

Mining Smart Learning Analytics Data Using Ensemble Classifiers

<https://doi.org/10.3991/ijet.v15i12.13455>

Samina Kausar

Shanghai University, Shanghai, China
University of Kotli Azad Jammu and Kashmir, Kotli Azad Kashmir, Pakistan

Solomon Sunday Oyelere

University of Eastern Finland, Kuopio, Finland

Yass Khudheir Salal

South Ural State University, Chelyabinsk, Russia

Sadiq Hussain ^(✉),

Dibrugarh University, Dibrugarh, India
sadiq@dibru.ac.in

Mehmet Akif Cifci

Istanbul Aydin University, Istanbul, Turkey

Slavoljub Hilcenko

College of Vocational Studies for Teacher and Trainers Education – Subotica,
Vojvodina, Serbia

Muhammad Shahid Iqbal

Anhui University, Hefei, China

Zhu Wenhao, Xu Huahu

Shanghai University, Shanghai, China

Abstract—Recent progress in technology has altered the learning behaviors of students; besides giving a new impulse that reshapes the education itself. It can easily be said that the improvements in technologies empower students to learn more efficiently, effectively, and contentedly. Smart Learning (SL), despite not being a new concept describing learning methods in the digital age, has caught the attention of researchers. Smart Learning Analytics (SLA) provides students of all ages with research-proven frameworks, helping students to benefit from all kinds of resources and intelligent tools. It aims to stimulate students to have a deep comprehension of the context and leads to higher levels of achievement. The transformation of education to smart learning will be realized by reengineering the fundamental structures and operations of conventional educational systems. Accordingly, students can learn the proper information yet to support to learn real-world context, more and more factors are needed to be taken into account. Learning has shifted from web-based dumb materials to con-

text-aware smart ubiquitous learning. In the study, SLA dataset was explored, and advanced ensemble techniques were applied for the classification task. Bagging Tree and Stacking Classifiers have outperformed other classical techniques with an accuracy of 79% and 78% respectively.

Keywords—Learning Analytics, Ensemble Classifiers, Educational Data Mining, Bagging, Boosting

1 Introduction

Learners are fostered with smart learning and intelligent design to impart education thanks to the progress in computer technologies. Learners using smart devices access the educational resources and plunge into seamless and ubiquitous learning. Learners can access the digital content wirelessly anywhere and anytime without time and location constraints. This 21st Century Education system, referred to as smart education, is also supported by new smart technologies such as cloud computing, IoT, and wearable technology. Wearable technology can infuse the learning materials for the learners, while IoT provides everything with the internet. Smart learning does not only focus on the utilization of devices, but it is a service-oriented, context-aware, and learner-centric educational paradigm. Smart learning features include content focus, personalized, technology-embedded learning [1, 2]. In a smart learning environment, learners can learn at their own pace and time. The recommender systems deliver content based on the learner's learning pattern and behavior. So, this learning environment is effective, engaging, and efficient [3]. Smart learning environment delivers guidance, support, and suggestions to the learners at the right time, in the right place, and form. This context-aware environment enables the system to adapt the subject contents and course contents. Kim et al. [4] devised such an environment based on cloud computing to push learning contents to the smart devices of the learners based on the learners' personal data. Thus, learning becomes personalized and customized. Huang et al. [5] considered the smart learning environment as the digital environment that records the learner's characteristics, learning process and outcome, using adaptive resources and interactive tools. Zhu et al. [6] summarized the smart learning environment as a personalized, learner-centric, adaptive, context-aware, and ubiquitous environment that supports high-end digital technology; moreover, a four-tier smart pedagogy framework to enrich learners with knowledge and skills was designed. This framework included class-based differentiated instruction, individual-based personalized learning, group-based collaborative learning, and mass-based generative learning. Nowadays, smart devices are smarter, smaller, and more affordable. These smart devices are used seamlessly to transfer data and location-based services. Smart computing utilizes the three-tier architecture comprising of fog, cloud, and swarm computing. In this architectural design, cloud computing supports auditability, action, and swarm computing facilitates awareness, while fog computing deals with analysis and alternatives. Awareness, analysis, actions, alternatives, and auditability are the key functions of smart computing. Therefore, smart learning is a new learning envi-

ronment in global education. It is based on contextual, seamless, personalized, and contextual learning to improve learner's knowledge and intelligence. "A smart learning environment not only enables learners to access digital resources and interact with learning systems in any place and at any time but also actively provides the necessary learning guidance, hints, supportive tools or learning suggestions to them in the right place, at the right time and in the right form" [7]. In recent years there were many educational projects based on smart learning. For instance, in Malaysia, smart schools education projects are implemented to improve the quality of education [8]. Moreover, Singapore has implemented the Intelligent Nation Master Plan, where learning is an integral part [9]. IBM is helping Australia to design a smart education system, which is student-centric in nature [10]. South Korea, the USA, and the UAE are the other such nations to implement such smart learning projects. Though having many chances, Smart learning comes with many challenges. Pedagogical theory, educational ideology, teacher's learning leadership, and educational structures are key challenges to mention in this domain [5].

Accordingly, authors [11] suggest a new computational research environment using mobile devices. In order to simplify the access of supercomputers, Science AppStore, task management, and virtualization technologies are developed on mobile devices. The authors [12] introduce a new paradigm called "Smart Learning" in South Korea. The adoption of highly advanced Information Technology (IT) in education has been a global issue for many years.

On the other end of the educational improvement, the spectrum is Learning Analytics (LA), which focused on the measurement, gathering, analyzing, and presenting learning data for the sole aim of improving learning outcomes [13]. An important aspect of learning analytics is the need to apply the knowledge obtained from the data towards improving learning, creating optimal and sustainable educational achievement. Several prerequisites of a smart learning environment, such as understanding the participation of learners, assessment measures, performance, collaboration, motivation, attitudes, satisfaction [14], and other indices of learning, can be determined through the support of learning analytics. The focus of research in learning analytics has already advanced beyond the customary analysis of numerical data about learners' footprints such as course grades, access to learning materials, amount of time spent on exploring learning contents and other data points. The focus nowadays is delving into the multi-faceted, intelligence-driven, ubiquitous digital environment characterized by a huge amount of data from multiple contexts, devices, and processes. Besides, the integration of the learning analytics process as a component of smart learning has gained useful momentum lately, an indicator of its relevance to improving educational goals. It is firmly believed that this integration will facilitate gaining insight into learning data within a smart learning environment. Educational experts, including teachers, are gradually using the large pool of data produced within smart learning environments, which give rise to the concept of SLA. Kinshuk and Kumar [15] emphasized that "educators can use SLA to discover and analyze data on students' behaviors, instruction merit and the suitability of learning environments, gathering information from various sources to discern learning traces that can facilitate instructional support." Therefore, SLA processes observe and trail the diverse digital foot-

prints associated with the learning environment, infer, and organize the data, use the data to support learning, and make decisions regarding the future. The aim of SLA is not only to gain knowledge and insight into the progress of students and teachers but also to evaluate the overall learning experience and justify the effectiveness of any learning intervention in real-time.

