



CRC Press  
Taylor & Francis Group

A CHAPMAN & HALL BOOK

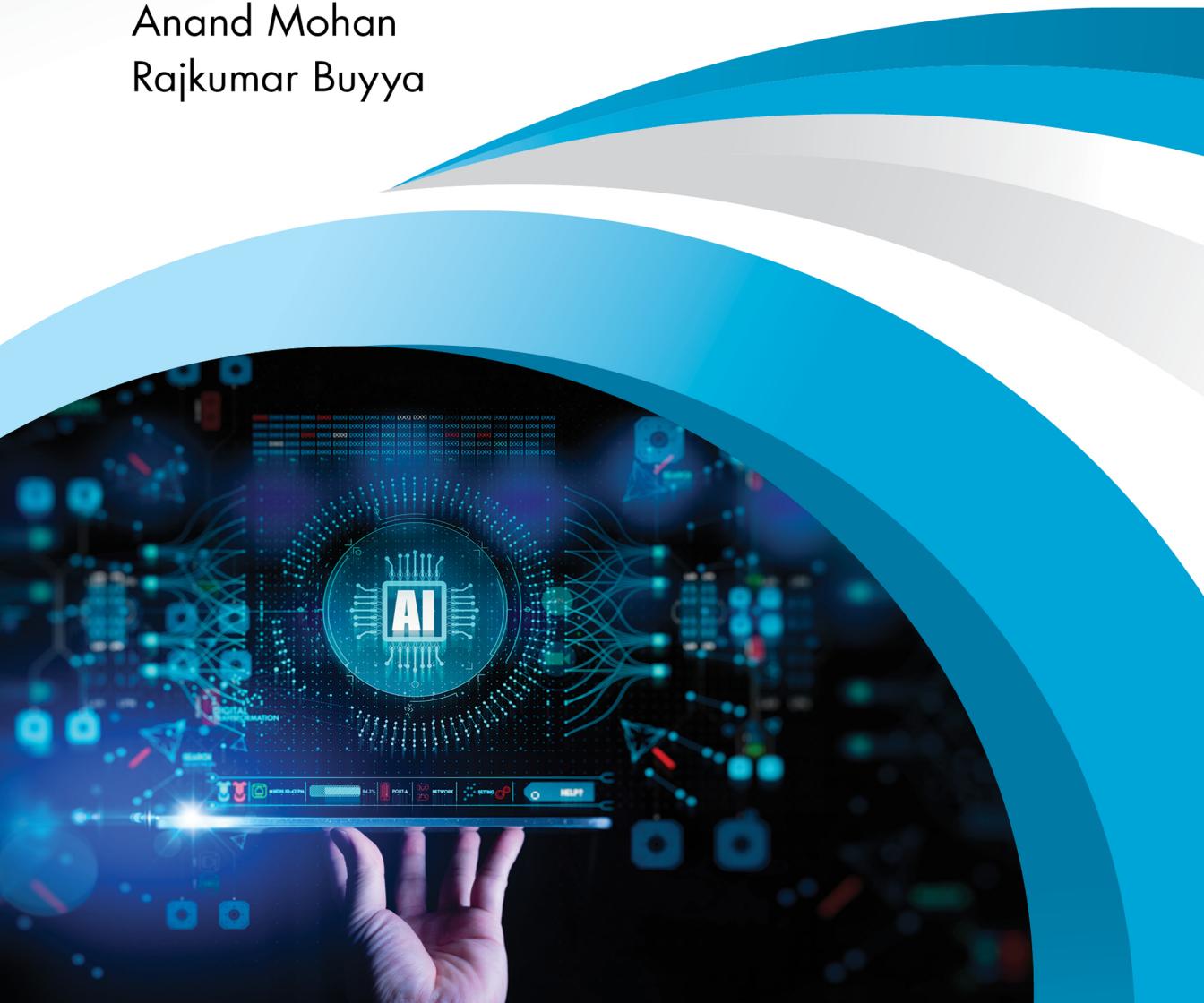
# Machine Learning for Cloud Management

Jitendra Kumar

Ashutosh Kumar Singh

Anand Mohan

Rajkumar Buyya



# Machine Learning for Cloud Management



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

# Machine Learning for Cloud Management

Jitendra Kumar  
Ashutosh Kumar Singh  
Anand Mohan  
Rajkumar Buyya



CRC Press

Taylor & Francis Group

Boca Raton London New York

---

CRC Press is an imprint of the  
Taylor & Francis Group, an **informa** business  
A CHAPMAN & HALL BOOK

