



Advances in Computational Collective Intelligence

SOCIAL MEDIA AND CROWDSOURCING

APPLICATION AND ANALYTICS

Edited by

Sujoy Chatterjee, Thipendra P. Singh,
Sunghoon Lim and Anirban Mukhopadhyay

An **Auerbach** Book



Social Media and Crowdsourcing

Social media and crowdsourcing are important tools for solving complex problems. The benefit of crowdsourcing is that it leverages the power of human intelligence cost effectively and with less time. *Social Media and Crowdsourcing: Application and Analytics* examines the concepts of social media and crowdsourcing as well as their analytical aspects. It explores how these technologies contribute to the real world and examines such applications as promoting social good, agriculture, healthcare, tourism, disaster management, education, crime control, and cultural heritage. The book also looks at ethical issues in crowdsourcing and future scenarios and challenges for policy. Highlights of the book include the following:

- A crowdsourcing application in agriculture
- Crowdsourcing outline for a contemporary aided medicinal backup system
- Crowdsourcing-based recommendation in the tourism industry
- Crowdsourcing mechanisms for reviving cultural heritage

Expanding the overarching concept of utilizing social media and crowdsourcing to solve various real-life problems, this book discusses how to bring together the wisdom of crowds for various decision-making problems in agriculture, disaster management, and healthcare. It addresses the various ethical issues arising out of various crowdsourcing-based applications. It puts forward diverse methodologies to involve crowdsourcing in education to implement new strategies to enhance learning outcomes. This book also addresses various problem-solving techniques for recommender applications in the travel and tourism industry.

Providing a systematic discussion of the many sectors using crowdsourcing as an essential part of social innovation, this book is a theoretical and methodological look at the application of social media.

Dr. Sujoy Chatterjee is an assistant professor, Department of Computer Science and Engineering, Amity University, Kolkata, India.

Dr. Thipendra P. Singh is currently a professor at the School of Computer Science Engineering and Technology, Bennett University, Greater Noida, NCR, India.

Dr. Sunghoon Lim is currently an Associate Professor in the Department of Industrial Engineering and the Graduate School of Artificial Intelligence, Ulsan National Institute of Science and Technology (UNIST), Ulsan, Republic of Korea.

Dr. Anirban Mukhopadhyay is currently a professor at the Department of Computer Science and Engineering, University of Kalyani, Kalyani, West Bengal, India.

Advances in Computational Collective Intelligence

Edited by Dr. Subhendu Kumar Pani

Principal, Krupajal Group of Institutions, India

Published

Technologies for Sustainable Global Higher Education

By Maria José Sousa, Andreia de Bem Machado, and Gertrudes Aparecida Dandolini

ISBN: 978-1-032-262895

Social Media and Crowdsourcing

By Sujoy Chatterjee, Thipendra P Singh, Sunghoon Lim, and Anirban Mukhopadhyay

ISBN: 978-1-032-386874

Forthcoming

Artificial Intelligence and Machine Learning for Risk Management of Natural Hazards and Disasters

By Cees van Westen, Romulus Costache, Dimitrios A. Karras, R. S. Ajin, and Sekhar L. Kuriakose

ISBN: 978-1-032-232768

Computational Intelligence in Industry 4.0 and 5.0 Applications: Challenges and Future Prospects

Joseph Bamidele Awotunde, Kamalakanta Muduli, and Biswajit Brahma

ISBN: 978-1-032-539225

Deep Learning for Smart Healthcare: Trends, Challenges and Applications

K. Murugeswari, B. Sundaravadivazhagan, S. Poonkuntran, and Thendral Puyalnithi

ISBN: 978-1-032-455815

Edge Computational Intelligence for AI-Enabled IoT Systems

By Shrikaant Kulkarni, Jaiprakash Narain Dwivedi, Dinda Pramanta, and Yuichiro Tanaka

ISBN: 978-1-032-207667

Explainable AI and Cybersecurity

By Mohammad Tabrez Quasim, Abdullah Alharthi, Ali Alqazzaz, Mohammed Mujib Alshahrani, Ali Falh Alshahrani, and Mohammad Ayoub Khan

ISBN: 978-1-032-422213

Machine Learning in Applied Sciences

By M. A. Jabbar, Shankru Guggari, Kingsley Okoye, and Houneida Sakly

ISBN: 978-1-032-251721

For more information about this series, please visit: <https://www.routledge.com/Advances-in-Computational-Collective-Intelligence/book-series/ACCICRC>

Social Media and Crowdsourcing Application and Analytics

Edited by

Dr. Sujoy Chatterjee

Dr. Thipendra P. Singh

Dr. Sunghoon Lim

Dr. Anirban Mukhopadhyay



CRC Press

Taylor & Francis Group

Boca Raton London New York

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

AN AUERBACH BOOK

First edition published 2024

by CRC Press

2385 Executive Center Drive, Suite 320, Boca Raton, FL 33431

and by CRC Press

4 Park Square, Milton Park, Abingdon, Oxon, OX14 4RN

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

© 2024 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 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

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

ISBN: 978-1-032-38687-4 (hbk)

ISBN: 978-1-032-38690-4 (pbk)

ISBN: 978-1-003-34632-6 (ebk)

DOI: 10.1201/9781003346326

Typeset in Minion

by MPS Limited, Dehradun

Contents

Preface.....	xi
--------------	----

Chapter 1	Introduction to Crowdsourcing.....	1
------------------	---	----------

Bikram Pratim Bhuyan and Manisha Singh

1.1	Introduction.....	1
1.2	Major Types of Crowdsourcing Platforms	3
1.2.1	Crowdfunding as a Type of Crowdsourcing.....	3
1.2.2	Microtasking as a Type of Crowdsourcing.....	6
1.2.3	Open Innovation as a Type of Crowdsourcing.....	10
1.2.4	Creative Sense as a Type of Crowdsourcing.....	14
1.2.5	Crowd Testing as a Type of Crowdsourcing.....	17
1.2.6	Collaborative Knowledge as a Type of Crowdsourcing	21
1.3	The Good, Bad, and Ugly of Crowdsourcing	24
1.4	Conclusion	27
	References.....	28

Chapter 2	Crowdsourcing through a Digital Ecosystem:
------------------	---

A Futuristic Tool of Remote Behavior

Modification.....	32
--------------------------	-----------

Rajesh Verma

2.1	Introduction.....	33
2.1.1	Definition.....	33
2.1.2	Crowdsourcing: Concept Note	34
2.2	Glance at a Primary Component of Crowdsourcing.....	35
2.3	Remote Behavior Regulation.....	37
2.4	Crowdsourcing and Behavior Modification	39

2.5	Case of Masked Strategy	40
2.6	Reverse Crowdsourcing (RC) and Behavior Modification	41
2.7	Social Media and Crowdsourcing	43
2.8	Futuristic Perspectives	43
2.8.1	Improved Predictive Accuracy	43
2.8.2	Improving Accuracy of Artificial Intelligence with Big Data	44
2.9	Futuristic Applications	44
2.10	Conclusion	45
	References.....	46

Chapter 3 Crowdsourcing Applications in Agriculture..... 49

*Achala Shakya, Gaurav Tripathi,
and Devarani Devi Ningombam*

3.1	Crowdsourcing Applications in the Agricultural Field and Farming Activities	49
3.2	Crowdsourcing of Data/Knowledge from Various Sensor Devices	53
3.2.1	Agricultural Land-Use/Land-Cover Data.....	54
3.2.2	Weather Information.....	55
3.2.3	Soil Data.....	56
3.2.4	Crop Phenology.....	56
3.2.5	Data Related to Diseases on Plants	58
3.2.6	Vegetation Yield Status.....	59
3.3	Recommendations for Data Quality Control.....	61
3.4	Comparison of Agricultural Crowdsourcing at the Global Level and Indian Perspective.....	63
	References.....	67

Chapter 4 Crowdsourcing Outline for Contemporary Aided Medicinal Backup Systems..... 71

Sonali Vyas, Shaurya Gupta, and Vinod Kumar Shukla

4.1	Introduction.....	71
4.2	Literature Review	72
4.3	Mental Health Supportive Architecture	73

4.4	Crowdsourcing Systems.....	74
4.5	Neighbor-Backed Medicinal Crisis Structure.....	76
4.6	Conclusion and Future Scope.....	79
	References.....	80

Chapter 5 Crowdsourcing-Based Recommendations in Travel and Tourism Market: A Review..... 84