The goal of this study is to review broadly and systematically the trends of SLA from the perspectives of different types of analytics and the several decision-making support systems in which learning analytics intervention is designed. This review of existing work is intended not only to unravel the direction of the pedagogical application of SLA but also to how relevant the solutions are provided by the analytics techniques. Furthermore, a study like this helps to prepare the educational stakeholders, including teachers, students, and administrators on how best the future educational data mining terrain should be designed to adequately meet the needs and expectations of stakeholders to achieve the desired result in support of decision-making in the education context.

An SLA dataset created in the Anhui University, Hefei, China, and was analyzed to perform classification. Different classification and ensemble methods were implemented to achieve better accuracy. The techniques applied were Decision Tree, K-NN, Bagging-KNN, Bagging Tree, Random Forest, Naïve Bayes, and Stacking Classifiers. Stacking classifier achieved an accuracy of 78% while Bagging Tree outperformed it with 79% accuracy. The average error recorded in the training data was 0.100. The base estimators beyond 15 did not achieve any accuracy gain in Bagging Tree ensemble for Student Performance.

The rest of the paper is organized as Section II presents Literature Review, while Section III depicts case studies of smart learning analytical tools, Section IV elaborates experiments and results, and Section V presents Discussion and Conclusion.

2 Literature Review

In this section, some research topics related to smart learning analytics and different dashboards related to it are studied. SLA involves context-awareness, i.e., the context-awareness means anything affecting learners' performance. The intelligent learning aims to give personalized learning for this, a study [16] develops a model consisting of four services such as pull, prospect, content, push. With pull service, learners acquire what content should be, and the prospect service is responsible for the preparation of the content service, which generates the content. Finally, the push service is responsible for the synchronization of delivery content to the target.

A conceptual educational dashboard design was proposed to evaluate the existing frameworks[17]. This framework tried to figure out whether the dashboards were fit in both pedagogically and visually for the stakeholders. The conceptual dashboard was based on a four-level evaluation model by Kirkpatrick and Few's models of educational dashboard design. Educational dashboards support the teachers and students by supplying personalized and auto-generated requirements and feedbacks in real-time. About ten case studies dealing with the design of such a tool were taken into

account. Educational data mining is an emerging area, and the institutes can gain insight into various hidden patterns and information out of it. The raw log files of the learning management system may be utilized to check the student drop out and excel in academic achievements. Educational dashboards can support the physically separated students and teachers to find out at-risk students, good performers, self-learning as well as act as a recommendation system based on the data generated. The educational dashboards they reviewed were GLASS, Gismo, SNAPP, SAM, Step-up, Narcissus, Course Signal, Students Success System, Student Inspector, and Loco analyst. They compared the case studies with their limitations. They evaluated the dashboards' designs and suggested effective ways to deliver information to the students visually at the right time.

In [18], it is studied that advanced information technologies provide an educational environment to broaden the place, content, and forms of education. Digital textbooks and smart devices have already taken paper books places, and the potential forms of education are becoming infinite. Likewise, the SNS has become an educational instrument.

In [19], it was figured out the features responsible for the smart implementations of learning analytics. In 2006, Retails et al. proposed the idea of learning analytics, and Long et al. predicted the future of higher education with big data and its deep analysis. The seven features they highlighted were learning awareness, privacy awareness, time awareness, visual feedback, pedagogical interventions, big data centralism, and knowledge structures acquisition. SLA should stimulate the learners about their learning process and behavior based on the data. In this environment, learners' privacy should not be compromised. The instructors and learners should be able to visualize their progress with time. Such illustrations should be simple and easy to understand. Such learning analytics should be clubbed with pedagogical interventions. Nowadays, students can learn anywhere, any time, and using any device, yet the data should be aggregated at a central place for further analyses. The LA software should be able to investigate the learning behavior, and pattern of the learners, and the policymakers should dig out the hidden knowledge from the big data repositories. The need of the learners should be addressed in real-time by such software.

Another study highlighted the revolution that occurred in the field of education with the help of technological advancement [20]. It is a paradigm shift from formal learning to more focus on informal learning. They suggested solutions to some of the current challenges of learning environments. They recommended a transformation from the present learning environment to a smart learning scenario. The fundamental operations and structures of the present education system need to be re-engineered with the pedagogical shift. They had reviewed the new innovations related to smart learning like game-based learning, portfolio maintenance, flipped classroom, gesture-based learning, lifelong learning, team teaching, peer learning, and competency assessment. Their futuristic vision document of the learning environment would help both the educators and learners. In these new learning environments, working and learning can be merged together, and learners can learn at their own pace and time. The learners can control their learning process, submit their assignments and assessment at their preferred time, and choose their own instructors.

The key issues and characteristics in designing smart learning environments based on personalization were discussed in [1, 2, 21]. Novel pedagogical approaches need to be orchestrated to integrate smart educational environments into the learning ecosystem. The author emphasized the need for future users of the smart learning systems to take active participation in designing such an ecosystem. Individual learners must be supported to create an adaptive learning environment. These environments searched for meaningful learning patterns by analyzing big data created by users. The system should be able to identify the learners' location and historical data to integrate real-time information. A new approach 'Education as a Service' is also coming into view to dealing with the challenges of smart learning. So, there would be remarkable changes in traditional education. A third party may conduct assessment tests instead of the tutor indulged in teaching in such a scenario. Therefore, there would be no distinguishing line between formal and informal learning in the future. That is why the developers of smart learning tools must take all the aspects of this new learning paradigm into account.

Models for implementation and planning of smart learning environments were proposed by R. Nikolov et al. [22]. Technology-enhanced learning had changed the learning behavior and teaching methodologies of an educational institute. New computing innovations such as Wearables, Internet of Things, Machine learning, Advanced Analytics had come up. The new generation learners should acquire these innovative skills to fulfill the requirements of the corporations and industries. So, the curriculum needs to be updated time to time basis to meet the requirements of these corporations and industries. Autonomous decision making, big data with learning analytics, stacking, and replacing the LMS and full context awareness are the drivers that make the SLA separable from the other learning models. Technology, psychology, and social constructivism are the main areas of deployment and development of SLA. They had demonstrated a case study for monitoring environmental parameters in a smart city by the students based on the Internet of Things. The project was powered by Forging Online Education through FIRE (FORGE) project under which the remote labs were built and delivered to the students for such kind of activities.

