights in various universities. His research interests include sustainable materials, construction materials, landslide mitigation and management, soil bioengineering, engineering education, geotechnology, environmental geotechnology, sustainable development, sustainable construction technologies; he has handled several projects related to these domains on various levels. He has executed a good number of consultancy projects related geotechnical and environmental engineering domains. His research profile is available at www.researchid.co/gobinathdpi

Dr. Arindam Pal is a senior research Scientist at Data61 in the Commonwealth Scientific and Industrial Research Organisation (CSIRO), and a conjoint senior lecturer in the School of Computer Science and Engineering at UNSW, Sydney. His research interests are in algorithms, artificial intelligence, data science, optimization, and machine learning. He works on business and research problems at CSIRO and collaborates with faculty members of universities, both in Australia and abroad.

He earned his PhD in Computer Science from the Indian Institute of Technology Delhi. He has over 15 years of research experience in software companies like Microsoft, Yahoo!, and Novell. He has published academic papers in reputed conferences and journals and was granted patents in India, the USA, and Europe. He is a technical program member for several reputed conferences and a technical reviewer for many renowned journals. He is a leader in the area of algorithms, data science, optimization, and machine learning. He is a senior member of both ACM and IEEE.

Dr. Ripon K. Chakrabortty (Senior Member, IEEE) is an assistant professor in Systems Engineering and Decision Analytics at the School of Engineering and Information Technology, University of New South Wales (UNSW), Canberra, Australia. He is also the team leader and founder of The Decision Support and Analytics Research Group at the School of Engineering and Information Technology, UNW. He serves as the program coordinator for the Master of Decision Analytics and Master of Engineering at UNSW in Canberra. He has written two book chapters and over 170 technical journals and conference papers in multiple prestigious venues. His research interests cover a wide range of topics in operations research, project management, supply chain management, artificial intelligence, and information systems management. Many organizations, such as the Department of Defence and the Commonwealth Government of Australia, have funded his research programs.

Contributors

Pooja Ahuja
SVN University
Sagar, India

Jacob Alexander
Resources 4 Solutions Inc.
Delta, British Columbia, Canada

Suman Ananthanarayan
University of Mumbai
Mumbai, India

Debarati Basak
Carpediem EdPsych Consultancy LLP
Bengaluru, India

Anudeep-Sri Bathina
University of Massachusetts
Dartmouth, Massachusetts
and
Nikita Kalloorthottiyil Chacko (RECE)
Dufferin Corporate Centre
London, Ontario, Canada

Varun Chhabra
Quantum Age Solutions Ltd.
Middlesex, United Kingdom

T. Kalavathi Devi
Kongu Engineering College
Erode, India

Amol S. Dhaigude
Manipal Academy of Higher Education
Manipal, India

Riju Ghosh
Amity Business School
Noida, India

Mukta Goyal
Guru Nanak Dev Institute of Technology
Delhi, India

Ekata Gupta
Guru Nanak Dev Institute of Technology
Delhi, India

Manju Jose
Majan University College
Muscat, Oman

S.K. Kabilesh
Jai Shriram Engineering College
Tirupur, India

Soni Lanka Karri
University Brunei Darussalam
Bandar Seri Begawan, Brunei

R. Karthikeyan
Varadhaman College of Engineering
Hyderabad, India

Arshia Kaul
Carpediem EdPsych Consultancy LLP
Bengaluru, India

Vasundhara Kaul
Carpediem EdPsych Consultancy LLP
Bengaluru, India

V. Kavitha
Hindusthan College of Arts and Science
Coimbatore, India

Prabha Kiran
Westminster International University
Tashkent, Uzbekistan

Sinimole K.R.
Rajagiri Business School
Kochi, India

Vijayalakshmi A. Lepakshi
Reva University
Bengaluru, India

Manasvi Logani
Indraprastha College for Women, University of Delhi
New Delhi, India