Sujoy Chatterjee and Thipendra P. Singh

5.1	Introduction.....	84
5.2	Tourism-Based Recommender System Using Crowdsourcing.....	85
5.2.1	Different Variants.....	86
5.2.2	Tourism Recommendation Using Crowdsourcing.....	86
5.3	Different Steps of Proposed Approaches.....	87
5.3.1	Different Steps of the Recommendation Algorithm.....	87
5.3.2	Leveraging Crowdsourcing to Improve the Amenities and Services Provided by Hotels.....	87
5.4	Cross-cultural Recommendation Using Crowdsourcing.....	87
5.5	Crowdsourcing-Based Streaming Data Processing for Better Travel.....	88
5.6	Crowdsourcing-Based Walking Routes.....	89
5.7	Conclusion.....	90
	References.....	91

Chapter 6 A Comparison of Capability Measurement between Various Incident Handling and Reporting Schemas..... 93

*Gyana Ranjana Panigrahi, Nalini Kanta Barpanda,
Prabira Kumar Sethy, and Debabrata Samantaray*

6.1	Introduction.....	93
6.2	A Structural Data-Assembly Model for Incident Detection.....	98

- 6.2.1 Resilient Incident Reporting Formats (Schemas).....99
- 6.3 Standards for Assessing the Forms of Incident Disclosure 100
 - 6.3.1 Parameters for Seismic Performance Evaluation.....101
 - 6.3.2 Assessing the Overall Quality of a Product.....101
- 6.4 Summary of the Analysis.....104
- 6.5 Assessment Standards.....106
- 6.6 Incident Gathering and Future Notification Model 110
- 6.7 Proposed Flow Chart.....110
- 6.8 Risk Evaluation and Discussion.....111
- 6.9 Research Methodology 114
 - 6.9.1 Hypothesis 115
 - 6.9.2 Sample Size and Data Collection 115
- 6.10 Descriptive Statistics of Catastrophic, Major, Moderate, Minor, and Insignificant.....116
- 6.11 Scope and Limitations 121
- 6.12 Conclusion 121
- Acknowledgments.....122
- References.....122

Chapter 7 Performance Evaluation of Support Vector Machine Kernel Functions on Students’ Educational Data Set..... 125

Sulaiman O. Abdulsalam, Rafiu A. Ganiyu, Elijah O. Omidiora, Stephen O. Olabiyisi, and Micheal O. Arowolo

- 7.1 Introduction.....125
- 7.2 Related Works 127
- 7.3 Methodology 136
 - 7.3.1 Data Acquisition and Description 136
 - 7.3.2 Support Vector Machine (SVM) 137
- 7.4 Results and Discussion 144
- 7.5 Conclusion 146
- References.....147

Chapter 8	Survey of Sensors and Techniques in Microwave Imaging for Through-Wall Imaging for Detecting Trapped Humans under Rubble	150
	<i>Jiya A. Enoch, Ilesanmi B. Oluwafemi, and Paul K. Olulope</i>	
8.1	Introduction.....	150
8.2	Building Collapse Information-Gathering Techniques	153
8.3	Review of Related Works.....	156
8.4	Conclusion	170
	References.....	171
 Chapter 9	 A Comparative Study on Available Fake Detection Apps.....	 178
	<i>Sakshi Gosain, Shrutika Saxena, Tapas Kumar, and Shweta Mongia</i>	
9.1	Introduction.....	178
9.2	Literature Survey.....	179
9.2.1	Features of Fake News	179
9.2.2	Identifying Fake News Methods	180
9.2.3	Conventional Techniques for Fake News.....	182
9.2.4	Machine Learning Technique	183
9.2.5	Fake News Detection Apps and Websites.....	186
9.2.6	Websites Used for Detecting Fake News.....	190
9.3	Conclusion	190
	References.....	191
 Chapter 10	 Crowdsourcing Mechanisms for Reviving Cultural Heritage	 194
	<i>Bikram Pratim Bhuyan and Ravi Tomar</i>	
10.1	Introduction.....	194
10.2	Advantages of Using Crowdsourcing-Based Techniques.....	195
10.3	Types of Crowdsourcing Mechanisms for Reviving Cultural Heritage	198
10.3.1	Crowdfunding.....	199
10.3.2	Crowdsourced Mapping	203
10.3.3	Citizen Science.....	206

10.3.4	Crowdsourced Translation	208
10.3.5	Crowdsourced Archiving.....	210
10.4	Challenges	211
10.5	Conclusion	213
	References.....	213
Chapter 11	Ethical Issues in Crowdsourcing.....	217
	<i>Bikram Pratim Bhuyan and Thipendra P. Singh</i>	
11.1	Introduction.....	217
11.2	Background Behind Crowdsourcing.....	218
11.3	Crowdsourcing's Coming of Age	223
11.4	Issues in Crowdsourcing	226
11.5	Ethical Issues in Crowdsourcing	229
11.6	Discussion	238
11.7	Conclusion and Future.....	241
	References.....	241

Preface

Recently, due to the development of the internet and mobile devices, such as smartphones and tablets, it has been easy to share information and communicate across the globe. Especially social media and crowdsourcing have significantly contributed to various traditional and new industries, such as manufacturing, social good, agriculture, healthcare, tourism, disaster management, education, crime control, and cultural heritage, as efficient and effective platforms for individuals and organizations to share information and communicate with each other.

Social media is defined as online platforms that allow users to communicate with each other, share information and knowledge, and express their opinions on various topics, regardless of their physical locations, in real time. Social media users can use texts, images, videos, and URLs to create content and communicate with each other as well as create their personal profiles. Over the last couple of years, social media platforms have been used not only for individual users but also for communities and enterprises for marketing and business purposes. Types of social media include social networking sites, blogs, photo-sharing sites, video-sharing sites, instant messaging, podcasts, and virtual worlds as well as social media users can use their computers and mobile devices. Popular social media platforms are Twitter, Facebook, Instagram, YouTube, Yelp, and LinkedIn and they have been widely used as powerful platforms for crowdsourcing recently.

Crowdsourcing is defined as a sourcing model in which individuals and/or organizations obtain goods, services, ideas, and finances from a large group of online users through an open call. Crowdsourcing is also considered a way to use the Internet to outsource work to a crowd, either paid or unpaid. It has been widely used for individuals and organizations to have collective knowledge and skills as well as create global communities to achieve the common goals. Especially, it contributes to create encyclopedias, fund projects, and address real-world complex problems, which traditional approaches struggle to solve. Crowdsourcing is based on the wisdom of crowds, which is the process of considering the collective opinion of a group of individuals, rather than a single expert to answer a question, as well as requires a

diversity of opinions, independence, decentralization, and aggregation. While outsourcing is commissioned from a specific and named group, crowdsourcing comes from a less specific and more public group. Compared to outsourcing, it is known that the advantages of crowdsourcing include improved profits, speed, quality, flexibility, scalability, and diversity. It is demonstrated that proper utilization of non-expert human beings can reduce the overall cost of hiring the experts while maintaining quality.

Types of crowdsourcing include collecting information, labor markets, collaborative creation, and smartest in the crowd and their popular examples are online commerce websites' reviews, Amazon Mechanical Turk, Wikipedia, and online contests, respectively. In addition, crowdfunding is one of the example of crowdsourcing, which is the collective effort of individuals who network and pool their money, usually via the Internet, to support efforts initiated by other people or organizations. Crowdfunding includes disaster relief, start-up company funding, free software development, and scientific research that supports individuals, non-profit organizations, start-ups, and small companies. Popular crowdfunding websites are GoFundMe and Kickstarter.

Social media can be used as an efficient and effective platform for crowdsourcing. Through social media platforms, individuals and organizations can use knowledge and opinions that are collected from the crowd in real time and can be supported to make decisions. For example, social media platforms, especially online communities, can be used for idea generation and brainstorming. Social media platforms including online communities also enable individuals and/or organizations to use online polls and surveys, which are necessary for crowdsourcing. Social media platforms can be used for collecting various contents, including texts, photos, and videos, which are generated by online users and used for marketing and/or business as well. In addition, social media platforms can be a powerful source for crowdfunding.

In this book, we have discussed the concept of social media and crowdsourcing as well as their analytical aspects. We have also explored their contributions to the real world, such as social good, agriculture, healthcare, tourism, disaster management, education, crime control, and cultural heritage. In addition, we discussed ethical issues in crowdsourcing and future scopes and challenges for policy makers who use social media and crowdsourced opinions. Due to the limited

scope of this book, some concepts and procedures have been introduced without theoretical explanations.

We would like to thank all the authors of this book for their valuable contributions. We hope that this book will provide practical knowledge and valuable insights to readers. In addition, we hope this book will contribute to the development of this research area.