First edition published 2022  
by CRC Press  
6000 Broken Sound Parkway NW, Suite 300, Boca Raton, FL 33487-2742

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

© 2022 Jitendra Kumar, Ashutosh Kumar Singh, Anand Mohan, Rajkumar Buyya

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

Reasonable efforts have been made to publish reliable data and information, but the author and publisher cannot assume responsibility for the validity of all materials or the consequences of their use. The authors and publishers have attempted to trace the copyright holders of all material reproduced in this publication and apologize to copyright holders if permission to publish in this form has not been obtained. If any copyright material has not been acknowledged please write and let us know so we may rectify in any future reprint.

Except as permitted under U.S. Copyright Law, no part of this book may be reprinted, reproduced, transmitted, or utilized in any form by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying, microfilming, and recording, or in any information storage or retrieval system, without written permission from the publishers.

For permission to photocopy or use material electronically from this work, access [www.copyright.com](http://www.copyright.com) or contact the Copyright Clearance Center, Inc. (CCC), 222 Rosewood Drive, Danvers, MA 01923, 978-750-8400. For works that are not available on CCC please contact [mpkbookspermissions@tandf.co.uk](mailto:mpkbookspermissions@tandf.co.uk)

*Trademark notice:* Product or corporate names may be trademarks or registered trademarks and are used only for identification and explanation without intent to infringe.

---

**Library of Congress Cataloging-in-Publication Data**

---

Names: Kumar, Jitendra, 1975- author. | Singh, Ashutosh Kumar, author. |  
Mohan, Anand (Of Indian Institute of Technology), author. | Buyya,  
Rajkumar, 1970- author.  
Title: Machine learning for cloud management / Jitendra Kumar, Ashutosh  
Kumar Singh, Anand Mohan, Rajkumar Buyya.  
Description: First edition. | Boca Raton : CRC Press, 2022. | Includes  
bibliographical references and index.  
Identifiers: LCCN 2021027713 | ISBN 9780367626488 (hardback) | ISBN  
9780367622565 (paperback) | ISBN 9781003110101 (ebook)  
Subjects: LCSH: Cloud computing. | Machine learning.  
Classification: LCC QA76.585 .K85 2022 | DDC 004.67/82--dc23  
LC record available at <https://lccn.loc.gov/2021027713>

---

ISBN: 978-0-367-62648-8 (hbk)  
ISBN: 978-0-367-62256-5 (pbk)  
ISBN: 978-1-003-11010-1 (ebk)

DOI: [10.1201/9781003110101](https://doi.org/10.1201/9781003110101)

*Publisher's note:* This book has been prepared from camera-ready copy provided by the authors.

*Dedicated to,*

*My wife: Gita, daughter: Aru, and Parents*

*~Jitendra Kumar*

*Anushka, Aakash, Akankshya, and Parents*

*~Ashutosh Kumar Singh*

*My wife: Sudha Mohan, son: Ashish Mohan, daughter: Amrita Mohan, and Late parents*

*~Anand Mohan*

*My international collaborators and team members in Melbourne CLOUDS Lab!*

*~Rajkumar Buyya*



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

---

# Contents

---

List of Figures	xi
List of Tables	xvii
Preface	xix
Author	xxi
Abbreviations	xxv
<b>CHAPTER 1 ■ Introduction</b>	<b>1</b>
1.1 CLOUD COMPUTING	1
1.2 CLOUD MANAGEMENT	2
1.2.1 Workload Forecasting	3
1.2.2 Load Balancing	4
1.3 MACHINE LEARNING	5
1.3.1 Artificial Neural Network	5
1.3.2 Metaheuristic Optimization Algorithms	6
1.3.3 Time Series Analysis	7
1.4 WORKLOAD TRACES	7
1.5 EXPERIMENTAL SETUP & EVALUATION METRICS	8
1.6 STATISTICAL TESTS	9
1.6.1 Wilcoxon Signed-Rank Test	10
1.6.2 Friedman Test	10
1.6.3 Finner Test	10
<b>CHAPTER 2 ■ Time Series Models</b>	<b>13</b>
2.1 AUTOREGRESSION	14
2.2 MOVING AVERAGE	14
2.3 AUTOREGRESSIVE MOVING AVERAGE	15