Sandhya Makkar
Sheffield Hallam University
Sheffield, United Kingdom

Seyed Jalaleddin Mousavirad
University of Beira Interior
Covilhã, Portugal

Mooon Paiithannkar
Sinhgad Technical Education Society
Pune, India

Roli Raghuvanshi
University of Delhi
New Delhi, India

Rashmi Rai
Christ University
Bangalore, Karnataka, India

Prashanth C. Ranga
Conestoga Institute of Technology
Kitchener, Ontario, Canada

Gobinath Ravindran
SR University
Hyderabad, India

Ravinder Rena
Durban University of Technology
Durban, South Africa

A. Stephen Sagayaraj
Bannari Amman Institute of Technology
Erode, India

Tanushree Sharma
Manipal University
Jaipur, India

Abhishek Srivastava
Gopal Narayan Singh University
Jamuhar, India

B. Sundaravadivazhagan
University of Technology and Applied Science
AI Mussanah, Oman

Pankaj Tiwari
SVN University
Sagar, India

Prakash Kumar Udupi
Middle East College
Muscat, Oman

Asad Ullah
Middle East College
Muscat, Oman

M. Veeramani
Hindusthan College of Arts and Science
Coimbatore, India

1 Machine Learning for Smart City AI Systems

Manasvi Logani[1] *and Sandhya Makkar*[2]
[1]Plaksha University, Mohali, Punjab, India
[2]Sheffield Hallam University, Sheffield, United Kingdom

1.1 INTRODUCTION TO SMART CITIES

While stating the obvious things, smart cities comprise two words, "smart" and "cities". Thinking of the word smart, the first thing that comes to mind is a competent, self-sustaining and mature environment or person. Combining all these qualities with a technological perspective and the concept of cities is what is meant by the term smart cities.

More formally, smart cities are those that integrate the use of information and communication technology, also known as ICT, to improve the overall quality of life and optimize infrastructure sustainably while promoting economic growth.

These cities consist of a well-planned urban environment to create sustainable living for all. Pollution control, parking and traffic management, public security, energy conservation, healthcare etc. are some of the focuses of a smart city.

1.2 IMPORTANCE OF SMART CITIES

As the world is exponentially progressing towards technological advancement, it makes sense to leverage this for economic growth and better lifestyles. With this aim, there are numerous significances of smart cities. A few of these points are as follows:

1. Improved Healthcare Systems
   With an increase in automation, healthcare will become easily accessible with a more accurate prediction of diseases. With the use of previous data and certain tools, the classification of diseases will be easier and they might get diagnosed at an early stage, leading to be lifesaving. Also, doctor consultation will become more widely accessible through video conferencing appointments.
2. Better Safety Measures
   One of the main advantages of a smart city will be an extensive installation of CCTV cameras to collect data for incident detection, gunshot detection, license plate recognition, face recognition etc. All these facilities will reduce the crime rate considerably because the chances of getting caught will immensely get more than before.
3. Environmentally Friendly
   When it comes to human development, one of the most neglected aspects is the environment. Over the years, there has been a huge depletion of resources, air and water quality index etc. The main aim of a smart city is not just

DOI: 10.1201/9781003225317-1

development but also sustainable development. Therefore, to achieve this, air and water quality detectors will be installed to keep the pollution level in check, energy conservation techniques will be adopted, and a thorough check on resources such as electricity and water will be maintained to limit wastage.

4. Organized Traffic Systems
   With the advent of the advanced age, commuting from one place to another has become easier. The term "easier" here refers to in terms of the availability and accessibility of modes of transportation and the technology used. This is because, ironically, automobiles were created to reduce travel time but instead due to high traffic congestion, it has become difficult to cut travel time. Smart cities will focus on reducing this traffic to increase comfort, along with other facilities such as smart streetlights and traffic rules detectors.
5. Improved Infrastructure
   As the cities are growing older, their infrastructures consisting of buildings, roads, bridges etc. are getting weaker. In the smart cities, there will be an installation of certain sensors to collect data and report if there are any cracks or considerable damage to the infrastructure. With the help of this, it will become easier to prioritize development and even save lives in certain situations.

1.3 BACKGROUND OF STUDY

ICT is a discipline integrated with unified communications, telecommunications and computers for creating an infrastructure that enables modern computing. It includes radios, televisions, copiers, fax machines, telephones, satellites, multimedia, and internet which are all communication mediums in some or the other way.

