World Scientific
www.worldscientific.com

Editorial

Special Issue on New Frontiers in Data Sciences and Data Analytics Tools and Applications

1. Introduction

The Special Issue on New Frontiers in Data Sciences and Data Analytics Tools and Applications comprises very selectively chosen papers that were rigorously peer-reviewed by at least three reviewers in the field of data sciences and data analytics tools and applications.

Submissions were received from researchers, engineers, and industry professionals. Each paper presents an improved approach in e-commerce, text classification, the internet of things (IoT), data transmission of IoT devices, deep learning, data mining, and intelligent decision support systems.

2. Overview of Submissions

Recommender Systems (RS) have become a fundamental piece of most internet business locales these days. Even though there are a few investigations directed on recommender frameworks, a mixture recommender framework for the genuine state internet searcher to discover fitting rental condo bringing clients' preference into account is still due. To address this issue, a half and half recommender framework is proposed by Neloy *et al.* The authors built two of the most well-known suggestions: Collaborative Filtering (CF) and Content-Based Recommender (CBR). CF-based strategies utilize the evaluations given to things by clients as the sole source of data for figuring out how to propose.

Currently, these evaluations are regularly extremely scanty in applications like search engines, causing CF-based strategies to let down accuracy and performance. To lower this sparsity issue in the CF strategy, the Cosine Similarity Score (CSS) between the client and predicted apartment, in light of their Feature Vectors (FV) from the CBR module is used. The proposed recommender was assessed utilizing the Statistical Cross-Validation comprising of Leave-One-Out Validation (LOOCV). Test results show that it essentially performs a benchmark arbitrary recommender as far as accuracy and review. Moreover, a graphical analysis of the connections between the precision and error minimization is introduced to give additional proof to the possibility of this hybrid recommender framework.

Kandhro *et al.* created roman Urdu feature news text characterization. The followed process is stepwise, in the initial step, it tokenizes the roman information, applies the preprocessing procedures including stop words expulsion and stemming utilizing various calculations, on the tokenized information.

The exploratory development of 10 distinctive news classes is demonstrating great correctness by utilizing three distinctive deep learning strategies. In this work, the deep learning models RNN, LSTM, and CNN are utilized for characterization Roman-Urdu news.

The testing exactness of models are RNN (80%), LSTM (82%), and CNN (79%). Results show that the LSTM strategy is nearly well on testing with certain ages and group size, the accuracy values are contrasted with RNN, LSTM, and CNN strategies with a few boundaries. In expansion to this, the presentation of the roman classifier has bargained someplace because of the inconsistencies in spellings. It can likewise be composed as khoobsurat, khubsoorat, and khobsorat. However, a few improvements are important in certain regions, for example, a solid stop words rundown, and more precise stemming classifier calculations utilized for mining of roman information.

To know the last known status of variables that are under monitoring, a real-time integrated approach can be used. Accordingly, the Internet-of-Thing (IoT) devices have given choices to address distributed information and data collection. In any case, the self-governance of IoT devices presents one of the fundamental difficulties to actualize the collection procedure. Battery autonomy is influenced straightforwardly by the vitality utilization from information transmissions.

The Data Stream Processing Strategy (DSPS) is a technique to implement project measurements from IoT devices that are linked with a part named Measurement Adapter (MA). Diván *et al.* presented a new information support association dependent on estimation metadata explained with online information filtering to advance the information transmissions from MA. Moreover, a weighted information change location approach is joined, while another local support dependent on coherent windows is proposed for MA. Likewise, an articulation among the information buffer, a temporal barrier, and data change detectors is presented. The information buffer consumed 568 kB for checking 100 concurrent measurements. The online assessment of the mean and change dependent on the statistical cycle control consumed 238 ns.

During the previous years, deep learning structures are being utilized in numerous potential territories, for example, object identification, natural language processing, face recognition, clinical investigation, and other related applications. In these applications, deep learning has accomplished noteworthy results in coordinating the performance of human specialists. Kumar *et al.* discovered a novel convolutional neural systems-based methodology for the location of breast malignant growth in obtrusive ductal carcinoma tissue areas utilizing entire slide pictures. It has been seen that breast malignant growth has been a driving reason for death among ladies. It likewise stays an endeavoring

task for pathologists to discover the harm districts from whole slide images. In this study, authors have actualized diverse convolutional neural systems (CNN) models that incorporate VGG16, VGG19, Xception, Inception V3, MobileNetV2, ResNet50, and DenseNet. The analyses were performed on standard whole slide images informational collection which incorporates 163 patients of invasive ductal carcinoma.

For experiment assessment, the same informational index was partitioned into 113 and 49 pictures for preparing and testing individually. The testing was completed independently over each model and the obtained results demonstrated that the proposed CNN model accomplished 83% precision which is superior to other models.

Numerous regressions are not dependable to recuperate indicator inclines inside homogeneous subgroups from heterogeneous examples. As opposed to Monte Carlo's investigation, which allots totally to the principal determined indicator the variety it imparts to remaining indicators, various relapse does not allocate to any indicator the common variety between or across indicators, which is sequestered in the residual term. This unassigned variety may correspond with determining indicators and leads to heteroscedasticity.

Francoeur developed and tested an iterative, successive algorithm to estimate a two-section arrangement of weighted least-squares relapses for recuperating Monte Carlo indicator inclines in three homogeneous subgroups of a heterogeneous example ($n = 1500$). Every indicator has an alternate nontypical appropriation. Inside every subgroup, the calculation changes inclination from heteroscedasticity identified with one, a few, or every predetermined indicator and from "unimportant" multicollinearity. It recovers each of the three indicated indicator inclines over the three subgroups in two situations, one of which is additionally impacted by two unspecified indicators.

The algorithm stretches out a versatile examination to find and assess designs in field exploration and machine learning when indicators are between related, even unspecified so that fair-minded result groups in homogeneous and heterogeneous examples can be uncovered.

Asadullah Shaikh
Department of Information Systems
College of Computer Science and Information System
Najran University, Najran 61441, Saudi Arabia
asshaikh@nu.edu.sa

Khairan D. Rajab
Department of Computer Science
College of Computer Science and Information System
Najran University, Najran 61441, Saudi Arabia
khairanr@gmail.com