Machine Learning for Data Mining, Data Science and Data Analytics

It has been a huge challenge successfully completing the special issue titled “Machine Learning for Data Mining, Data Science, and Data Analytics” under the reputed journal Recent Advances in Computer Science and Communications, formerly Recent Patents on Computer Science.

Most research contributions related to machine learning are primarily devoted to improving the prediction accuracy and performance of the learning model. However, it is a fact that less attention is paid towards mining health care data, hybrid machine learning models, detecting low-frequency attacks [1-3], and addressing optimization. To address these issues, the focus of this special issue in Recent Advances in Computer Science and Communications is on machine learning for data mining, data science, and data analytics.

This special issue was planned to receive quality submissions from different parts of the country and the world. Nearly 12 research contributions were received with great enthusiasm from the research and scientific community. Out of these, 4 articles were accepted for inclusion in the special issue after a stringent and thorough review process, at least by two independent referees.

The special issue includes four contributions, which were selected after reviews were conducted by experts working in these areas. Here, we present a very brief summary of the selected papers for this special issue and concluding remarks.

SUMMARY

The special issue contains 4 papers, briefly discussed as follows:

The first article, “Mathura (MBI) - A Novel Imputation Measure for Imputation of Missing Values in Medical Datasets,” is written by B. Mathura Bai, N. Mangathayaru, B. Padmaja Rani, and Shadi Aljawarneh [4], proposes an imputation measure for filling missing data values to make the incomplete medical datasets as complete datasets. Then the approach is to apply this imputation measure on imputed datasets to achieve improved classifier accuracies. The basic intention of the research study is to present an imputation measure to find the proximity between medical records and an approach for imputation of missing values in medical datasets to improve the accuracy of existing classifiers. The performance of the proposed approach is compared to existing approaches with respect to classifier accuracy and also by performing a non-parametric test called the Wilcoxon test.

The second article, “E-mail Fraud Attack Detection Using Hybrid Machine Learning Approach,” is written by Yousef A. Yaseen, Malik Qasaimeh, Raad S. Al-Qassas, and Mustafa Al-Fayoumi [5]. E-mail is an efficient way to communicate. It is one of the most commonly used communication methods, and it can be used for achieving legitimate and illegitimate activities. Many features that can be effective in detecting e-mail fraud attacks are still under investigation. This paper proposes an improved classification accuracy for fraudulent e-mails implemented through feature extraction and hybrid Machine Learning (ML) classifier that combines Adaboost and Majority Voting. Eleven machine learning classifiers are evaluated experimentally within the hybrid classifier. The performance of the e-mail fraud filtering is evaluated by using WEKA and R tool on a data set of 9298 e-mail messages. The utilized proposed e-mail features with the combination of Adaboost and Voting algorithms prove the efficiency of fraud e-mail detection.

Intrusion detection systems play a key role in system security by identifying potential attacks and giving appropriate responses. As new attacks are always emerging, intrusion detection systems must adapt to these attacks. More work is continuously needed to develop and propose new methods and techniques to improve efficient and effective adaptive intrusion systems. Feature selection [1, 2, 6, 7] is one of the challenging areas that need more work because of its importance and impact on the performance of intrusion detection systems. The third article, “Intrusion Detection System for Malicious Traffic Using Evolutionary Search Algorithm” is written by Samar Al-Saqqaa, Mustafa Al-Fayoumi, Malik Qasaimeh, and Raad S. Al-Qassasb [8]. It applies an evolutionary search algorithm in feature subset selection for intrusion detection systems. The evolutionary search algorithm is applied to find the best subset of features for the intrusion detection system. The experimental study proved their approach promising to be used as a feature selection method for intrusion detection. The results showed better performance for the intrusion detection system in terms of accuracy and detection rate.

The app stores, for example, Google Play and Apple Play Store, provide a platform that allows users to provide feedback on the apps in the form of reviews. An app review typically includes a star rating followed by a comment. Recent studies have shown that these reviews possess a vital source of information that can be used by the app developers and vendors for improving the future versions of an app. However, in most cases, these reviews are present in unstructured form, and extracting useful information from them requires a great effort. The fourth article, “An Optimized Classification of Apps Reviews for Improving Requirement Engineering” is written by MPS Bhatia, Akshi Kumar, and Rohit Beniwal [9]. It provides an optimized classification approach that automatically classifies the reviews into a bug report, feature request, shortcoming, and improvement request relevant to requirement engineering. The optimized automatic classification improves the requirement engineering where the developer straightforwardly knows what to improve further in the concerned app.
CONCLUSION

The lead guest editor, Vangipuram Radhakrishna, and other guest editors, Gunupudi Gali Suresh Reddy, Rajesh Kumar, and Dammavalam Srinivasa Rao, of the special issue, are heartfully thankful to the Editor-in-Chief of the journal, Francesco Bene- detto, and also to the previous Editor-in-Chief of the journal, Professor Hamid Mcheick for their valuable support. Our special thanks to the whole editorial staff, especially Wajeeha Syed, Ahmed, and Raheela Anjum, for their continuous support in the preparation and publication of this special issue. Finally, the lead guest editor and guest editors of the special issue heartfully express their gratitude to the authors for their quality contributions and the reviewers for providing timely reviews and their dedication.

REFERENCES


Vangipuram Radhakrishna
(Lead Guest Editor)
Department of Information Technology,
VNR Vignana Jyothi Institute of Engineering and Technology, Telangana, India
E-mail: vrkrishna@acm.org; radhakrishna_v@vnrvjiet.in

Gunupudi Rajesh Kumar
(Guest Editor)
Department of Information Technology,
VNR Vignana Jyothi Institute of Engineering and Technology, Telangana, India
E-mail: rajeshkumar_g@vnrvjiet.in

Gali Suresh Reddy
(Guest Editor)
Department of Information Technology,
VNR Vignana Jyothi Institute of Engineering and Technology, Telangana, India
E-mail: sureshreddy_g@vnrvjiet.in

Dammavalam Srinivasa Rao
(Guest Editor)
Department of Information Technology,
VNR Vignana Jyothi Institute of Engineering and Technology, Telangana, India
E-mail: srinivasarao_d@vnrvjiet.in