2.4	AUTOREGRESSIVE INTEGRATED MOVING AVERAGE	15
2.5	EXPONENTIAL SMOOTHING	17
2.6	EXPERIMENTAL ANALYSIS	17
2.6.1	Forecast Evaluation	17
2.6.2	Statistical Analysis	21
<b>CHAPTER</b>	<b>3 ■ Error Preventive Time Series Models</b>	<b>25</b>
3.1	ERROR PREVENTION SCHEME	25
3.2	PREDICTIONS IN ERROR RANGE	27
3.3	MAGNITUDE OF PREDICTIONS	28
3.4	ERROR PREVENTIVE TIME SERIES MODELS	29
3.4.1	Error Preventive Autoregressive Moving Average	29
3.4.2	Error Preventive Autoregressive Integrated Moving Average	35
3.4.3	Error Preventive Exponential Smoothing	38
3.5	PERFORMANCE EVALUATION	47
3.5.1	Comparative Analysis	47
3.5.2	Statistical Analysis	53
<b>CHAPTER</b>	<b>4 ■ Metaheuristic Optimization Algorithms</b>	<b>59</b>
4.1	SWARM INTELLIGENCE ALGORITHMS IN PREDICTIVE MODEL	59
4.1.1	Particle Swarm Optimization	60
4.1.2	Firefly Search Algorithm	61
4.2	EVOLUTIONARY ALGORITHMS IN PREDICTIVE MODEL	62
4.2.1	Genetic Algorithm	63
4.2.2	Differential Evolution	63
4.3	NATURE INSPIRED ALGORITHMS IN PREDICTIVE MODEL	64
4.3.1	Harmony Search	64
4.3.2	Teaching Learning Based Optimization	66
4.4	PHYSICS INSPIRED ALGORITHMS IN PREDICTIVE MODEL	67
4.4.1	Gravitational Search Algorithm	68
4.4.2	Blackhole Algorithm	69
4.5	STATISTICAL PERFORMANCE ASSESSMENT	71

---

<b>CHAPTER</b>	<b>5 ■ Evolutionary Neural Networks</b>	<b>75</b>
5.1	NEURAL NETWORK PREDICTION FRAMEWORK DESIGN	75
5.2	NETWORK LEARNING	77
5.3	RECOMBINATION OPERATOR STRATEGY LEARNING	78
5.3.1	Mutation Operator	78
5.3.1.1	DE/current to best/1	78
5.3.1.2	DE/best/1	78
5.3.1.3	DE/rand/1	78
5.3.2	Crossover Operator	79
5.3.2.1	Ring Crossover	79
5.3.2.2	Heuristic Crossover	79
5.3.2.3	Uniform Crossover	80
5.3.3	Operator Learning Process	80
5.4	ALGORITHMS AND ANALYSIS	83
5.5	FORECAST ASSESSMENT	86
5.5.1	Short Term Forecast	87
5.5.2	Long Term Forecast	88
5.6	COMPARATIVE ANALYSIS	92
<b>CHAPTER</b>	<b>6 ■ Self Directed Learning</b>	<b>97</b>
6.1	NON-DIRECTED LEARNING-BASED FRAMEWORK	97
6.1.1	Non-Directed Learning	98
6.2	SELF-DIRECTED LEARNING-BASED FRAMEWORK	99
6.2.1	Self-Directed Learning	100
6.2.2	Cluster-Based Learning	100
6.2.3	Complexity analysis	102
6.3	FORECAST ASSESSMENT	103
6.3.1	Short Term Forecast	103
6.3.1.1	Web Server Workloads	104
6.3.1.2	Cloud Workloads	104
6.4	LONG TERM FORECAST	106
6.4.0.1	Web Server Workloads	106
6.4.0.2	Cloud Workloads	107
6.5	COMPARATIVE & STATISTICAL ANALYSIS	108