Performance Evaluation of Support Vector Machine Kernel Functions on Students' Educational Data Set

*Sulaiman O. Abdulsalam¹, Rafiu A. Ganiyu²,
Elijah O. Omidiora², Stephen O. Olabiyisi²,
and Micheal O. Arowolo³*

¹Department of Computer Science, Kwara State University, Malete, Nigeria

²Department of Computer Science and Engineering, Ladoke Akintola University of Technology, Ogbomoso, Nigeria

³Department of Electrical Engineering and Computer Science, University of Missouri, Columbia, Missouri, United States of America

7.1 INTRODUCTION

Data mining, when applied to academic domain brings innovative methodologies, techniques, and technologies to the development of the student learning. In order to acquire an improved understanding of the undergraduates' educational process, the most recent advancement offers helpful tools for analyzing and interpreting large data sets employing knowledge discovery techniques. Today's academic institutions function in a challenging and competitive world. Academic institutions experience many problems, such as assessing and comparing, ensuring excellent education, devising methods for measuring student progress, and envisioning the future. These establishments must enhance student preventative strategies in order to address difficulties students face throughout their schooling (Albreiki et al., 2021).

Education is one of society's most important pillars. It controls students' character and intellect. The present educational scheme may not be adequate to meet society's changing and dynamic needs. One of the most significant features of the new educational paradigm is the skill to forecast student performance in advance. Because students are the primary stakeholders in educational systems, academic institutions can meet the changing needs of society by evaluating student data and emerging various predictions based on it. Furthermore, the outcomes of predictions can be useful in developing approaches to advance educational quality. A higher-quality education aids in the development of skilled and distinctive students. This lets you analyze academic information. Utilizing machine learning techniques, student achievement classifiers analyze information of students. Many academic achievement classifiers exist. Both the research and treatment communities have considered students achievement predictive models. Academic achievement predictions scientists suggest semesters, GPA, CGPA, and pass/fail courses (Zaffar et al., 2022).

In EDM (educational data mining), the goal of developing computer models for academics is not simply to obtain optimal precision of prediction, but additionally to aid instructional authorities in evaluating academic achievement. Learners are a society's most important asset, and the primary objective of every academic institution is to provide them with a higher educational attainment. Moreover, great education contributes to the development of talented and unique students. Models for predicting student achievement facilitate the examination of data collected, utilizing numerous data mining techniques. The creation of models for predicting student achievement has required considerable time and effort. There are two methods for developing standardized testing prediction models. The first technique is supervised, but the second is unsupervised. Classification is a form of supervised learning methodology. The parameter out of which we want to infer whether or not a student passed or failed was provided explicitly as the input vector in the clearly indicate. Therefore, we decided to use the classification approach in a model that predicts students' grades (Guo et al., 2015).

Evaluating students' work helps teachers and other educators draw useful conclusions and make necessary adjustments so that they can better serve students and the evolving needs of society. When trying to improve forecast accuracy and create actionable plans for raising students' academic performance, the attributes chosen for use in making those predictions are crucial. The literature suggests that present methods have pros and cons

when it comes to determining ideal attributes for machine learning models utilized to predict student performance (Zaffar et al., 2022).

The performance of four Support Vector Machine (SVM) kernel functions, polynomial function (PF), linear function (LF), radial basis function (RBF), and sigmoid function (SF), are investigated using the grid search method, with the objective of determining the best kernel function, using students' educational data set obtained from the Department of Computer Science in Nigerian universities in the north-central region from 2009 to 2014.

7.2 RELATED WORKS

Students' performance prediction model using Iterative Dichotomiser 3 (ID3) and C4.5 classification algorithms in a Nigerian University was developed. The data set used contained students' academic records such as gender, SSCE grade, age, post/Unified Tertiary Matriculation Examination (UMTE) scores, and the first-year student grade. From the result of their work, ID3 algorithm outperforms the C4.5 with respect to identification of significant attributes that contributed greatly to the determination of students' academic performance (Afeni 2019).

An artificial neural network (ANN) is utilized to analyze and forecast students' CGPAs based on their socioeconomic level and entrance exam scores for undergraduate education at a Chinese institution. Through the use of confusion matrix, error histogram, regression analysis, and mean square error, ANN performance was examined. ANN attained an accuracy of 84.8 percent and an Area under the curve (AUC) of 0.86. Despite this, ANN performed badly in identifying undergraduates due to their gender and produced a high rate of false negatives as a result of a high imbalanced ratio data sample (Lau et al., 2019).

A log-data segmentation approach was suggested for determining student disengagement in introductory programming programs by identifying individuals who really remain focused with programming lectures. During a 90-minute examination of 39 beginning-level computer program learners, the research employed a way to predict undergraduate withdrawals utilizing the virtual terminal connection input sequence, outliers classification, and an unsupervised machine learning methodology (Oeda & Hashimoto, 2017).

A classification model for predicting critical courses that affects student's performance using ID3 decision tree initiation algorithm was developed (Altujjar et al., 2016). The researcher model was built using 100 records of feminine students in the King Saud University, Information Technology Department. Their result divulged those dependable predictions are attained due to the performance of scholars in the second year progresses and also discovered significant courses that help as pointers of scholar academic performance.

Employing a Waikato Environment for Knowledge Analysis (WEKA) information mining application, a comprehensive review of classification models, Nave Bayes (NB), a Bayesian network, Instance-based learning with parameter k (IBk) approach, a slow predictor, and Classification and Regression Tree (CART), a tree-based technique, was conducted using the true positive rate, false positive rate, and productivity accuracy. They utilized the educational data of 385 individuals with seven parameters (sex, age, entrance method, city of residence, denomination, 200-level CGPA, and final examination) from the Computer Science Department and Technology at LAUTECH. In contrast to NB and CART, which had performance levels of 64 percent and 62 percent, separately, IBk was capable of making predictions for all stages of education with a highest accuracy of 86 percent. Furthermore, the outcomes of the performance analysis revealed that Naïve Bayes, Instant Based, as well as CART had comparable results. They discovered that IBk provided the performance accuracy (PA) with the smallest normality test and also performed the greatest with increased true positive rating (TPR) intensity (Ayinde et al., 2015).

The influence of the quality for predicting the academic performance of first college freshmen premised on their degree relatives and previous educational attainment was suggested. They examined the data of 1,500 computer science majors at Babcock Nigerian University between 2001 and 2010. In developing a prediction model, ten learning algorithms (five classification tree methods: Random Forest, J48, Decision tree [dt], REPTree, and decision tree trunk, and five supervised learning procedures: OneR, JRip, PART, ZeroR, and Classification model) and a multilayered perceptual were utilized. Utilizing the tenfold cross-validation (10-FCV) and retention approaches, the accuracy, confusion matrix, and computational efficiency of the produced models were measured. It was discovered that the Random tree method outperformed other strategies employed in their research (Kuyoro et al., 2013).

An improved SVM feature selection algorithm is widely employed for the large-scale email dataset. Particle swarm optimization (PSO) procedure was employed for finding the optimal parameters values and also for obtaining optimal feature subset. The result of this learning revealed that the improved SVM procedure generated a classification accuracy while the standalone SVM produced an accuracy of 68 percent in 6.33 seconds for e-mail data set of 1,000. In addition, with the use of 3,000 e-mail data set, the optimized SVM and standalone SVM yielded classification accuracy and computation time of 91 percent, 0.60 second and 47 percent, and 60 seconds, respectively. Likewise, using 6,000 e-mail data set, 93 percent, 0.20 second and 18 percent, 92 seconds were achieved for augmented SVM method and standalone SVM individually. The researchers argued that the experimental results acquired establish that for each rise in data set dimensions, there is constantly a lessening in the classification of SVM classifier with a significant rise in computational time. This study concluded that the optimized SVM technique outperformed the standalone SVM with respect to classification accuracy and time consumption in the face of huge e-mail data set (Oludare et al., 2014).

An optimization of RBF kernel parameters to attain maximum accuracy value in nonlinear multiclass SVM classifier for the recognition of images of Indonesian Batik data set with geometric decorative patterns. The researcher applied grid search method with cross-validation (CV) process to optimize a range of values of parameters C (Cost) and γ (Gamma) of SVM-RBF kernel. Additionally, the discrete wavelet transform (DWT) algorithm was used for feature extraction and the extracted feature data set was employed for training and testing sets. The experimental result produced a maximum accuracy value for the Batik data set when using the constraints $C = 27$ and $\gamma = 2-15$. The researcher concluded that the obtained optimal values and reduced variety of constraint principles can be utilized as a reference for numerary image recognition with qualities keeping geometric enhancing patterns with SVM-RBF kernel classification (Budiman, 2019).

Afeni et al. (2019) developed a students' performance prediction model using ID3 and C4.5 classification procedures at a Nigerian university. The data set used contained students' academic records such as gender, SSCE grade, age, post/UMTE scores, and the first-year student grade. From the result of their work, ID3 algorithm outperforms the C4.5 with respect to identification of significant attributes that contributed greatly to the determination of students' academic performance.