A contextualization framework called CoALA was designed to improve the effectiveness, performance, and scalability of learning analytics [23]. The important learning analytics requirements may be categorized as personalization, data collection, data integration, student profiling, insight extraction and projection, and decision making. Educational Institutes may generate dynamically learning materials for each student by utilizing the recommendation systems. In their IoT based model, three types of analytics viz., descriptive, predictive, and prescriptive analytics is done. They utilized the OODA loop of the IoT applications for their framework. OODA involved observation (data collection), Orientation (integration and processing of data), Decision (decision making), and Action (action based on that decision). They described the effectiveness of their model in collaborating with the learning environment in one scenario.

The key areas and features taken into consideration while developing smart learning tools were discussed in [1, 2, 24]. One of the important aspects of a smart city is technology-enhanced learning and SLA. The main characteristics of a smart city are

smart cities, smart energy, smart technology, smart infrastructure, smart mobility, smart building, smart healthcare, smart governance, and education. They suggested effective measures to access learning outcomes properly. The relationship between learning analytics, learning outcomes, and smart learning environments were also demonstrated. They figured out openness, benchmarks identification, qualitative analysis, course map designing, and mapping LMS activities with learning outcomes as the major features for measuring learning outcomes. To extract intelligent data from the LMS system is a challenge for the data miners. They suggested both qualitative and quantitative analytics to bring in-depth research and interpretations in this area.

The state-of-the-art learning analytics tools and techniques were summarised by Peter Mikulecky [25] and presented a vision for future endeavors. He also analyzed the successful projects in this area and jolt down the key features. He reviewed learning analytics from multi-agent to cloud-based architecture perspective. The ubiquitous and context-aware learning is the foundation of a smart learning environment. In this environment, the right information must be delivered to the right person right time in the right way. The challenges and future directions in this area were also addressed.

An Internet of Everything supported Educational Model using deep learning system was proposed by [26]. This model could be used to access and evaluate learners' learning process, outcome, and achievements and suggest corrective measures for the betterment of the learners. They discussed the challenges and advantages of such a model and how to analyze the data generated from the Internet of Everything devices in educational domain perspectives. They compared the performance parameters such as attainment, retention, attention, and cognition of learners of their learning analytics model with the traditional approach of teaching-learning. Some of the advantages of such an ecosystem based on educational paradigms are student and staff attendance management, smart learning and teaching, automated library management, reduced power consumption, security, and authentication. The challenges include the high velocity of data, mining relevant and valuable patterns, security, heterogeneous data types, data storage, the conventional mindset of stakeholders, varied data formats, and cost of the infrastructure.

The characteristics of the smart learning environment and smart education and the pedagogical approaches were elaborated by Elena Shoikova et al. [27] for the same. The learners can learn more effectively, flexibly, efficiently, and comfortably in the digital era of education. The fusion of pedagogy and technology make the learners achieve their learning objectives. In the context-aware approach, learners can learn seamlessly and have personalized settings. They suggested the opinion of future users in the design and implementation of smart learning tools and software. The design of such tools orchestrates the content, practice, pedagogical theory, and use of technological resources in a proper manner.

3 Case Studies of Smart Learning Analytics Tools

3.1 Case study: MENSA Serbia

One of the possible examples for the learning concepts as for per models of “Nikola Tesla Center” (NTC). “NTC”-learning system implies learning based on the theoretical basis of neurology, neuropsychiatry, and other sciences, especially pedagogical – family pedagogy, didactics, and methods for pre-school and elementary school age. That learning system represents the operationalization of theoretical knowledge of the mentioned scientific disciplines, which is a good indicator of how to connect theory with educational practice. Thereby, it is about a new approach to learning, which dominates the thoughtful activity of the child and its successful development. The system is very well and thoroughly elaborated, applicable in the family, preschool institutions, and in the elementary school” [28].

Colossal data is inevitable to be seen publicly. In only 2 seconds, any information will appear on the screen of our computers or mobile phones. It can be seen the year of any great battle or the Latin name of a plant... So, getting the data –is no longer a challenge. But how to use this data, apply it, connect two or more and conclude the third, is the method of the "NTC" learning system, which can be adopted by any educator, teacher, or parent in order to exercise children. No specific knowledge is required for that! It can be the first preparation for life and work challenges to solve problems in a creative and successful way with what is currently had as disposal!

By applying program Mensa – “NTC” – learning system:

The level of intellectual abilities of all children participating in the program is increasing; Concentration and attention disorders are prevented, later in the school period (dyslexia); Movement coordination and mobility are developing; The speed of thinking and conclusion are developing (functional thinking); All children benefit from the program, and it is especially useful for detecting gifted children and encouraging the development of giftedness; The number of neuronal connections increases, which increases the brain's capacity for information processing with specific exercises.

The methodology of work within – “NTC”-learning system has a wide range of impacts on total psycho-physical development of children, from which, the standardized test-questions were as pointed out, which aim to encourage (deep, analytical thinking) logically-functional knowledge like:

Connecting abstract symbols – I degree and

Connecting abstract symbols and logic – II degree.

The essential difference between the enigmatic logical-functional question and the common question in education is that the usual school question will be answered only by one who has learned, memorized, heard by whispering, or acquired the knowledge required to answer in some other way. Such a child will respond immediately. Children who have not heard, learned, or get the information needed to answer in some other way will not respond, will not think, will not give their idea, will not play, will not develop their neural network at that moment. By applying enigmatic questions in work, everything of that will exist, and to a great extent! By applying enigmatic ques-

tions in work, everything of that will exist, and to a great extent! Those who will be trained to make these kinds of questions will be able, whenever they wish, to revive the atmosphere in leisure and to socialize (at home), at activities (in kindergarten), in-class (in school) and at least for a few minutes, they create a dynamic, interesting process in which all children are thinking and trying to apply everything they know to get the correct answer! [29].

It has been proven that pre-school and low-school age is the most important and most effective period for early stimulation of synapses, directly responsible for the enormous intellectual potential of children. This model of work should be practiced in conjunction with motor exercises, which further stimulate these predispositions [30].