---

CHAPTER	7 ■ Ensemble Learning	121
7.1	EXTREME LEARNING MACHINE	121
7.2	WORKLOAD DECOMPOSITION PREDICTIVE FRAMEWORK	122
7.2.1	Framework Design	122
7.3	ELM ENSEMBLE PREDICTIVE FRAMEWORK	125
7.3.1	Ensemble Learning	126
7.3.2	Expert Architecture Learning	127
7.3.3	Expert Weight Allocation	129
7.4	SHORT TERM FORECAST EVALUATION	130
7.5	LONG TERM FORECAST EVALUATION	133
7.6	COMPARATIVE ANALYSIS	137
CHAPTER	8 ■ Load Balancing	141
8.1	MULTI-OBJECTIVE OPTIMIZATION	141
8.2	RESOURCE-EFFICIENT LOAD BALANCING FRAMEWORK	142
8.3	SECURE AND ENERGY-AWARE LOAD BALANCING FRAMEWORK	146
8.3.1	Side-Channel Attacks	147
8.3.2	Ternary Objective VM Placement	148
8.4	SIMULATION SETUP	151
8.5	HOMOGENEOUS VM PLACEMENT ANALYSIS	151
8.6	HETEROGENEOUS VM PLACEMENT ANALYSIS	152
CHAPTER	9 ■ Summary	155
	Bibliography	159
	Index	171

---

# List of Figures

---

1.1	Service model view of cloud computing	2
1.2	Cloud resource management view	2
1.3	Cloud resource management approaches	3
1.4	Schematic representation of workload forecasting	4
1.5	Load balancing	5
1.6	Artificial neural network	6
1.7	An arbitrary optimization function with multiple local optima	7
2.1	Autoregression process	14
2.2	Moving average process	15
2.3	Autoregressive moving average process	15
2.4	Autoregressive integrated moving average process	16
2.5	Exponential smoothing process	17
2.6	Autoregression forecast results on MAE	18
2.7	Autoregression forecast results on MASE	18
2.8	Moving average forecast results on MAE	19
2.9	Moving average forecast results on MASE	19
2.10	Autoregressive moving average forecast results on MAE	20
2.11	Autoregressive moving average forecast results on MASE	20
2.12	Autoregressive integrated moving average forecast results on MAE	21
2.13	Autoregressive integrated moving average forecast results on MASE	21
2.14	Exponential smoothing forecast results on MAE	22
2.15	Exponential smoothing forecast results on MASE	22
3.1	Error preventive workload forecasting model	26
3.2	An example of error preventive and non-error preventive forecast	27
3.3	Prediction error range and segments	28
3.4	Error preventive ARMA forecast analysis for 10-minute prediction interval	30

3.5	Error preventive ARMA forecast analysis for 20-minute prediction interval	31
3.6	Error preventive ARMA forecast analysis for 30-minute prediction interval	32
3.7	Error preventive ARMA forecast analysis for 60-minute prediction interval	33
3.8	R-score comparison of non-error preventive and error preventive ARMA	34
3.9	SEI comparison of non-error preventive and error preventive ARMA	35
3.10	MPE comparison of non-error preventive and error preventive ARMA	35
3.11	PER comparison of non-error preventive and error preventive ARMA	36
3.12	Positive magnitude comparison of non-error preventive and error preventive ARMA	37
3.13	Negative magnitude comparison of non-error preventive and error preventive ARMA	37
3.14	Error preventive ARIMA forecast analysis for 10-minute prediction interval	39
3.15	Error preventive ARIMA forecast analysis for 20-minute prediction interval	40
3.16	Error preventive ARIMA forecast analysis for 30-minute prediction interval	41
3.17	Error preventive ARIMA forecast analysis for 60-minute prediction interval	42
3.18	R-score comparison of non-error preventive and error preventive ARIMA	43
3.19	SEI comparison of non-error preventive and error preventive ARIMA	43
3.20	MPE comparison of non-error preventive and error preventive ARIMA	44
3.21	Positive magnitude comparison of non-error preventive and error preventive ARIMA	45
3.22	Negative magnitude comparison of non-error preventive and error preventive ARIMA	45
3.23	PER comparison of non-error preventive and error preventive ARIMA	46
3.24	Error preventive ES forecast analysis for 10-minute prediction interval	47
3.25	Error preventive ES forecast analysis for 20-minute prediction interval	48
3.26	Error preventive ES forecast analysis for 30-minute prediction interval	49
3.27	Error preventive ES forecast analysis for 60-minute prediction interval	50
3.28	R-score comparison of non-error preventive and error preventive ES	51
3.29	SEI comparison of non-error preventive and error preventive ES	51
3.30	MPE comparison of non-error preventive and error preventive ES	52
3.31	Positive magnitude comparison of non-error preventive and error preventive ES	52