Lau et al. (2019) utilized the ANN to analyze and evaluate the individuals' cumulative grade point average based on their social economic level and entry-standardized test scores for undergraduate studies at a Chinese institution. Through using a prediction model, multiple regressions, inaccuracy distribution, and root-mean-square average, the functionality of ANN was examined. ANN achieved an accuracy rate of 85 percent and an AUC of 0.90. Nevertheless, the performance of ANN was poor in categorizing scholars with respect to their gender and achieved a high false negative rate result due to a high imbalance ratio data sample.

In light of assessing standardized testing data, Oeda and Hashimoto (2017) suggested utilizing a file segmentation technique to predict learner dropouts in introductory training. This approach might be necessary for universal screening, especially for individuals who struggle with computational training. During such a 90-minute examination of 39 beginning programmers, the researchers used UNIX command input records to develop a technique for anticipating learner attrition by applying feature extraction to clustering records with an unsupervised learning approach.

Altujjar et al. (2016) developed a classification model for envisaging critical courses that affects student's performance using ID3 decision tree induction algorithm. The researcher model was built using 100 records of feminine students in the King Saud University, Information Technology Department. Their result divulged that dependable predictions are attained due to the performance of scholars in the second year progresses and also discovered important developments that serve as pointers of student academic performance.

Ayinde et al. (2015) performed a comparative analysis of three classifiers, Naïve Bayes, a Bayesian network, IBk procedure, a lazy classifier, and CART, a decision tree procedure in terms of true positive rate, false positive rate, and performance accuracy in mining students' educational data using WEKA data mining software. They used 385 students' academic records, with seven attributes each (gender, age, admission mode, state of origin, religion, 200 level CGPA, and students' final grade), of the Computer Science and Engineering, Department at LAUTECH. Their results revealed that IBk was able to expect for all the classes of grades with an overall performance rate of 86 percent, unlike NB and CART with a performance rate of 64 percent and 62 percent, respectively. The average TPR for NB, IBk, and CART was 55.4 percent, 79.4 percent, and 59 percent, consecutively; the average FPR was 58 percent, 42 percent,

and 66 percent; and the average PA score was 59 percent, 82 percent, and 69 percent. They concluded that IBk gave the PA with the lowermost insignificant hypothesis and also performs best with increasing value of TPR.

The best method for predicting the hypothetical success of first-year postsecondary students based on their family history and past academic achievements was discovered by Kuyoro et al. (2013). They analyzed information from 1,500 CS majors at Nigeria's Babcock University from 2001 to 2010. Ten classification methods (five classification tree methods: Random Forest, J48, Random tree, REPTree, and Decision root, and five supervised learning algorithms: OneR, JRip, PART, ZeroR, and Decision tree) and a multilayered perceptual were utilized in generating classification techniques. The resulting models were verified both 10-FCV and retention methods to estimate of average accuracy, confusion matrices and computational efficiency. Results exhibited that the Random tree method achieved improved than the others tested.

Oyedele et al. (2010) predicted students' performance with the K-means clustering procedure integrated with the deterministic model based on the data set of a private institution in Nigeria. A total number of 79 students that offered nine courses each in a semester was sampled and a scheme for examining students' outcomes-based clustering analysis and utilized statistical processes to position their scores' information conferring to the degree of performance was demonstrated.

Ayinde et al. (2013) applied Naïve Bayes and Decision Stump procedures in predicting the scholars' concluding year mark at Ladoko Akintola University of Technology, Ogbomoso, Nigeria. A total of 473 students comparing 10 attributes namely, gender, age, admission mode, faith, pre-degree grade, scholar enrolment number, origin, 200-level CGPA, 500-level GPA, and students' final grades were used. The outcomes of their study divulged that the Naïve Bayes with an average precision of 77 percent performed better than the decision stump algorithm with an average precision of 57 percent.

Akinola et al. (2012) utilized the ANN to advance a data mining model for performance of computer science students' prediction in computer programming development taken at the 200 level using their pre-qualification ordinary level results and their results obtained in 100-level mathematics and physics courses. Two hundred records of scholars in the Computer Science Department, University of Ibadan, were used for the study and it was discovered that a previous background knowledge of

mathematics and physics were vital in instruction for a scholar excelling in computer programming.

Minaei-Bidgoli's (2003) research attempted a method to predict students' final evaluation by classifying them based on extracted features from recorded information in a teaching web-based system. An ensemble of classifiers was designed, implemented, and evaluated on an online programmed data set at Michigan State University (USA), relating to 2002 data. Conclusions from a methodological viewpoint show that the use of multiple classifiers resulted in a major upsurge in classification accuracy (as research in ensemble methods confirms) and that this method could be extremely helpful in recognizing scholars at risk early on, particularly in large classes, and allowing the teacher to deliver timely advising.

Barker et al. (2004) studied the problem in a perspective of classifying student graduation performance from numerous demographic, academic, and attitudinal variables. They used neural networks and SVM algorithms over data from the first-year advising center of the University of Oklahoma (USA), composed via a review assumed to all inbound freshmen. An overall conclusion of this study is the lack of area information about the data set and that underprivileged outcomes (one-third of the students were misclassified) are not the fault of the algorithm. Delivering a pessimistic view, the authors realized that it may be difficult to capture appropriate and suitable information about student completion.

Herzog (2005) research was more specific as it measured the influence of high college training, first-year academic performance, multi-institute enrollment, and economic aid sustenance on second-year determination. A methodology was developed using multinomial logistic regression. The recognized scholar data system included data on scholar demographics, academics, economic support, and student on-campus jobs data, as well as the institution's payroll system. The findings show the importance of looking at risk issues from both a dropout and a transfer-out viewpoint, while taking into account the possibility that a scholar might be enrolled at added institution at the same time. When evaluating freshmen retention, the findings show that first-year mathematics knowledge, math strength of the acknowledged major, concurrent enrollment at extra higher education institutions, and second-year economic aid suggestions are all important factors to consider. Furthermore, middle-income undergraduates with more unmet needs are more probable to drop out, while educationally well-prepared freshers with unmet needs are more probable to move.

Herzog (2006), in further research, took another perspective on predicting retaining and graduation time using decision trees, neural networks, and logistic regression. Like previous research, data from the institutional student information system was used, together with parent income data, and a system to identify transfer-out students. The analysis revealed that when analyzing the generalization ability of numerous data mining procedures to regression models, the complexities of the data set utilized and the primary objective may greatly guide the selection of an evaluation method. Decision tree algorithm and neural networks can help analysts discover which qualities improve prediction performance when interacting with huge amounts of information.

Romero et al. (2008) compared various data mining approaches for classifying scholars based on their Moodle “e-learning system” use information and concluding characters in their individual programs. The methodology relied on a particular mining tool to make data mining technique setup and execution simpler for instructors. The data included students’ usage log in a web-based program of seven Moodle programs in the University of Córdoba (Spain) and the final marks obtained. The algorithms used were neural networks and genetic and decision trees. When preprocessing tasks (discretization and rebalancing data) were applied, some algorithms improved their classification efficiency, while others did not. A classification algorithm intended for pedagogical usage must be exact and straightforward for instructors to be beneficial for strategic planning.

Dekker et al. (2009) research aimed to predict which electrical engineering scholars would drop out after their initial semester and to define success factors unique to that program. Over the years 2000 to 2009, data was collected on 648 Eindhoven University of Technology freshman students (Netherlands). A Bayesian classifier, logistical algorithm, principal classifier, and decision tree model were examined. The OneR predictor also functioned as a measure and estimator of variable predictiveness. Decision trees have valuable insights with accuracies between 75 and 80 percent, according to experimental results. Conclusions delve into showing ways of further improvement in prediction without having to collect additional data about the students.

Zhang et al. (2010) wanted to recognize potential difficulties as primary as conceivable and to trail with involvement decisions to improve student retention. For that purpose, a system was built, including data mining techniques, using Naïve Bayes, SVM, and decision trees algorithms. This

system monitored and analyzed student academic behavior with the aim to provide a basis for efficient intervention strategies. The data used was from the university system of the Thames Valley University (UK). The main conclusions were that Naïve Bayes attained the maximum prediction accuracy outperforming decision trees and SVM.

Socio-demographic factors (such as gender, age, ethnicity, employment position, educational attainment, and disability) and the learning environment were investigated by Kovacic's (2010) research. From 2006 to 2009, 450 scholars at the Open Polytechnic of Wellington (New Zealand) were included in the data set. The strategy involved a combination of cross-tabulation analysis (situational tables), feature selection, and prediction using four distinct kinds of decision trees. The most important criteria between successful and unsuccessful students were found to be students' race and their participation in a particular program. After making some methodological considerations, we found that the CART model achieved the highest accuracy (60.5 percent total) in our classification task.