Since the advent of ICT, there has been a gradual change in how the world is perceived. Initially, these technologies were used only for military and research purposes and then by technology enthusiasts and pioneer governments. It was only later when dynamic and innovative cities started involving these technologies. IT companies like IBM, Cisco, Siemens and General Electric started to employ their technology and investing in the improvement of cities (Barrera, 2020). Due to the widespread use of internet, computers, smartphones etc., there has been an enormous amount of data generated and even awareness; thus, citizens are also interested in deploying changes into a city to create a "smart city".

Over the years, the cities which incorporate technology into their management regulatory systems have been given many names. In 1987, the term *wired cities* (Dutton, Blumler, & Kraemer, 1987) was coined for a city with telecommunications between households and businesses. Later on, the concept of *cyber cities* (Graham & Marvin, 1999) was stated as a city with complex interdependence between transport and telecommunications, which link between urban and cyber cultures and economies and interdependently linked IT-based communities. Further, the term *digital city* (Ishida & Isbister, 2000) came into the picture and it "is defined as a city that integrates urban information (both achievable and real time) and creates public spaces for people living in the cities". In 2002, the term *intelligent cities* (Komninos, 2002) was referred to as "where the innovation processes meet the digital world and the applications of the information society". Next was *the sentient city* (Shepard, 2011), wherein it can hear and feel things but doesn't have any particular information; it can remember, correlate and anticipate using certain technology like sensors and algorithms. Now, coming to *information city* (Chourabi, et al., 2012), it collects information from localities and delivers it to the public via the internet.

The cities combined with the ICT technologies have been referred using various terms over the years, each with a different underlying concept, but share a focus on the effects of ICT on urban form, processes and modes of living. Therefore, all can be combined under an umbrella term of *smart city* which is the topic of concern in the chapter.

In 1955, a new concept of artificial intelligence (AI), coined by Dr. John McCarthy, emerged enabling to explore areas which were never thought of before. Its concepts consisting of machine learning, deep learning, computer vision, Internet of Things (IoT) etc. merged with the ICT technologies will bring revolutionary changes in the cities, aiming to become a smart city.

1.4 INTRODUCTION TO MACHINE LEARNING

The focus now shifts to a very critical concept and its need for building a smart city. However, first, to understand this, it is important to talk about "artificial intelligence".

Every species survives based on the type of intelligence they possess. *Homo sapiens* is termed the most "intelligent" species of all because their intelligence consists of many additional and more advanced qualities as compared to others.

These qualities are as follows:

1. *Estimation or prediction*: It is the quality of forecasting a result based on the factors affecting them.
2. *Differentiation or classification*: It is the quality of categorizing different items into definite classes, for example, cats and dogs.
3. *Segmentation*: It is the quality of grouping similar things together, for example, plants and animals.
4. *Recommendation*: It is the quality of recommending something to someone according to their interests.
5. *Decision-making*: It is the quality of choosing based on the information gathered.

Now, when a machine is made or taught to mimic the human tendency to fulfil the previous five qualities of intelligence, it is termed AI.

Shifting focus to the whole basis of the chapter, we can see that machine learning is a subset of AI. In simple words, it consists of algorithms or models which help in achieving AI. It is an intersection of the field of mathematics, computer science and optimization.

Two sub-parts of machine learning will be discussed further.

1.5 INTRODUCTION TO MACHINE LEARNING ALGORITHMS

As discussed earlier, machine learning helps achieve the various characteristics of AI. As the words suggest, it is all about the machines "learning". But what is meant by this? How is a machine supposed to learn? The answer to this is simple, through machine learning algorithms.

Machine learning algorithms are certain mathematical programs that are programmed in a way to provide basic guidelines to create a model. They are accustomed to adjust to the data provided by the user.

Now, before going further to understand the types of machine learning algorithms, it is important to first discuss some important terms which will be used throughout the chapter.