3.32	Negative magnitude comparison of non-error preventive and error preventive ES	53
3.33	PER comparison of non-error preventive and error preventive ES	54
3.34	R-score performance relative improvement	55
3.35	SEI performance relative improvement	55
3.36	MPE performance relative improvement	56
3.37	Friedman test ranks of non-error preventive and error preventive models	56
4.1	Population-based metaheuristic optimization algorithms' taxonomy	59
4.2	Forecast accuracy comparison of swarm intelligence based prediction models	62
4.3	Wilcoxon test statistics of swarm intelligence based prediction models	62
4.4	Forecast accuracy comparison of evolutionary algorithms based prediction models	65
4.5	Wilcoxon test statistics of evolutionary algorithms based prediction models	65
4.6	Forecast accuracy comparison of nature-inspired algorithms based prediction models	67
4.7	Wilcoxon test statistics of nature-inspired algorithms based prediction models	68
4.8	Forecast accuracy comparison of physics-inspired algorithms based prediction models	70
4.9	Wilcoxon test statistics of physics-inspired algorithms based prediction models	71
4.10	Friedman test ranks of metaheuristic algorithms based prediction models	71
5.1	Neural network-based workload prediction model	76
5.2	Ring crossover	79
5.3	Heuristic crossover	80
5.4	Uniform crossover	81
5.5	Learning period effect on forecast accuracy of self-adaptive differential evolution algorithm based workload prediction model	86
5.6	Short term forecast assessment of self-adaptive differential evolution algorithm based prediction model on 1-minute prediction interval	87
5.7	Short term forecast residuals of self-adaptive differential evolution algorithm based prediction model on 1-minute prediction interval	88
5.8	Short term forecast assessment of biphasic adaptive differential evolution algorithm based prediction model on 1-minute prediction interval	89

5.9	Short term forecast residual of biphasic adaptive differential evolution algorithm based prediction model on 1-minute prediction interval	89
5.10	Comparing short term forecast errors of SaDE and BaDE based predictive frameworks	90
5.11	Long term forecast assessment of self-adaptive differential evolution algorithm based prediction model on 60-minute prediction interval	91
5.12	Long term forecast residuals of self-adaptive differential evolution algorithm based prediction model on 60-minute prediction interval	91
5.13	Long term forecast assessment of biphasic adaptive differential evolution algorithm based prediction model on 60-minute prediction interval	92
5.14	Long term forecast residuals of biphasic adaptive differential evolution algorithm based prediction model on 60-minute prediction interval	92
5.15	Comparing long term forecast errors of SaDE and BaDE based predictive frameworks	93
5.16	Comparing forecast accuracy of Maximum, Average, BPNN, SaDE, BaDE based predictive frameworks on NASA Trace	93
5.17	Comparing forecast accuracy of Maximum, Average, BPNN, SaDE, BaDE based predictive frameworks on Saskatchewan Trace	94
6.1	Non-directed predictive framework	98
6.2	Self-directed predictive framework	99
6.3	Self-directed learning process	100
6.4	Position update procedures in blackhole algorithm (standard vs. modified)	101
6.5	Web server workload prediction results of self-directed learning predictive framework on 5-minute prediction interval	105
6.6	Web server workload prediction residuals auto-correlation of self-directed learning predictive framework on 5-minute prediction interval	106
6.7	Mean squared error of non-directed and self-directed predictive frameworks on short term forecasts of web server workloads	107
6.8	Training time (min) of non-directed and self-directed predictive frameworks on short term forecasts of web server workloads	108
6.9	Cloud server workload prediction results of self-directed learning predictive framework on 5-minute prediction interval	109
6.10	Cloud server workload prediction residuals auto-correlation of self-directed learning predictive framework on 5-minute prediction interval	110
6.11	Web server workload prediction results of self-directed learning predictive framework on 60-minute prediction interval	111
6.12	Web server workload prediction residuals auto-correlation of self-directed learning predictive framework on 60-minute prediction interval	112