Kabakchieva (2012) aims to illustrate the promising importance of data mining systems for school administrators in order to enable academic institutions run increasingly profitable enrollment campaigns and recruit the brightest students. The determination of this study was to advance data-mining algorithms that might predict a student's future abstract performance based on their unique set of pre-university, collegiate, and post-collegiate experiences. Many different algorithms, including the neural networks, OneR rule learning, decision trees, and k-nearest neighbor (k-NN) classifiers, were put to the test. The data set used in this research is comprised of three years' worth of admissions data from the University of National and Financial System in Sofia. Specifically, the outcomes exhibited that the neural network model had the precise results, followed by the prediction model and the k-NN model. Attributes of data pertaining to students' university entrance scores and the number of unsuccessful attempts on that first university studies are two of the most significant criteria in the classification stage.

Decision tree algorithms, ANNs, naive classification algorithms, instance-based learning, regression models, and SVMs were among the supervised machine learning algorithms developed by Kotsiantis et al. (2004) to expect scholar success at the Hellenic Open University based on key demographic variable measure and assignment marks. The percentage of correct predictions wide-ranging from 58 percent (when a

neural network was utilized) to 65 percent (when only demographic variables were included) (when using SVMs). The Naïve Bayes classifier proved to be the greatest effective method for expecting student success when additional data were included in addition to demographics.

Students' risk levels were determined using a combination of neural networks, decision trees, and linear discriminant analysis by Vandamme et al. (2007). The academic success of first-year scholars at French colleges in the kingdom of Belgium was found to be significantly correlated with a variety of demographic and personal factors. Students' academic success was best predicted by their level of education, the number of hours they spent on mathematics, their financial stability, and their age, but not by their gender, their parents' education or career, or whether or not they were married. But none of the three techniques used to foretell moot success actually worked. Overall, the classification accuracy of decision trees was 40.63 percent, that of neural networks was 51.88 percent, and that of discriminant analysis was 57.35 percent.

Secondary student grades were estimated in two groups by Cortez and Silva (2008) using a variety of factors including prior academic performance, demographics, social information, and other school-related information. To get there, we used a variety of machine learning systems like neural networks, random forests, decision trees, and SVMs. When taking into account prior grades, they proved to be rather accurate in predicting future performance. In certain situations, they incorporated socioeconomic factors into their models in addition to demographic factors (such as a student's age, their parents' occupation, and their level of education). There were several factors that Open Polytechnic students simply couldn't get to (for example, previous student grades).

Predictive model Chi-square automatic interaction detector (CHAID) with a seven-class response variable was developed by Ramaswami and Bhaskaran (2010) to evaluate students' performance in Indian higher secondary institutions. The model relied on highly impacting predictive factors obtained through feature selection. The CHAID prediction model was built using data from several different Tamilnadu universities and 772 student records. Therefore, the CHAID prediction model was used to generate a set of rules, and the system's efficacy was evaluated.

The Course Registration Planning Model (CRPM) was first proposed by Pathom et al. (2008), and its development was proposed by a classifier algorithm. The four classifiers that were considered for this task were the Bayesian network, C4.5, the decision forest, and the NB tree.

Undergraduate enrollments, including GPAs and grades, were used to generate the data.

Lykourantzou et al. (2009) used three algorithms of machine learning and individual information to identify e-learning withdrawals. The Municipal Tech College of Athens' Collaborative Communication Institute sponsors the e-learning crew (Greece). Feed-forward neural pathways, SVMs, and deterministic ensembles fuzzified ARTMAP were in use. On the whole, the strategy's consistency, responsivity, and specificity were significantly comparable to those shown in the field.

The phrases "stop out," "maintained," and "transferable" were defined by Sujitparapitaya (2006). The determination of this case study was to recognize the most important factors that freshmen consider while making their final decision. Based on the model, incoming freshmen can choose between staying at their current school, transferring to a different post-secondary institution, or dropping out altogether. The data used came from the National Student Clearinghouse, which is widely recognized as an authoritative repository for academic credentialing and student performance statistics across American universities. The algorithms deployed included logistic regression, neural networks, and C5.0 (decision tree). In terms of predicting first-year retention, the C5.0 rule induction model looks to do slightly better than the other two. The findings are noteworthy because they show that changes in campus, scholar, and parental values are just as important as changes in basic norms and rewards for effective retention programs that increase student learning.

7.3 METHODOLOGY

7.3.1 Data Acquisition and Description

The experimental set of data used in this research is a multi-class student educational data set obtained from the Department of Computer Science in a university in north-central region of Nigeria. The data set is made up of records of 153 graduated students with variables sex, age, and their 64 respective courses offered as the predictor variables and their obtained final year grades as the class label. The descriptions of predictor variables (sex, age, and students' scores in different courses offered) and class label (final grade obtained at graduation) used in the data sets, the final grade

obtained at graduation had five distinct values: (First, Second Upper, Second Lower, Third Classes, and Pass).

The students' educational data set was first separated into two portions of 80:20 ratios. Eighty percent of the data went through 10-FCV, while the remaining 20 percent served as the validation set and was put through its paces in the testing phase. In other words, out of a total of 153 samples, 122 were used to train/build the four SVM models (with a linear, polynomial, radial basis, and sigmoidal SVM Kernel function), and the outstanding 31 examples were used as a validation test set to determine how well the developed hybrid rule extraction method from SVM performed.

7.3.2 Support Vector Machine (SVM)

The SVM is an educational method that is founded on the statistical learning model (Vapnik, 1995). SVM has gained popularity as a potent data mining technology that can be used to address the challenge of knowledge extraction (Borges-Christopher, 1998). By making use of kernel functions, SVM is able to change input qualities from lower to higher dimensions. A wide variety of applications have been utilized, such as the forecasting of financial time series (Kim, 2003), marketing (Ben-David & Lindenbaum, 1997), face recognition (Chowdhury et al., 2011), cervical cancer diagnosis, and the classification of gene expression data (Das et al., 2020).

For classification techniques, SVM aims to discover an ideal hyperplane that maximizes the margin that accurately distinguishes data points as much as feasible and divides class labels to the greatest extent practicable, while minimizing the likelihood of mishandling training data set and unlabeled testing set. The purpose is to discover a hyperplane that maximizes the boundary that meaningfulness information points and separates them as much as feasible (Vapnik, 1998). Other training instances can't define binary class bounds. Support vectors are training spots near the optimal separation hyperplane. SVM projects nonlinearly recoverable information to a higher dimensional space and then looks for the linear boundary. SVMs can tackle linear and nonlinear sorting issues, but their bid to multi-class challenges is restricted because of its one-to-one methodology. This offers SVMs a competitive advantage (Sampon et al., 2011). Numerous individuals present a multiclass challenge as many problems involve binary classification (one-to-all difficulties) (Luts et al., 2010).

Provided a training set of instance-label pair (x_i, y_i) , $i = 1, 2, \dots, m$ where $x_i \in R^n$ and $y_i \in \{+1, -1\}$, SVM identifies an ideal extrication hyperplane with the determined boundary by resolving the subsequent optimization task:

$$\text{Min}_{w,b} \quad \frac{1}{2} w^T w \quad (7.1)$$

Subject to: $y_i(\langle w \cdot x_i \rangle + b) - 1 \geq 0$

where x_i is an actual-valued n -dimensional input trajectory and y_i is a tag that defines the class of x_i . An untying hyperplane is achieved through an impertinent vector w and a partiality b .

It is well understood that the burden point of the Lagrange role is found to resolve this quadratic equation optimization issue:

$$L_P(w, b, \alpha) = \frac{1}{2} w^T \cdot w - \sum_{i=1}^m (\alpha_i y_i (\langle w \cdot x_i \rangle + b) - 1) \quad (7.2)$$

where the α_i denotes Lagrange multipliers, hence $\alpha_i \geq 0$. The exploration for an optimum burden opinion is essential since the L_P necessitate is diminished with reverence to the primal variables w and b and exploited with reverence to the non-negative double variable α_i . By distinguishing with reverence to w and b , and familiarizing the Karush Kuhn–Tucker (KKT) approval for the optimal inhibited function, then L_P is distorted to the double Lagrangian $L_D(\alpha)$:

$$\max_{\alpha} L_D(\alpha) = \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i=1}^m \alpha_i \alpha_j y_i y_j \langle x_i, x_j \rangle \quad (7.3)$$

Subject to: $\alpha_i \geq 0 \ i = 1, \dots, m$ and $\sum_{i=1}^m \alpha_i y_i = 0$

Finding the best hyperplane, a twofold Lagrange $L_D(\alpha)$ necessitates exploiting with reverence to non-negative α_i . The answer α_i for the twofold optimization issue describes the constraints w^* and b^* of the ideal hyperplane. Hence, the ideal hyperplane result function $f(x) = \text{sgn}(\langle w^* \cdot x \rangle + b^*)$ can be described as

$$f(x) = \text{sgn} \left(\sum_{i=1}^m y_i \alpha_i^* \langle x_i, x \rangle + b^* \right) \quad (7.4)$$

Small subclass of the Lagrange multipliers α_i is typically superior than zero in a normal classification task. These vectors are the most geometrically similar to the ideal hyperplane. The relevant training vectors having non-zero α_i are termed support vectors, as the best verdict hyperplane $f(x, \alpha^*, b^*)$ depends on them entirely.