1. *Dataset*: It is a set of data, i.e., collection of data which can be in the form of tables, SQL queries, images and videos.
2. *Training dataset*: This dataset is used to train the machine learning algorithm to create a model.
3. *Validation dataset*: It is a part of the training dataset which is set aside to get an estimate of the performance of the model and tune the hyper-parameters accordingly.
4. *Testing dataset*: It is also a part of the training dataset, and its use is quite similar to the validation data except here the hyper-parameter tuning doesn't take place and gives the results of the final fine-tuned model.

1.5.1 Supervised Machine Learning Algorithms

Supervised learning is derived from the word supervisor who is a person "supervising" a particular task and providing guidance.

For example, when the students are first taught addition, they are provided with a method and some examples to understand the concept. When using this, they understand the task at hand; they are provided with questions similar to the examples to apply what they had learnt.

Similarly, in supervised learning, the machine is taught using some examples called a labelled *training dataset*, through a method or an algorithm. After this, the *validation dataset* is used to fine-tune the hyper-parameters to create an optimum model. When the machine is done learning, unknown data called the *test dataset* is introduced for the machine to provide an unbiased evaluation of a final model fit on the training dataset.

The supervised machine learning algorithms are categorized into two types:

1. *Regression*: This type of analysis is used to map the dependent variables to independent variables to predict the output in the form of continuous values such as a floating point or an integer.
2. *Classification*: In this type of analysis, the prediction output is in the form of discrete values such as classifying if a mail is a spam or ham.

Now, it is the time to discuss some algorithms which can be used to fulfil the tasks of regression, classification or both.

1.5.1.1 Random Forest

To first understand the random forest algorithm, it is imperative to know what are *Decision Trees* and define *Ensemble methods*.

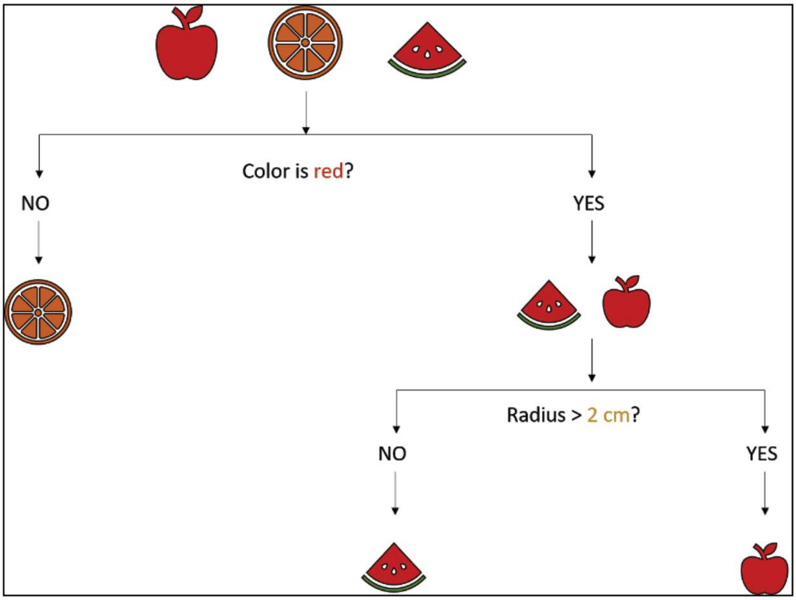

FIGURE 1.1 Decision tree example.

To get a clear idea, an example of a classification problem is provided (Figure 1.1). There is a basket of fruits having apples, oranges and watermelon slices, which need to be segregated into three different baskets with the same type of fruits. This can be done with the help of a decision tree as given in the following.

To get results from a decision tree, a feature split is done to move closer to the classification. So, to split our first node, the question "is the colour of the fruit red?" can be used. As one can observe, watermelon slices and apples are classified as red and orange as not red. Therefore, for oranges, it's clear how to classify them. The next node split can be done by answering the question "radius is greater than 2 cm". These watermelon slices cannot have a radius greater than 2 cm, therefore, are categorized accordingly.

Now, through the example mentioned previously (Figure 1.1), the contents of the fruit basket have been classified into oranges, apples and watermelon slices using two node splits. This was a really simple example to understand decision trees, but in reality, there can be a greater number of nodes used for prediction.