6.13	Mean squared error of non-directed and self-directed predictive frameworks on long term forecasts of web server workloads	113
6.14	Training time (min) of non-directed and self-directed predictive frameworks on long term forecasts of web server workloads	114
6.15	Cloud server workload prediction results of self-directed learning predictive framework on 60-minute prediction interval	115
6.16	Cloud server workload prediction residuals auto-correlation of self-directed learning predictive framework on 60-minute prediction interval	116
7.1	Decomposition based predictive framework	122
7.2	Decomposition of CPU requests trace	123
7.3	A conceptual view of ensemble stacking	125
7.4	An ensemble of ELMs in workload prediction	127
7.5	Input node selection	128
7.6	Network architecture analysis for short term forecast of decomposition predictive framework	131
7.7	CPU and Memory data-trace auto-correlation for 5-minute prediction interval	132
7.8	Short term forecast accuracy of ensemble prediction framework on CPU Trace	133
7.9	Short term forecast accuracy of ensemble prediction framework on Memory Trace	134
7.10	Network architecture analysis for long term forecast of decomposition predictive framework	135
7.11	Long term forecast accuracy of ensemble prediction framework on CPU Trace	136
7.12	Long term forecast accuracy of ensemble prediction framework on Memory Trace	136
8.1	An illustration of virtual machine placement scenarios	142
8.2	Resource-efficient load balancing framework design	143
8.3	Chromosome encoding for VM placement	144
8.4	Single point crossover operator for VM placement	145
8.5	Swapping based mutation operator for VM placement	145
8.6	Three cases of VM allocation	149
8.7	Secure and energy-aware load balancing framework design	150
8.8	Power consumption (W) for homogeneous VM requests	152
8.9	Resource utilization for homogeneous VM requests	152
8.10	Presence of conflicting servers (%) for homogeneous VM requests	153

**xvi ■ LIST OF FIGURES**

8.11	Power consumption (W) for heterogeneous VM requests	153
8.12	Resource utilization for heterogeneous VM requests	154
8.13	Presence of conflicting servers (%) for heterogeneous VM requests	154

---

# List of Tables

---

2.1	Friedman test statistics for time series forecasting models	23
2.2	Friedman test ranks for time series forecasting models	23
2.3	Finner test post-hoc analysis of time series forecasting models	23
3.1	Wilcoxon test statistics for error preventive and non-error preventive time series forecasting model	57
3.2	Finner test post-hoc analysis of error preventive and non-error preventive time series forecasting models	57
4.1	Friedman test statistics of metaheuristic algorithms based prediction models	71
4.2	Finner test post-hoc analysis statistics of metaheuristic algorithms based prediction models ( $\alpha = 0.05$ )	73
4.3	Finner test post-hoc analysis results on the null hypothesis of metaheuristic algorithms based prediction model	74
5.1	Number of iterations and time elapsed in the training of differential evolution based predictive models for short term forecasts	90
5.2	Number of iterations and time elapsed in the training of differential evolution based predictive models for long term forecasts	90
5.3	Iterations elapsed for the training of predictive models using backpropagation, self-adaptive differential evolution, and biphasic adaptive differential evolution	94
5.4	Training time (sec) elapsed in the training of predictive models using backpropagation, self-adaptive differential evolution, and biphasic adaptive differential evolution	95
6.1	Mean squared error of non-directed and self-directed predictive frameworks on short term forecasts of cloud server workloads	110
6.2	Training time (min) of non-directed and self-directed predictive frameworks on short term forecasts of cloud server workloads	111
6.3	Mean squared error of non directed and self-directed predictive frameworks on long term forecasts of cloud server workloads	117