A part of short instructions of MENSA Serbia, which make the assembling of enigmatic logical-functional questions much easier is provided:

Rule 1: Good enigmatic question should have only one correct and meaningful answer! When the enigmatic question is made, first, it needs to be carefully rethought. In case there are two or more correct and meaningful answers, it is necessary to adjust the question so that there is only one possible answer.

Rule 2: A question, which will inevitably lead to a mere enumeration, is avoided. Even more undesirable question is when there is a short-range of possible answers, so it inevitable to correctly answer on it with short enumeration.

Rule 3: Remember! A question on which the answer will be given only by the one who has learned is not desired, memorized, heard by whispering, or acquired the knowledge required to answer in some other way! Therefore, questions about the exact data are out of question, such as names, years, or professional expressions. They were not required in answers either.

Rule 4: Do not be afraid of very difficult questions! If during the work with the target group, it turns out that the question is too difficult, and that the attempt to find the right answer lasts for more than 4 minutes, it will be made easier with additional information or imaginative guidance. Please remember, if the question is too easy, and the answer comes in a few seconds, the questions and a few minutes of a valuable atmosphere of intense thinking was lost. An example of our interactive multimedia e-application for the encouragement of a deep analytical (functional) thinking that includes 100 questions is given, designed in accordance with the requirements and respecting the characteristics of the target group, pupils of IV grade class of the elementary school. The application will be able to be upgraded (expand, aggravate, facilitate with additional questions...) with new content. In figures 1 and 2, one task is presented in a virtual environment.



Fig. 1. Primeval enigmatic question

14. TASK: Enigmatic Primeval question – “Which came first, the chicken or the egg?”

Kindly write on the line under question mark WHO or WHAT is missing, and then write ordinal numbers (from 1 to 5) in squares, so that the images are sorted in the exact order and to get the correct answer!

Why it is important to encourage deep analytical thinking among children. First, because it is the right time for that when we are children, and another very important reason is that employers give more and more attention to this component of future candidates for the job.

Thus, even from 2011, Glassdoor.com website, dedicated to providing business pieces of advice, gathered at one place the most unusual questions which employers give to candidates, in order to show to unemployed people, who are seeking for jobs, on what challenges they can encounter in one interview. For example:

“How would you measure the weight of an elephant without using a beam scale?” – IBM.

“How many basketball balls can fit in this room?” – Google, these are some of the “apparently” bizarre questions which managers are asking, not expecting a correct answer, but to see how a person behaves in difficult situations, said experts.



Fig. 2. The answer to the enigmatic question

“The essence is to show a sense of humor, interest for studying, teamwork and to like intellectual challenges,” said John M McKee, founder of website BusinessSuccessCoach.net and author of the book called “The Plan” [30].

3.2 Case study: Student success system

Student Success System is a product of D2L, which was founded by John Baker in 1999. D2L Technology is used by 15 million students worldwide in the higher student sector. Student Success System (S3) can help the at-risk students by their predictive models based on the learning engagement and behavior patterns. S3 is an SLA tool for

seamless and context-aware learning. The educators may intervene in the learners to help them to achieve better performance. Administrators of the system can create target courses. The instructors may analyze the weekly predictions generated by the system for the students. Five domains in the system are taken into account. They are content access, preparedness, assessments, social learning, and course access. The class dashboard can visualize the performance of the student with the rest of the class. It also issued a success index. On the student dashboard, a win-loss chart is also displayed. The students are categorized into three classes at risk, potential risk, and success. The instructor may filter out one particular category of students and emailed them as a group.

3.3 Case study: Loco analyst

Loco analyst another smart learning tool specially designed for the instructors to get feedback on different leaning aspects of the system. The teachers may take the help of the system to design better course content and structure. Loco analyst provides feedback with the teachers about the social interactions among the learners, the usage of the web-based course contents, and the learning activities of the students. Learning Object Context (LOC) framework is orchestrated to track the student activities with the system for the learning process. This open-source is an extension of RCP Editor for creating online courses in compliance with IMS CP specification.

3.4 Case study: GLASS (Gradient’s Learning Analytics SyStem)

To track and analyze the different events of the educators and learners, a graphical learning analytics tool was developed and called GLASS [31]. Several recorded events are explored and stored in a database to support modular visualization. Bottom-up technology was employed in architecture. Web interface, visualization parameters, module management, and database access are the prime functionality of the tool. Contextualized Attention Metadata (CAM) schema was utilized for storing in the dataset. A facility of incorporating a new module can be achieved through an API with a different set of filters and common graphical settings. Visualization, modules, codebase, and data layers are the four layers of the design. The data layer is responsible for storing parameters and comprised of CAM databases. The codebase is com-

posed of interfaces and user management. Modules are at par with the specifications of the platform and settings for visualization and appearance.

3.5 Case study: GISMO (Graphical Interactive Student Monitoring Tool for Course Management Systems)

GISMO is an interactive graphical student tracking and monitoring tool for a course management system [32]. The tool generates a visual representation of the online data of the course and student management. The course instructors may be benefited by using these records to track the progress and different aspects of the students. The system that generated complex multidimensional data of the learners can be rendered by the Information Visualization techniques of the GISMO. Hence this tool helps educators to impart their teaching in a better way and provide a better learning environment for the students as well. Data Explorer module tracks the learners' events and exports the logs to MySQL database. To represent data visually, some transformation and computations on the data are required. It is done in four stages. They are the selection of significant data, the transformation of data, visualization mapping, and transformation view. The three different areas of application are time, list, and chart panel. There are four overviews viz. Login, Discussion, Students accesses to resources, resource details under these areas.

4 Experiments and Results

4.1 Dataset

Student results dataset for one semester are used to inspect the association between students' semester course and results (pass & fail). The data dimension is represented by 37 columns and 121 rows (tuples) because there are different types of exams (Quizzes, Assignments, Mid-term, and Final-term) student-level results. The variables used in this study are the number of exams in a student semester course, course quizzes output (0), course assignments output (0), Mid-term course output (0), and course Final-term output (0). Accuracy can be applied as evaluation metrics to test the efficacy of the system [33].

4.2 Methods

Ensemble methods consist of meta-algorithms that incorporate various machine learning techniques into one predictive model to minimize variability (bagging), bias (boosting), or improve results (stacking). Ensemble techniques are defined in two types: linear ensemble methods (e.g., AdaBoost) and simultaneous ensemble techniques (e.g., Random Forest). Sequential techniques utilize the reliance between the base learners to enhance general efficiency while, by averaging, parallel techniques utilize independence between the base learners to decrease inaccuracies. Most ensemble techniques use a single base learning algorithm to generate homogeneous base

learners that result in inhomogeneous groups, while some techniques use heterogeneous learners. A 10-fold cross-validation technique is applied to test the validity of the results. The accuracy of ensemble methods depends upon the accuracy and diversity of its individual members, the base learners.