Next, the topic of conversation is *Ensemble Learning*. Sometimes, one type of algorithm cannot generate the desired level of accuracy and to solve this, predictions from models created by multiple machine learning algorithms are combined to improve the predictive performance.

Here, in a random forest, a particular type of ensemble learning is used, known as bagging. Using this method, many decision trees are used to make predictions on different samples of the same dataset and an average of these predictions is considered output for *Random Forest*. The main advantage of this widely used algorithm is reflected when there is a low correlation among the different decision tree models.

The main reason for this is, the lower the correlation, the higher the chances of individual errors of each decision tree not affecting the other tree.

Also, random forest can handle missing data well, robust to outliers, less impacted by noise and is extremely stable, which makes it a very useful algorithm to apply in a smart city.

1.5.1.2 Support Vector Machines (SVMs)

Now, to understand this algorithm, a classification problem statement is used. Along with which some relevant terms will also be defined.

There are two class points represented by an addition operator and a triangle as can be seen in Figure 1.2. In support vector machine (SVM) algorithm, one can classify these points using a hyperplane. Hyperplanes are decision boundaries used to classify data points. As can be observed in the diagram, the hyperplane classifies the two points into different classes. But the story doesn't end with just a simple hyperplane. To make the data points easily linearly separable, the algorithm also creates two margin lines parallel to the hyperplane having some distance from each other. These hyperplanes (on the margin lines) are not only parallel to the original hyperplane but also pass through one or more data points in their class. The main significance of creating these hyperplanes apart from the original hyperplane is to create a more generalized model giving comparatively accurate results. The whole distance between the two hyperplanes created is called the *Margin Distance* and the data points through which these planes pass are called *Support Vectors,* hence, the name, SVMs.

The example discussed earlier gives a basic intuition of how the algorithm works. In reality, one has to work with n-dimensional data which cannot even be visualized and can be non-linearly separable. Another example of non-linearly separable data to understand this better is shown in Figure 1.3.

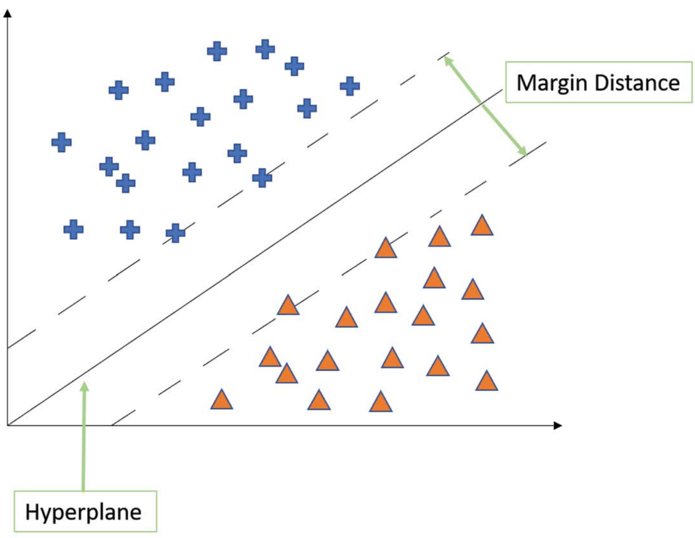

FIGURE 1.2 SVM example.

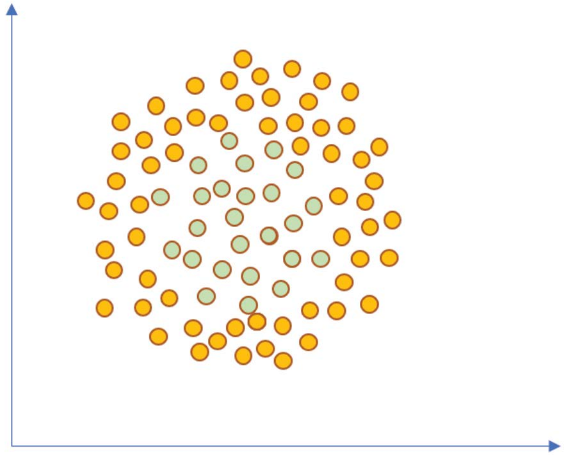

FIGURE 1.3 Non-linearly separable data.