These notions also apply to non-separable cases (linear comprehensive SVM). In terms of these new inequality constraints, the task of discovering the hyperplane with the fewest training mistakes (i.e., minimizing constriction error) is expressed as follows:

$$\min_{w, b, \xi} \frac{1}{2} w^T w + C \sum_{i=1}^m \xi_i \quad (7.5)$$

Subject to: $y_i(\langle w, x_i \rangle + b) + \xi_i - 1 \geq 0 \quad \xi_i \geq 0$

where C is the non-negative relaxed variables and is a penalized variable on the training error. By analyzing all possible hyperplanes, SVM finds the one that results in the fewest blunders during training (that is, to retain the limit error as minor as conceivable).

This optimization model can be solved using the Lagrangian approach, which is corresponding to the method used to resolve the optimization technique in the independent assortment. The Lagrangian with two independent variables must be optimized:

$$\max_{\alpha} L_D(\alpha) = \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i=1}^m \alpha_i \alpha_j y_i y_j \langle x_i, x_j \rangle \quad (7.6)$$

Subject to: $0 \leq \alpha_i \leq C \quad i = 1, \dots, m$ and $\sum_{i=1}^m \alpha_i y_i = 0$

Finding an optimum hyperplane, a twofold Lagrange $L_D(\alpha)$ needs to be exploited with reverence to non-negative α_i under the constraint $\sum_{i=1}^m \alpha_i y_i = 0$ and $0 \leq \alpha_i \leq C$. The consequence constraint C , which is the higher bound on α_i , is resolute by the user. Lastly, the procedure of best hyperplane result function is the same as (7.4).

The nonlinear SVM plans the training models from the input space into a higher-dimensional feature planetary via a plotting purpose ϕ . In the twofold Lagrange (7.6), the internal products are swapped by the kernel function (7.7), and the nonlinear SVM dual Lagrange $L_D(\alpha)$ (7.8) is related with the linear global instance.

$$(\Phi(x_i) \cdot \Phi(x_j)) := k(x_i, x_j) \quad (7.7)$$

$$L_D(\alpha) = \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i=1}^m \alpha_i \alpha_j y_i y_j k(x_i, x_j) \quad (7.8)$$

Subject to: $0 \leq \alpha_i \leq C$ $i = 1, \dots, m$ and $\sum_{i=1}^m \alpha_i y_i = 0$.

The decision function is obtained in the following form after the measures defined in the linear generalized case:

$$\begin{aligned} f(x) &= \text{sgn} \left(\sum_{i=1}^m y_i \alpha_i^* \langle \Phi(x), \Phi(x_i) \rangle + b^* \right) \\ &= \text{sgn} \left(\sum_{i=1}^m y_i \alpha_i^* \langle x_i, x \rangle + b^* \right) \end{aligned} \quad (7.9)$$

By selectively plotting their involvements into high-dimensional feature spaces, SVM is able to efficiently execute nonlinear classification thanks to the kernel trick. The essential roles that a system should serve have been proposed in abundance. The computational efficiency of SVM can be enhanced by replacing its fundamental functions with those optimized for specific data attributes.

7.3.2.1 SVM Kernel Functions

To keep the computational complexity acceptable, SVM systems define dot products in terms of a kernel function so that they may be easily calculated in terms of the parameters in the innovative feature space $k(x,y)$ (Teukolsky et al., 2007). To generalize the optimal hyperplane with the biggest margin between two sets, an SVM uses a nonlinear kernel function to translate information from an input space to a potentially high-dimensional feature space (Chapelle et al., 2002). Kernels aid in the definition of a set of features that may be utilized to do a classification operation with ease. To improve classification precision, it is important to adjust the appropriate kernel parameters in the relevant kernel functions. There are two merits to using the kernel function. To begin with, it can be used in conjunction with linear classifiers to generate nonlinear decision boundaries. Furthermore, it allows classifiers to be applied to data that initially lacks a clear representation in a fixed-dimensional vector space (Huang et al., 2014). Generally accepted SVM kernel functions include the four possible forms:

Linear kernel function:

$$K(x_i, y_j) = x_i^t \cdot y_j \quad (7.10)$$

Polynomial kernel function:

$$K(x_i, y_j) = (\gamma x_i^t x_j + r)^m, \quad \gamma > 0 \quad (7.11)$$

Radial basis/Gaussian kernel function:

$$K(x_i, y_j) = \exp(-\gamma \|x_i - y_j\|^2), \quad \gamma > 0 \quad (7.12)$$

Sigmoid kernel function:

$$K(x_i, y_j) = \tanh(\gamma x_i^t \cdot y_j + r), \quad (7.13)$$

The kernel function is used to introduce nonlinearity and deal with arbitrary data structures. In most cases, successful classification results require fine-tuning not only the kernel function's parameters but also the other SVM parameters.

Choosing the right kernel and adjusting the right kernel parameters, as well as narrowing down the collection of functions to use, are essential steps toward regaining SVM's classification accuracy (Akay, 2009). Therefore, it begs the question, "Which SVM kernel delivers the best classification accuracy among various kernels?" Exploring the several SVM kernels to find the one with the best classification accuracy for a specific task is the only way to get an answer to this topic (Omidiora et al., 2014; Ovidiu, 2007). Finding the optimal kernel function and its appropriate kernel's parameters based on the students' educational data set, this study experiments with the four common SVM kernel functions, with linear, polynomial, radial basis, and sigmoid purposes.

7.3.2.2 Factors that Influence SVM Model Prediction Accuracy

The penalty factor C and the kernel parameters are constraints that have the greatest impact on SVM prediction accuracy. The penalty factor C calculates the penalty for data with greater deviations than precision ε . They have an

effect on classification accuracy and the capacity of the SVM model to generalize. There is, however, no universal formula for enhancing the two parameters. The grid search method is the most commonly used technique (Chen et al., 2004; Friedrichs & Igel, 2005), where the pair (C , and other kernel parameters) are checked, and the CV accuracy of the best one is selected. As a result, the grid search method employs a CV procedure to increase SVM prediction accuracy. The grid search method was used to determine the appropriate kernel's parameter and best kernel function among the four SVM kernel functions investigated in this study.

7.3.2.3 Applications of SVM Models

Pattern recognition and image processing are two examples of applications where SVM has been used (Romon & Christodoulou, 2006). SVM has been used in control engineering in recent years. SVM, on the other hand, has yet to be widely used in educational science (Huang, 2011). Kotsiantis et al. (2003) conducted a study using SVM to predict the attrition rate of first-year scholars. The data was gathered through four written assignments, one-on-one advising sessions with instructors, and comprehensive tests. Scholars at peril of falling out of college were classified using the collected data and additional variables (such as gender, age, and parent work). The results demonstrated that Support Vector Machines (SVM) outperformed neural networks after the third training stage, which incorporated data from both the second stage and the initially recorded requirements. Individual ordinal data were used in this learning by Kotsiantis et al. (2003). Moreover, no education has yet been shown on the prediction accuracy and understandability of SVM in informative investigations employing students' educational data set.

7.3.2.4 SVM Kernel Parameters Optimization Using a Grid Search Method

To specify the best hyperplane with the biggest boundary amongst the two classes, SVMs begin by constructing a nonlinear kernel function to translate information from the input space into a presumably high-dimensional feature space (Chapelle et al., 2002). When using a kernel, defining a set of features that facilitates categorization is simplified. Optimizing the kernel parameters in the kernel functions can boost classification precision. First, the kernel function aids in the creation of

nonlinear corresponding points inside the framework of linear classifier techniques, which is a great benefit. Second, it lets people classify information that doesn't seem to fit neatly into any predetermined vector space. From what I've read (Huang et al., 2014), polynomial kernel functions, linear kernel functions, radial basis/Gaussian kernel functions, and sigmoid kernel functions are the most often used SVM kernel functions.

To confirm that the existing results are reliable and comprehensive for making predictions regarding original data, a k -fold CV is used to randomly split the data set into training and independent testing sets. Individual k subgroups perform as its own holdout testing set for the training model using the remaining $k-1$ subcategories. Tenfold CV was employed for the analysis in this study. CV helps since it lessens the impact of data dependence and boosts the credibility of the results.

In this research, we used a grid exploration to invent the optimal kernel function for the SVM and its optimal parameters. Finding the optimal values for C , the penalty cost, and the other kernel parameters of the SVM may be done quickly and easily using the grid search technique. The original grid search was a comprehensive scan of a small region of the full hyper-parameter space. The hyper-parameters can be set by specifying a minimum value, maximum value, and the desired number of steps. Utilizing a CV technique as a performance measured, grid search optimizes the SVM parameters (C , degree, and coefficient (r)). The objective is to optimize the classifier's hyper-parameters so that it can reliably predict future outcomes.