4.3 Advanced ensemble techniques

It is needed to choose a base learner algorithm to use bagging and boosting methods [34]. Hence different ensemble techniques are chosen for bagging and boosting.

Bagging (Bootstrap Aggregating): Bagging is the aggregation of bootstrap. One way of reducing an estimate's variance is to average various estimates together. Bagging uses bootstrap sampling to get the base learner training data subsets. Bagging uses voting for classification and averaging for regression to aggregate the outputs of base learners. Bootstrap Aggregation is an ensemble technique. First, random samples of the training dataset with substitute (sub-sets of training dataset) were generated. A model (classifier or decision tree) for each sample is constructed. Finally, the outcomes of these various schemes are merged using the median or majority vote. Multiple model combinations decrease variance, particularly in case of unstable models

Boosting: Boosting is a sequential method where the first algorithm is educated in the whole dataset to suit the remaining components of the first algorithms, and the subsequent algorithms are engineered so that the findings poorly predicted by the previous model become more effective. It depends on a number of weak learners who might not be good for the whole set of data but are good for some part of the collection. Therefore, each model improves the production of the ensemble.

Decision tree: A decision tree determines the predictive value based on a series of questions and conditions [35]. Decision Trees can also solve quantitative problems as well with the same format. In the student course process, there are three different types of courses, such as undergraduate, masters, and doctoral (Ph.D.), is it an undergraduate degree course? Master's degree course? A doctoral degree course? Pass course? Fail course? These questions are answered and solved using this decision tree. In the decision tree method, the performance of student courses through smart learning can be predicted.

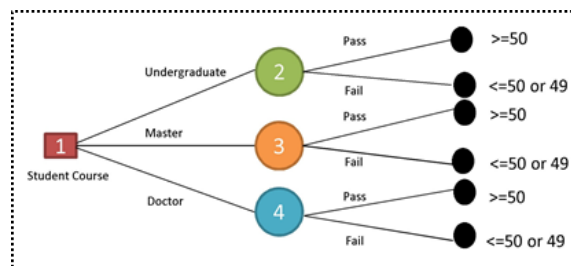


Fig. 3. Decision Tree of Student Performance

K-Nearest Neighbours: K-nearest neighbor (KNN) is an algorithm that uses monitored sample data classified by the majority class [36]. This algorithm is designed to classify new products based on training data and features. The KNN principle is to find k products from the training data closest to the information testing algorithm is very simple, it works on the nearest range between test documents for training data to determine KNN, followed by the test information prediction for the majority of KNN. The KNN method provides countless advantages over other methods, such as solid, too noisy training data and effective when the information is small. This method is extremely prone to irrelevant or redundant variables since all variables are comparable and categorized. KNN uses a similarity measure when comparing the testing information to the training data. One of the similarity measures used is the Euclidean distance between training data (xtrain) and data testing (xtest).

Random Forest: Random forest is a group approach for building predictive models [37]. In this sense, the random forest is a series of decision-making bodies acting as weak categorizers, which are poor predictors but a strong general prediction. Random forests do not heavily assume the scale and normality of incoming data. It is done well with a combination of numbers and categories, need a little tuning to obtain a good first version of a predictive model, are fast to train, are simple to understand, have functional importance as a function of the model, can manage the missing data inherently and have been implemented in all languages. These characteristics are a nice starting point for random trees in any project where a predictive model is generated or explore the option of applying machine learning to a new field.

Naive Bayes classification: Naive Bayes classification scheme is a simple probabilistic classification based on Bayes theorem and particularly suitable if the sizes of the input are big [36]. A class in text classification is assigned to the designated document.

Stacking: Stacking involves training a learning algorithm, which combines the forecasts of several other learning algorithms and sometimes called stacked generalizations. All other algorithms are first trained using the available data, and then a mixed algorithm is trained to generate a final prediction by all the other algorithms. The stacking method generates better performance than any trained models. It has been used efficiently for both supervised learning operations (regression, classification, distance learning, and uncontrolled density assessment).

Advanced ensemble techniques results: Unlike the K-NN-bagging ensemble, the decision-tree bagging ensemble has achieved higher accuracy because K-NN is less susceptible to sample disruption and thus is stable learners [38]. Combining stable learners is less advantageous as the ensemble does not increase the effectiveness of generalization. Four different methods are implemented viz. decision tree, K-NN, bagging K-NN, and bagging tree, and each method achieves different accuracy on student data. The accuracies for decision tree and bagging K-NN are 72% and K-NN 74% respectively. The best performance is achieved on student data by bagging tree classifier, and it achieves an accuracy of 79%, which is the proposed method. Table 1 shows the detail description of the accuracy of methods, and Figure 4 is showing the training and test performance on student data. Figure 5 is showing the graphical representation of different methods.

Table 1. Predicted accuracy of different methods.

Implemented Methods	
Methods	Accuracy
Decision Tree	0.72
K-NN	0.74
Bagging K-NN	0.72
Bagging Tree	0.79

Figure 4 shows the learning curves for the bagging tree ensemble. An average error of 0.100 on the training data and a polynomial-shaped error curve for the testing data can be seen. The smallest gap between training and test errors occurs at around 90% of the training set size.

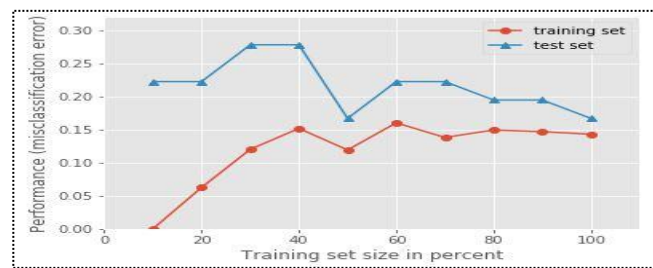


Fig. 4. Bagging Tree for Student Performance

Figure 5 shows that the accuracy increases significantly between ensemble sizes 0 to 15, but does not show much improvement after that, so adding base estimators beyond 15 only increases computational complexity without accuracy gain for the dataset. Figure 6 shows the comparison of accuracies of different bagging methods applied in the dataset.