6.4	Training time (min) of non-directed and self-directed predictive frameworks on long term forecasts of cloud server workloads	117
6.5	Mean squared error comparison of non-directed and self-directed learning based models' NASA Trace forecasts with state-of-art models	117
6.6	Mean squared error comparison of non-directed and self-directed learning based models' Calgary Trace forecasts with state-of-art models	117
6.7	Mean squared error comparison of non-directed and self-directed learning based models' Saskatchewan Trace forecasts with state-of-art models	118
6.8	Mean squared error comparison of non-directed and self-directed learning based models' CPU Trace forecasts with state-of-art models	118
6.9	Mean squared error comparison of non-directed and self-directed learning based models' Memory Trace forecasts with state-of-art models	118
6.10	Mean squared error comparison of self-directed learning-based model's PlanetLab Trace forecasts with deep learning model	118
6.11	Friedman test ranks of non-directed, self-directed, LSTM, SaDE, and backpropagation based predictive models	119
6.12	Friedman test statistics of non-directed and self-directed learning predictive frameworks	119
6.13	Wilcoxon signed test ranks for self-directed learning predictive framework	119
7.1	ARIMA analysis orders for cloud resource demand traces	124
7.2	Network configuration parameter choices for decomposition predictive framework	124
7.3	List of experiments selected by D-Optimal Design	124
7.4	Mean squared error of short term forecast of ELM based predictive framework	134
7.5	Mean squared error of long term forecast of ELM based predictive framework	137
7.6	Forecast accuracy comparison of ELM based predictive models on CPU trace with state-of-art models	138
7.7	Forecast accuracy comparison of ELM based predictive models on Memory trace with state-of-art models	138
7.8	Forecast accuracy comparison of ELM based predictive models on Google cluster trace and PlanetLab Trace with state-of-art models	139
8.1	Virtual machine details for illustration	148

---

# Preface

---

Cloud computing has become one of the revolutionary technology in the history of the computing world. It offers subscription-based on-demand services and has emerged as the backbone of the computing industry. It has enabled us to share resources among multiple users through virtualization by the means of creating a virtual instance of a computer system running in an abstracted hardware layer. Unlike early distributed computing models, it assures limitless computing resources through its large-scale cloud data centers. It has gained wide popularity over the past few years, with an ever-increasing infrastructure, number of users, and amount of hosted data. The large and complex workloads hosted on these data centers introduce several challenges: resource utilization, power consumption, scalability, operational cost, and many others. Therefore, a practical resource management scheme is essential to bring operational efficiency with improved elasticity. The elasticity of a system depends on several factors such as the accuracy of anticipated workload information, performance behavior of applications in different scenarios communicating the forecast results, use of the anticipated information, and many others.

Effective resource management can be achieved through workload prediction, resource scheduling, and provisioning, virtual machine placement, or a combination of these approaches. The workload prediction has been widely explored and a number of methods are presented. However, the existing methods suffer from various issues including the incapability of capturing the non-linearity of workloads and iterative training that consumes huge computing resources and time. This book discusses the machine learning-based approaches to address the above-mentioned issues. The highlights of the discussed models are continuous learning from error feedback, adaptive nature, decomposition of workload traces, and ensemble learning. Detailed analysis of predictive methods on different workload traces is also included and their performance is compared with state-of-art models. Virtual machine placement is another aspect that is explored to achieve efficient resource management. In general, virtual machine placement is a multiobjective problem that involves multiple conflicting objectives to be optimized simultaneously. The frameworks discussed in this book address the issues of resource utilization, power consumption, and security while placing the workloads on servers.

The remainder of the book is organized as follows: [Chapter 1](#) briefs the basic cloud computing concepts. The discussion on the workload prediction mechanisms begins in [chapter 2](#). First, the basic time series forecasting models are discussed with their performance on different workload traces. [Chapter 3](#) discusses the error preventive time series forecasting models which significantly improve the performance over classical time series models. Then, a discussion on various nature-inspired algorithms is included