When using k -fold CV to select C and, the educational data set for students is first partitioned into k subgroups (in this research, k is set to 10). The remaining $k-1$ training subsets are used for analysis, while one is used as testing input. Then, for each value of C , and other parameters, the corresponding CV error for the SVM classifier is computed. A number of hyper-parameter clusters are compared, and the one with the best CV accurateness (smallest CV error) is picked for full-data set SVM training. The linear kernel has a single, crucial optimization parameter called C ; the RBF kernel has two, called C ; and the sigmoid kernel has three, called C , and coefficient (r); and the polynomial kernel has four, called C , degree, and coefficient (r). Figure 3.3 depicts a flowchart for optimizing the parameters of SVM kernels using a grid search. Parameter optimization using SVMs does not rely on hard-and-fast sets of parameter values. When there are more parameters to choose from, the

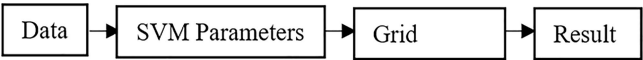


FIGURE 7.1
Framework of the study.

grid search strategy has a better chance of locating the optimal combination parameter. Thus, we employed the ranges $[C \text{ and } \gamma] = 0.0001$ to 10,000, $[\text{degree}] = 1$ to 5, and $[r] = 1$ to 10 in this investigation. The research design is depicted in Figure 7.1.

7.4 RESULTS AND DISCUSSION

The grid search method was used to optimize SVM parameters (C , γ , degree, and coefficient (r)) using CV. Various combinations of hyperparameters rate were entered and the one with the top CV precision (or the lowest CV error) was designated. The range of values used in parameter tuning of each of the four SVM kernel functions (LF, RBF, PF, and SF) considered in this research is shown in Table 4.2. In addition, Table 4.3 revealed the optimal parameter values of each of the four SVM kernel functions after parameter optimization.

The results obtained in the parameter optimization experimentation on the performance of each of the four SVM kernel functions in terms of classification accuracy, sensitivity, specificity, number of support vectors, and computation time are shown in Tables 4.4 and 4.5. As shown in Tables 4.4 and 4.5, the evaluation results of LF, RBF, PF, and SF yielded classification accuracies of 79.31, 86.21, 82.76, and 34.5 percent, respectively. Also, LF, RBF, PF, and SF recorded sensitivities of 71.12, 75.56, 65.56, and 22.20 percent, respectively. Moreover, LF, RBF, PF, and SF produced specificities of 94.44, 96.34, 95.36, and 74.06 percent, respectively. Likewise, the number of support vectors generated for LF, RBF, PF, and SF were 78, 80, 82, and 121, respectively. Furthermore, LF, RBF, and PF utilized the same computation time of 0.03 seconds, while SF executed for 0.06 seconds.

The findings from the results divulged that the RBF kernel was the best in terms of classification accuracy, sensitivity, and specificity. In addition, the RBF kernel has the same computation time as that of LF and PF. Likewise, RBF generated 80 support vectors, which is the closest to that of the LF kernel that produced the least number of support vectors (78).

From the result of parameter optimization, it can be deduced that the RBF kernel outperformed the three other SVM kernel functions and that SF performed least in terms of performance evaluation metrics and the data set used in this study.

Additionally, Tables 7.1, 7.2, 7.3, and 7.4 depict the comparative analysis of each of the four SVM kernel functions in terms of classification accuracy, sensitivity, specificity, number of support vectors, and computation time, respectively.

This study provides a machine learning technique for tuning SVM kernel functions using a data set acquired from the education domain and also focuses on accurate prediction of learners at peril of dropping out. With an 86 percent advantage, RBF-SVM outperforms the three kernel functions.

TABLE 7.1

Range of Values Used in Setting the Parameters for the SVM Kernel Functions

Parameters	Values
Cost (C):	0.00001, 0.0001, $\dots$, 0.1, 1, 10, 100, $\dots$, 10000
Gamma (γ)	0.00001, 0.0001, $\dots$, 0.1, 1, 10, 100, $\dots$, 10000
Coefficient (r):	1, 2, 3, 4, 5, 6, 7, 8, 9, 10
Degree ($^{\circ}$)	1, 2, 3, 4, 5

TABLE 7.2

Optimal Parameter Values of SVM Kernel Functions after Optimization

SVM Kernel Function	Optimal Parameter Values
Linear	Cost C: 10
Radial Basis	Cost C: 100 Gamma (γ): 0.0001
Polynomial	Cost C: 1 Gamma (γ): 0.001 Degree ($^{\circ}$): 1 Coefficient (r): 1
Sigmoid	Cost C: 100 Gamma (γ): 0.0001 Coefficient (r): 1

TABLE 7.3

Computation Time for Tuning SVM Kernel Functions

SVM Kernel Function	Computation Time (Seconds)
Linear	0.03
Radial Basis	0.03
Polynomial	0.03
Sigmoid	0.06

TABLE 7.4

Accuracy, Sensitivity, Specificity, and Amount of Support Vectors of SVM Kernel Functions

SVM Kernel Function	Metrics			
	Accuracy (Percent)	Sensitivity (Percent)	Specificity (Percent)	Number of Support Vectors
Linear	79.31	71.12	94.44	78
Radial Basis	86.21	75.56	96.34	80
Polynomial	82.76	65.56	95.36	82
Sigmoid	34.5	22.20	74.06	121

7.5 CONCLUSION

Education institutions should be able to predict their students’ academic performance. It is a necessity for educators in any domain of study to provide adequate teaching resources which propel a better learning environment. SVM kernel functions fine-tuning employing grid search approach were carried out in this work. The classification accuracies of LF, RBF, PF, and SF were 79.31, 86.21, 82.76, and 34.5 percent, respectively, based on the evaluation results. In addition, the sensitivities of LF, RBF, PF, and SF were 71.12, 75.56, 65.56, and 22.20 percent, respectively. Furthermore, the specificities for LF, RBF, PF, and SF were 94.44, 96.34, 95.36, and 74.06 percent, respectively. Based on the performance measures and data set used in this study, the results demonstrated that the RBF kernel is more efficient than the three other kernel functions. In the future, this research suggests using a hybridized optimization strategy.

REFERENCES

- Afeni, B.O., Oloyede, I.A., & Okurinboye, D. (2019). Students' Performance Prediction Using Classification Algorithms, *Journal of Advances in Mathematics and Computer Science*, 30(2), 1–9. 10.9734/JAMCS/2019/45438
- Akay, M.F. (2009). Support Vector Machines Combined with Feature Selection for Breast Cancer Diagnosis, *Expert Systems with Applications, Elsevier*, 36, 3240–3247.
- Akinola, O.S., Akinkunmi, B.O., & Alo, T.S. (2012). A Data Mining Model for Predicting Computer Proficiency of Computer Science Undergraduate Students, *African Journal of Computing and ICT*, 5(1), 43–52.
- Albreiki, B., Zaki, N., & Alashwal, H. (2021). A Systematic Literature Review of Student' Performance Prediction Using Machine Learning Techniques, *Education Sciences*, 11(9), 552. 10.3390/educsci11090552
- Altujjar, Y., Altamimi, W., Al-Turaiki, I., & Al-Razgan, M. (2016). Predicting Critical Courses Affecting Students Performance: A Case Study, *Procedia Computer Science*, 82, 65–71. 10.1016/j.procs.2016.04.010
- Ayinde, A.Q., Adetunji, A.B., Odeniyi, O.A., & Bello, M. (2013). Performance Evaluation of Naïve Bayes and Decision Stumps Algorithms in Mining Students' Educational Data, *International Journal of Computer Science Issues*, 10(4), 147–151.
- Ayinde, A.Q., Omidiora, E.O., & Adetunji, A.B. (2015). Comparative Analysis of Selected Classifiers in Mining Students', *Educational Data, Communication on Applied Electronics, Foundation of Computer Science, New York, USA*, 1(5), 5–8.
- Barker, K., Trafalis, T., & Rhoads, T. (2004). Learning from Student Data, *In 2004 System and Information Engineering Design Symposium*.
- Ben-David, S. & Lindenbaum, Y. (1997). Learning Distributions by Their Density Levels: A Paradigm for Learning without a Teacher, *Journal of Computer and System Sciences*, 55, 171–182.
- Budiman, F. (2019). SVM-RBF Parameters Testing Optimization Using Cross Validation and Grid Search to Improve Multiclass Classification, *Scientific Visualization*, 11(1), 80–90.
- Burges-Christopher, J.C. (1998). A Tutorial on Support Vector Machines for Pattern Recognition, *Data Mining and Knowledge Discovery*, 2(2), 121–167.
- Chapelle, O., Vapnik, V., Bousquet, O., & Mukherjee, S. (2002). Choosing Multiple Parameters for Support Vector Machines, *Machine Learning*, 46, 131–159.
- Chen, P.W., Wang, J.Y., & Lee, H.M. (2004). Model Selection of SVMs Using GA Approach, *Proceedings of the International Joint Conference on Neural Networks*, Budapest, Hungary, 2035–2040.
- Chowdhury, S., Sing, J.K, Basu, D.K, & Nasipuri, M. (2011). Face Recognition by Generalized Two-Dimensional FLD Method and Multi-Class Support Vector Machines, *Applied Soft Computing*, 11(7), 4282–4292.
- Cortez, P., & Silva, A. (2008). Using Data Mining to Predict Secondary School Student Performance, *In Proceedings of 5th Annual Future Business Technology Conference*, Porto, Portugal, 5–12.
- Das, P., Roychowdhury, A., Das, S., Roychowdhury, S., & Tripathy S. (2020). sigFeature: Novel Significant Feature Selection Method for Classification of Gene Expression Data Using Support Vector Machine and t Statistic, *Frontiers in Genetics*, 11(247), 1–12.
- Dekker, G., Pechenizkiy, M., & Vleeshouwers, J. (2009). Predicting Students Dropout: A Case Study, *In 2nd International Educational Data Mining Conference (EDMOG)*, 41–50.