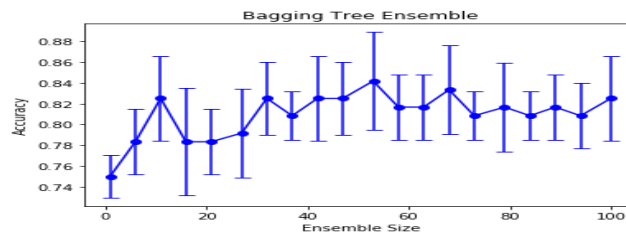


Fig. 5. Accuracy of Bagging Tree Ensemble of Student Performance

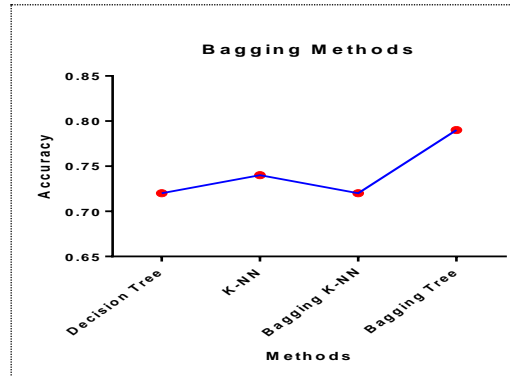


Fig. 6. Comparison of Bagging Methods on Student Performance

Stacking is an ensemble learning method that uses a meta-classifier or meta-regressor to combine various classification or regression designs. The base-level models are trained on the basis of a complete training set, and the meta-model is trained as features on the outputs of the base-level models. The base-level often comprises distinct learning algorithms, and thus stacking groups are often heterogeneous. Four different methods are implemented viz. stacking, random forest, K-NN, naïve Bayes, and stacking classification, and each method achieves different accuracy on student data. The accuracies for both naïve Bayes and K-NN are 74%, and for the random forest, it is 77%. The best performance is achieved by stacking classification with an accuracy of 78%, which is the proposed method. Table 2 shows the detail description of the accuracy of methods, and Figure 7 is showing the training and test performance on student data. Figures 8 and 9 are showing the graphical representation of different methods.

Table 2. Predicted accuracy of different methods

Implemented Methods	
<i>Methods</i>	<i>Accuracy</i>
Random Forest	0.77
K-NN	0.74
Naive Bayes	0.74
Stacking Classifier	0.78

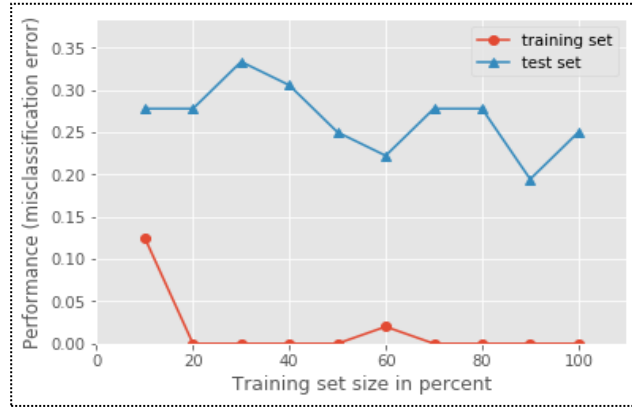


Fig. 7. Stacking Ensemble of Student Performance

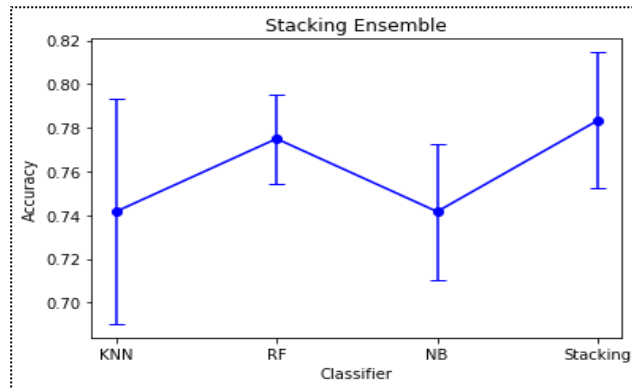


Fig. 8. Accuracy of Stacking Ensemble for Student Performance.

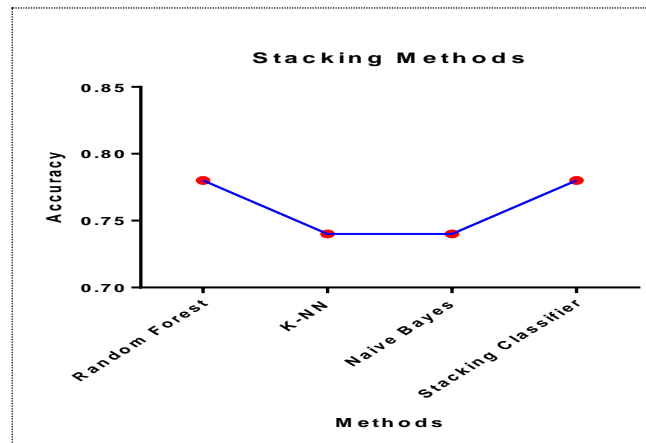


Fig. 9. Comparison of Stacking methods for Student Performance.

5 Discussion and Conclusion

As studied without learning and practicing in real-time and the real world, learners always find it challenging to learn and solve the problems they face, traditional methods or learning from the textbooks do not meet students' requirements today. In SLA, competency and proficiency are two main key factors that will lead to performance appraisal in technology-enabled and resource-rich industries. Now that drawing private sector attention education is becoming a global commodity making the future educational process more viable for individual requirements. Moreover, learners will be able to decide what to learn and how to learn, or they will have control of their own learning process, choosing their own instructors and have their own assessments. This will be inevitable with SLA. While SLA holds the key for the future of education, there are some obstacles that researchers and practitioners have to overcome challenges such as technological or social. When processing a big chunk of data, which helps learners and teachers on context-based tutoring is taken into account, it can be easily understood that SLA is still very much in its early phase, which is also promising for the better future use. Environments are a challenge to researchers, practitioners; educational experts feel safe in their comfort zone. SLA is such a new concept for developing and improving, creating a more stable and more helpful learning environment than its traditional counterpart showing what is more technologically advanced. An SLA dataset with the help of advanced ensemble techniques is successfully analyzed. The techniques that achieved higher accuracies were Bagging Tree and Stacking Classifiers. So, advanced ensemble techniques may be applied in such a dataset for prediction and classification tasks.

6 Acknowledgement

Work by author, Solomon Sunday Oyelere was supported by the Jenny and Antti Wihuri Foundation.