If one treats the non-linear separable data given in Figure 1.3 as linearly separable and creates a hyperplane or line (as its 2D form), it will only give an accuracy of ≤50%. So, how to solve this problem?

To solve this, the SVM algorithm uses the SVM kernels. The SVM kernels convert low-dimensional data into a higher dimension which in this case is from 2D to 3D as shown in Figure 1.4.

As can be observed (Figure 1.4), a hyperplane can be easily created to classify the two types of data points with much higher accuracy than by just creating a line in its 2D form.

In smart cities, the data collected has multiple attributes which have to be processed to proceed to an accurate solution; therefore, SVM is a good choice in such cases as it performs profoundly well with higher-dimensional data.

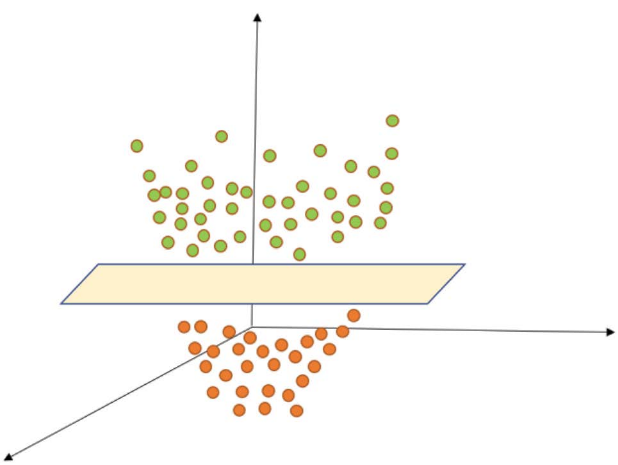

FIGURE 1.4 3D for non-linear separable data.

1.5.1.3 Naïve Bayes

To understand the Naïve Bayes algorithm, it is crucial to understand some statistical concepts called *Conditional Probability* and *Bayes' Theorem*.

Conditional probability is the one of an event A occurring such that event B has already occurred. The formula for the same can be observed as given in the following:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Bayes' theorem gives the probability of an event based on the probabilities of the data provided in the hypothesis. It assumes that every input variable is dependent on all the variables. The formula for the same is as follows:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

The concept can be applied in an algorithm in the following manner. Let there be a set of inputs $X = x_1, x_2, x_3, \ldots, x_n$ and y be the output to be classified. The probability of x is known and the probability of y is to be calculated. This can be represented as follows:

$$P(y|x_1 \cap x_2 \cap x_3 \cdots \cap x_n) = \frac{P(x_1|y) * P(x_2|y) * P(x_3|y) \cdots * P(x_n|y) * P(y)}{P(x_1) * P(x_2) * P(x_3) * \cdots * P(x_n)}$$

Just as observed in Bayes' theorem, the event which has occurred has been replaced with the input features x.

Now, assume that the event has to be classified as True or False and by the algorithm formula applied earlier, the resulting probability is 0.8 for True and 0.2 for False. To get the results, one more formula has to be applied, which will give the final classification result:

$$y = argmax\left[P(y)\prod_{i=1}^{n} \Pi(P(x_i) \mid y)\right]$$

where the argmax operation returns the result with the maximum probability, which in this case is 0.8 to classify the set of features as True. Hence, this is the intuition behind the working of the Naïve Bayes algorithm.

The Naïve Bayes algorithm is quite useful in a smart city setup as it is very simple to implement. Also, it requires a less amount of training data to train, thus making it a computationally fast algorithm.

1.5.1.4 K-Nearest Neighbours

This algorithm is explained using two use cases, classification and regression.

First focus of attention is classification. As can be observed from Figure 1.5, the data points belong to two classes, blue and green. A new data point indicated by the red star has been introduced, and it has to be classified in either blue or green class. But the question here is, how to solve using the K-nearest neighbours (KNN) algorithm?