- Friedrichs, F., & Igel, C. (2005). Evolutionary Tuning of Multiple Sum Parameters, *Neurocomputing*, 64, 107–117.
- Guo, B., Zhang, R., Xu, G., Shi, C., & Yang, L. (2015). Predicting Students Performance in Educational Data Mining, In *2015 International Symposium on Educational Technology (ISET)*, 125–128. 10.1109/ISET.2015.33
- Herzog, S. (2005). Measuring Determinants of Students Return Vs. Dropout/Stopout Vs. Transfer: A First-to-Second Year Analysis of New Freshmen, *Research in Higher Education*, 46, 883–928.
- Herzog, S. (2006). Estimating Student Retention and Degree Completion Time: Decision Trees and Neural Network Vis-à-Vis Regression, *New Directions for Institutional Research*, 131, 17–33.
- Huang, M., Hung, Y., Lee, W. M., Li, R. K., & Jiang, B. (2014). SVM-RFE Based Feature Selection and Taguchi Parameters Optimization for Multiclass SVM Classifier, *The Scientific World Journal*, Hindawi Publishing Corporation, 1–10.
- Huang, S. (2011). A Predictive Modeling and Analysis of Student Academic Performance in an Engineering Dynamics Course, *PhD Thesis*, Department of Engineering and Technology Education, Utah State University.
- Kabakchieva, D. (2012). Student Performance Prediction Using Data Mining Classification Algorithms, *International Journal of Computer Science and Management Research*, 1(4), 686–690.
- Kim, K.J. (2003). Financial Time Series Forecasting Using Support Vector Machines, *Neuro-Computing*. 55(1/2), 307–319.
- Kotsiantis, S.B., Pierrakeas, C.J., & Pintelas, P.E. (2003). Preventing Student Dropout in Distance Learning Using Machine Learning Technique, Paper Presented at the *AI Techniques in Web-Based Educational System at Seventh International Conference on Knowledge-Based Intelligent Information Engineering Systems*, Oxford, England.
- Kotsiantis, S.B., Pierrakeas, C.J., & Pintelas, P.E. (2004). Predicting Students' Performance in Distance Learning Techniques, *Applied Artificial Intelligence*.
- Kovacic, Z.J. (2010). Early Prediction of Student Success Mining Student Enrolment Data, In *Proceedings of Informing Science and IT Education Conference (INSITE) 2010*.
- Kuyoro, S.O., Nicolae, G., Oludele, A., & Samuel, O. (2013). Optimal Algorithm for Predicting Students Academic Performance. *International Journal of Computer and Technology*, 4(1), 63–75.
- Lau, E.T., Sun, L., & Yang, Q. (2019). Modelling, Prediction and Classification of Student Academic Performance Using Artificial Neural Networks. *SN Applied Sciences*, 1(9), 982. 10.1007/s42452-019-0884-7
- Luts, J., Ojeda, F., Van de Plos R., De Moor, B., & Van Huffel, S. (2010). A Tutorial on Support Vector Machine-Based Methods for Classification Problems in Chemometrics, *Anal Chim Acta*, 665, 129–145.
- Lykourantzou, I., GiannouKos, I., Nikolopoulos, V., Mpardis, G., & Loumos, V. (2009). Dropout Prediction in E-Learning Course through the Combination of Machine Learning Techniques, *Computers and Education*, 53, 950–965.
- Minaei-Bidgoli, B. (2003). Predicting Student Performance: An Application of Data Mining Methods with an Educational Web-Based System, In *33rd ASEE/IEEE Frontiers in Education Conference*, 1–6.
- Oeda, S., & Hashimoto, G. (2017). Log-Data Clustering Analysis for Dropout Prediction in Beginner Programming Classes, *Proceedings of International Conference on Knowledge Based and Intelligent Information and Engineering Systems, Procedia Computer Science*, Elsevier, Marseille, France, 112, 614–621.

- Oludare, O., Olabiyisi, S., Olaniyan, A., & Fagbola, T. (2014). An Optimized Feature Selection Technique for Email Classification, *International Journal of Scientific and Technology Research*, 3(10), 286–293.
- Omidiora, E.O., Fenwa, O.D., Adeyanju, F.A., Babalola, O.V., Opeyemi, E.D. & Jimoh, H.O. (2014). Performance Evaluation of Support Vector Machines and Multilayer Perception for Recognition of Handwritten Digits, *Proceedings of Third International Conference on Engineering and Technology Research, Ladoke Akintola University of Technology, Ogbomosho*, 3, 141–148.
- Ovidiu, I. (2007). Application of Support Vector Machines in Chemistry, *In Reviews in Computational Chemistry*, 23, 291–400.
- Oyedele, O.J., Oladipupo, O.O., & Obagbuwa, I.C. (2010). Application of K-Means Clustering Algorithm for Prediction of Students’ Application Performance, *International Journal of Computer Science and Information Security*, 71(1), 292–295.
- Pathom, P., Anongnart, S., & Prasong, P. (2008). Comparisons of Classifiers Algorithms: Bayesian Network, C4.5, Decision Forest and NBTree for Course Registration Planning Model of Undergraduate Students’ Scripatum University Chonburi Campus, Chonburi, Thailand, *IEEE*.
- Ramaswami M., & Bhaskaran, R. (2010). CHAID-Based Performance Prediction Model in Education Data Mining, *International Journal of Computer Science Issue*, 7(1), 10–18.
- Romero, C., Ventura S., Hervás, C., & Gonzales, P. (2008). Data Mining Algorithms to Classify Students, *In Proceeding of International Conference on Education Data Mining Canada*.
- Romon, M.M., & Christodoulou, C. (2006). *Support Vector Machines for Antenna Array Processing and Electromagnetics*, San Rafael, CA: Morgan and Claypool.
- Sampon, D.L., Parker, T.J., Upton, Z., & Hurst, C.P. (2011). A Comparison of Methods for Classifying Clinical Samples Based on Proteomics Data: A Case Study for Statistical and Machine Learning Approaches, *PLoS One*, 6(9): e24073. 10.1371/journal.Pone.0024973.
- Sujitparapitaya, S. (2006). Considering Student Mobility in Retention Outcomes, *New Directions for Institution Research*, 131, 35–51.
- Teukolsky, H.W., Vetterling, A.S., & Flannery, T.W. (2007). *Section 16.5 Support Vector Machines: Numerical Recipe: The Art of Scientific Computing*, 3rd ed., New York: Cambridge University.
- Vandamme, J.P., Meskens, N., & Superby, J.F. (2007). Predicting Academic Performance by Data Mining Methods, *Education Economics*, 15(4), 405–419.
- Vapnik V.N. (1995). *The Nature of Statistical Learning Theory*, New York: Springer.
- Vapnik, V.N. (1998). *Statistical Learning Theory*, New York, USA: John Wiley & Sons.
- Zaffar, M., Ahmed Hashmani, M., Habib, R., Quraishi, K., Irfan, M., Alqhtani, S., & Hamdi, M. (2022). A Hybrid Feature Selection Framework for Predicting Students Performance. *Computers, Materials & Continua*, 70(1), 1893–1920. 10.32604/cmc.2022.018295
- Zhang, Y., Qussena, S., Clark, T., & Kim, H. (2010). Use Data Mining to Improve Student Retention in Higher Education: A Case Study, *In 12th International Conference on Enterprise Information Systems (ICEIS)*.