7 References

- [1] Agbo F. J., Oyelere S. S., Suhonen J., & Tukiainen M. (2019). Smart learning environment for computing education: readiness for implementation in Nigeria. In J. Theo Bastiaens (Ed.), *Proceedings of EdMedia + Innovate Learning 2019* (pp. 1382-1391). Waynesville, NC: Association for the Advancement of Computing in Education (AACE). <https://doi.org/10.1504/ijlt.2019.106551>
- [2] Agbo F.J., Oyelere S.S. (2019) Smart Mobile Learning Environment for Programming Education in Nigeria: Adaptivity and Context-Aware Features. In: Arai K., Bhatia R., Kapoor S. (eds) *Intelligent Computing. CompCom 2019. Advances in Intelligent Systems and Computing*, vol 998. Springer, Cham. https://doi.org/10.1007/978-3-030-22868-2_71
- [3] Merrill, M. D. (2013). *First Principles of Instruction: Identifying and Designing Effective, Efficient, and Engaging Instruction*. San Francisco, CA: Pfeiffer.

- [4] Kim J, et al. (2018). Bre1 mediates the ubiquitination of histone H2B by regulating Lge1 stability. *FEBS Lett* 592(9):1565-1574. <https://doi.org/10.1002/1873-3468.13049>
- [5] R. Huang, J. Yang, Y. Hu, From digital to smart: the evolution and trends of the learning environment. *Open Educ. Res.* 1, 75-84 (2012)
- [6] Zhu, Z. T., Yu, M. H., & Riezebos, P. (2016). A research framework of smart education. *Smart learning environments*, 3(1), 4. <https://doi.org/10.1186/s40561-016-0026-2>
- [7] Hwang, G-J. (2014). Definition, framework and research issues of smart learning environments – a context-aware ubiquitous learning perspective. *Smart Learning Environments*, Vol. 1, No. 4, pp. 1-14. <https://doi.org/10.1186/s40561-014-0004-5>
- [8] Chan, F. M. (2002, October). ICT in Malaysian schools: Policy and strategies. In a Workshop on the Promotion of ICT in Education to Narrow the Digital Divide (pp. 15-22).
- [9] Hua, M. T. A. (2012). Promises and threats: iN2015 Masterplan to pervasive computing in Singapore. *Science, Technology and Society*, 17(1), 37-56. <https://doi.org/10.1177/097172181101700103>
- [10] IBM, SmartEducation(2012).https://www.ibm.com/smarterplanet/global/files/au_en_uk_cities_ibm_smarter_education_now.pdf. Accessed 20 Mar 2015.
- [11] Park, Sun-Rae, et al. "Intelligent u-Learning and Research Environment for Computational Science on Mobile Device." *KSII Transactions on Internet & Information Systems* 8.2 (2014). <https://doi.org/10.3837/tiis.2014.02.023>
- [12] Kim, Taisiya, Ji Yeon Cho, and Bong Gyou Lee. "Evolution to smart learning in public education: a case study of Korean public education." *IFIP WG 3.4 International Conference on Open and Social Technologies for Networked Learning*. Springer, Berlin, Heidelberg, 2012. https://doi.org/10.1007/978-3-642-37285-8_18
- [13] Long, P., & Siemens, G. (2011). Penetrating the fog: Analytics in learning and education. *Educause Review Online*, Vol. 46, No. 5, pp. 31-40.
- [14] Papamitsiou, Z. and Economides, A.A. (2016). Learning Analytics for Smart Learning Environments: A Meta-Analysis of Empirical Research Results from 2009 to 2015. J.M. Spector et al. (eds.), *Learning, Design, and Technology*. Springer International Publishing, Switzerland. https://doi.org/10.1007/978-3-319-17727-4_15-1
- [15] Kinshuk, Vivekanandan Kumar, (2018). Advancing learning through smart learning analytics: a review of case studies, *Asian Association of Open Universities Journal*, Vol. 13 Issue: 1, pp.1-12. <https://doi.org/10.1108/aaouj-12-2017-0039>
- [16] Kim, S.; Song, S.M.; Yoon, Y.I. Smart Learning Services Based on Smart Cloud Computing. *Sensors* 2011, 11, 7835-7850. <https://doi.org/10.3390/s110807835>
- [17] Yoo Y., Lee H., Jo IH., Park Y. (2015). Educational Dashboards for Smart Learning: Review of Case Studies. In: Chen G., Kumar V., Kinshuk, Huang R., Kong S. (eds) *Emerging Issues in Smart Learning. Lecture Notes in Educational Technology*. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-44188-6_21
- [18] Kim, T., Cho, J. Y., & Lee, B. G. (2012, July). Evolution to smart learning in public education: a case study of Korean public education. In *IFIP WG 3.4 International Conference on Open and Social Technologies for Networked Learning* (pp. 170-178). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-37285-8_18
- [19] Ebner, M., Taraghi, B., Saranti, A. and Schön, S. (2015). "From the field: seven features of smart learning analytics – lessons learned from four years of research with learning analytics", *eLearning Papers* no. 40, pp. 51-55.
- [20] Kinshuk, Chen, NS. Cheng, IL. et al. (2016). Evolution Is not enough: Revolutionizing Current Learning Environments to Smart Learning Environments, *Int J ArtifIntellEduc*, 26: 561, <https://doi.org/10.1007/s40593-016-0108-x>.

