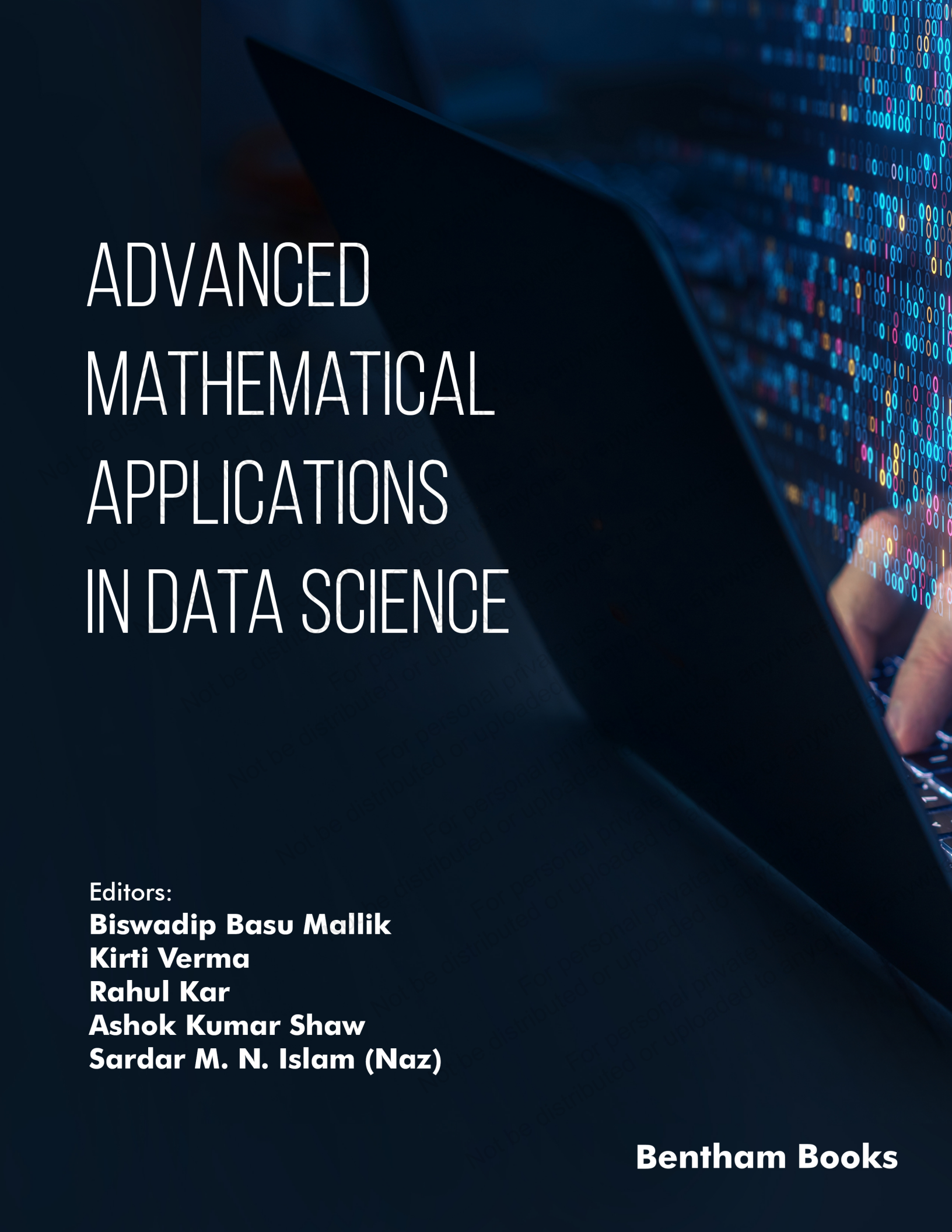

ADVANCED MATHEMATICAL APPLICATIONS IN DATA SCIENCE

Editors:
Biswadip Basu Mallik
Kirti Verma
Rahul Kar
Ashok Kumar Shaw
Sardar M. N. Islam (Naz)

Bentham Books

Advanced Mathematical Applications in Data Science

Editors: Biswadip Basu Mallik, Kirti Verma, Rahul Kar, Ashok Kumar Shaw & Sardar M. N. Islam (Naz)

ISBN (Online): 978-981-5124-84-2

ISBN (Print):978-981-5124-85-9

ISBN (Paperback): 978-981-5124-86-6

© 2023, Bentham Books imprint.

Published by Bentham Science Publishers Pte. Ltd. Singapore. All Rights Reserved.

First published in 2023.

For personal private use only
Not be distributed or uploaded to anyone or anywhere

APPLICATION OF STATISTICS IN DATA SCIENCE ANALYZING AND CATEGORIZING DATA 58
NUMERIC DATA & CATEGORICAL DATA 58
EXPLORATORY KNOWLEDGE ANALYSIS 58
SIGNIFICANCE TESTS 58
Null Hypotheses 59
Alternative Hypotheses 59
CHI-SQUARED CHECK 59
STUDENT'S T-TEST 59
ANALYSIS OF VARIANCE CHECK (ANOVA) 59
Unidirectional 60
Two-ways 60
RESERVATION AND PREDICTION 60
Linear Regression 60
Logistic Regression 60
CLASSIFICATION OF KNOWLEDGE SCIENCE IN STATISTICS 60
Naive Mathematician 61
K-nearest Neighbors 61
PROBABILITY 61
FREQUENCY TABLES 61
HISTOGRAM 62
CONTINUOUS RANDOM VARIABLES 63
SKEWNESS DISTRIBUTION 63
RIGHT SKEW DISTRIBUTION 63
LEFT SKEW DISTRIBUTION 63
NORMAL DISTRIBUTION 63
EXPONENTIAL DISTRIBUTION 64
UNIFORM DISTRIBUTION 64
POISSON DISTRIBUTION 64
IMPORTANT OF INFORMATION SCIENCE 64
DATA WHILE NOT KNOWLEDGE SCIENCE 65
DATA CAN PRODUCE HIGHER CLIENT EXPERTISE 65
DATA USED ACROSS VERTICALS 65
POWER OF INFORMATION SCIENCE 65
FUTURE OF INFORMATION SCIENCE 65
DATA SCIENCE IN TRADE 66
BENEFITS OF KNOWLEDGE SCIENCE 66
STATISTICAL INFORMATION 66
DATA SCIENCE IS VERY IMPORTANT IN THE MODERN WORLD 66
DATA INDIVIDUAL 66
DATA SCIENCE WORKS 67
CONCLUDING REMARKS 67
REFERENCES 67

CHAPTER 4 BAG OF VISUAL WORDS MODEL - A MATHEMATICAL APPROACH 68
Maheswari
INTRODUCTION 68
HISTOGRAM REWEIGHTING – TF – IDF APPROACH 69
COST MATRIX GENERATION 69
EUCLIDEAN DISTANCE AND COSINE DISTANCE 70
MODEL DESCRIPTION 71

Not be distributed or uploaded to anyone or anywhere
For personal private use only

CHAPTER 4

Bag of Visual Words Model - A Mathematical Approach

Maheswari[1,*]

[1] *Department of Computer Applications, Fatima College, Madurai, India*

Abstract: Information extraction from images is now incredibly valuable for many new inventions. Even though there are several simple methods for extracting information from the images, feasibility and accuracy are critical. One of the simplest and most significant processes is feature extraction from the images. Many scientific approaches are derived by the experts based on the extracted features for a better conclusion of their work. Mathematical procedures, like Scientific methods, play an important role in image analysis. The Bag of Visual Words (BoVW) [1, 2, 3] is one of them, and it is helpful to figure out how similar a group of images is. A set of visual words characterises the images in the Bag of Visual Words model, which are subsequently aggregated in a histogram per image [4]. The histogram difference depicts the similarities among the images. The reweighting methodology known as Term Frequency – Inverse Document Frequency (TF-IDF) [5] refines this procedure. The overall weighting [6] for all words in each histogram is calculated before reweighting. As per the traditional way, the images are transformed into the matrix called as Cost matrix. It is constructed through two mathematical: Euclidean distances and Cosine distances. The main purpose of finding these distances is to detect similarity between the histograms. Further the histograms are normalized and both distances are calculated. The visual representation is also generated. The two mathematical methods are compared to see which one is appropriate for checking resemblance. The strategy identified as the optimum solution based on the findings aids in fraud detection in digital signature, Image Processing, and classification of images.

Keywords: Bag of visual words, Cost matrix, Cosine distance, Euclidean distance.

INTRODUCTION

Recent advances in information retrieval from images, in particular with mathematical methods, are helpful for research [7, 8]. The dissimilarities among a

[*] **Corresponding author Maheswari:** Department of Computer Applications, Fatima College, Madurai, India; E-mail: kpmshri123@gmail.com

Biswadip Basu Mallik, Kirti Verma, Rahul Kar, Ashok Kumar Shaw & Sardar M. N. Islam (Naz) (Eds.)
All rights reserved-© 2023 Bentham Science Publishers