- [21] Gros, B., (2016). The design of smart educational environments, *Smart Learn. Environ.* 3: 15. <https://doi.org/10.1186/s40561-016-0039-x>.
- [22] Nikolov, R., Shoikova, M., Kovatcheva, E., Dimitrov, V., & Shikalanov, A. (2016). Learning in a smart city environment. *Journal of Communication and Computer*, 13, 338-350. <https://doi.org/10.17265/1548-7709/2016.07.003>.
- [23] Soltanpoor R. and Yavari A., (2017). CoALA: Contextualization Framework for Smart Learning Analytics, *IEEE 37th International Conference on Distributed Computing Systems Workshops (ICDCSW)*, Atlanta, GA, 2017, pp. 226-231, <https://doi.org/10.1109/icdcs.2017.58>
- [24] Yassine, Sahar & Kadry, Seifedine & Sicilia, Miguel-Angel. (2016). Measuring learning outcomes effectively in smart learning environments. 1-5. <https://doi.org/10.1109/ssfc.2016.7447877>.
- [25] Mikulecky Peter (2016). Smart Learning Environments Revisited, *The 11th International Scientific Conference on Distance Learning in Applied Informatics*, pp 33-42.
- [26] Ahad, M.A., Tripathi, G. & Agarwal, P. (2018). Learning analytics for IoE based educational model using deep learning techniques: architecture, challenges, and applications, *Smart Learn. Environ.* 5: 7. <https://doi.org/10.1186/s40561-018-0057-y>
- [27] Shoikova Elena, Nikolov, Roumen, Kovatcheva, Eugenia, (2017). Conceptualizing of Smart Education, *ELECTROTECHNICA & ELECTRONICA*, 52(3-4), pp 29-37.
- [28] Rajović, R. (2009). Nikola Tesla Centar (NTC) IQ deteta – brigaroditeljja, autorskoizdanje, Novi Sad
- [29] Petrovic, N., Jeremic, V., Petrovic, D., & Cirovic, M. (2014). Modeling the use of Facebook in environmental higher education. In *The social classroom: Integrating social network use in education* (pp. 100-119). IGI Global. <https://doi.org/10.4018/978-1-4666-4904-0.ch006>
- [30] Hilčenko, S. (2015). A model: animated logical mathematical tasks for the younger elementary school children, *Journal of Research & Method in Education (IOSRJRME)*, Volume 5, Issue 4, pp. 39-42.
- [31] Leony, D., Pardo, A., de la Fuente Valentín, L., de Castro, D. S., & Kloos, C. D. (2012, April). GLASS: a learning analytics visualization tool. In *Proceedings of the 2nd international conference on learning analytics and knowledge* (pp. 162-163). ACM. <https://doi.org/10.1145/2330601.2330642>
- [32] Mazza, R., & Milani, C. (2004, November). Gismo: a graphical interactive student monitoring tool for course management systems. In *International Conference on Technology Enhanced Learning*, Milan (pp. 1-8).
- [33] Hussain, S., Muhsin, Z. F., Salal, Y. K., Theodorou, P., Kurtoğlu, F., & Hazarika, G. C. (2019). Prediction Model on Student Performance based on Internal Assessment using Deep Learning. *International Journal of Emerging Technologies in Learning*, 14(8), 4-22. <https://doi.org/10.3991/ijet.v14i08.10001>
- [34] Dietterich, T. G. (2000). An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization. *Machine learning*, 40(2), 139-157.
- [35] Sunday, K., Ocheja, P., Hussain, S., Oyelere, S. S., Samson, B. O., & Agbo, F. J. (2020). Analyzing Student Performance in Programming Education Using Classification Techniques. *International Journal of Emerging Technologies in Learning (iJET)*, 15(02), 127-144. <https://doi.org/10.3991/ijet.v15i02.11527>
- [36] Hussain, S. (2017). Survey on current trends and techniques of data mining research. *London Journal of Research in Computer Science and Technology*.

- [37] Hussain, S., Atallah, R., Kamsin, A., & Hazarika, J. (2018, April). Classification, clustering and association rule mining in educational datasets using data mining tools: A case study. In *Computer Science On-line Conference* (pp. 196-211). Springer, Cham. https://doi.org/10.1007/978-3-319-91192-2_21
- [38] Dietterich, T. G. (1998). An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting and randomization. *Machine learning*, 32, 1-22.

8 Authors

Samina Kausar is currently working as a researcher and also a PhD. scholar at School of Computer Engineering and Science, Shanghai University, China. She received the M.S. degree in computer science from the International Islamic University Islamabad, Pakistan. She is also working as Assistant Professor at University of Kotli, Azad Jammu and Kashmir, Pakistan. Her research interests are in the fields of big data, bioinformatics, computer networks, cloud computing, data mining, E-learning, and machine learning.

Solomon Sunday Oyelere received a PhD. degree in Computer science from the University of Eastern Finland, MSc degree in Research in Computer and Systems Engineering from Ilmenau University of Technology and B.Tech (Hons) degree in Computer science from Federal University of Technology Yola. Solomon's research interest focuses mainly on improving learning environments through smart technology, pedagogy, and content. He applies techniques from multidisciplinary areas including educational technology, interactive mobile computing, web technology, computing education, and educational data mining.

Yass Khudheir Salal was born in Al-Qadisiyah, Iraq on Sep. 06, 1983. He received the B.E. degree in Computer Science from Al-Qadisiyah University, Faculty of Science, Al-Qadisiyah, Iraq, in 2004. He received the M.E. degree from Belarus State University of Electronic and radio communications, security department of information and data protection, Minsk, Belarus, in 2014, He is currently pursuing PhD. degree in the department of system programming in Graduate School of Electronics and Computer Science, South Ural State University (SUSU), National Research University, Chelyabinsk, Russia. His research interest includes Big Data, Machine learning.

Sadiq Hussain is System Administrator at Dibrugarh University, Assam, India. He received his PhD. degree from Dibrugarh University, India. His research interest includes data mining and machine learning. He is associated with Computerization Examination System and Management Information System of Dibrugarh University.

Mehmet Akif Cifci received the BSc degree in the English language from the University of 9 September, Izmir, Turkey and the MSc from Istanbul Aydin University, Istanbul, Turkey. He is now working toward the PhD. degree in computer science in Istanbul, Turkey. His research focuses on applied machine learning, Networks science and mining, and machine learning.

Slavoljub J Hilcenko was born in Crvenka in 1962. The elementary school he finished partly in Crvenka/Serbia and partly in Wien/Austria. He finished Art High

school in Novi Sad and Pedagogic academy in Sombor in 1985. In 1997 he graduated at the Faculty of Pedagogy in Sombor. From 1994 to 2008 he was employed as a teacher in elementary school "Nikola Vukićević" in Sombor. From the year of 2008, he works as associate professor at the Vocational College for Education of Preschool-Teachers and Sports Trainers in Subotica. He teaches at the subjects of Informatics, Elementary forming of mathematical notions, IKT in educations. Postgraduate studies at the Technical Faculty in Zrenjanin of the University of Novi Sad he started in 1998 and received his magister degree in 2000. His doctoral thesis he defended in 2003.

Muhammad Shahid Iqbal is currently working towards his Ph.D. at the School of Computer Science and Application Technology at Anhui University Hefei China. His research interest includes computer vision, image processing, machine learning, automated software engineering, and software testing.

Zhu Wenhao was born in Sep. 1979 and currently working as an Associate Professor in the School of Computer Engineering and Science, Shanghai University, China. His research is in the areas of text representation, information extraction, and web data mining. He received his bachelor's, master's, and Ph.D. degrees from Zhejiang University in 2002, 2006, 2009 respectively. He also worked as a research scholar for one year (from 2012 to 2013) at the Computer Laboratory, University of Cambridge.

Xu Huahu is a Doctoral Supervisor and a Professor of the School of Computer Engineering and Science, Shanghai University. At the same time, he is the Director of the Information Office of Shanghai University and chairman of Shanghai Security and Technology Association. His research interests include multimedia technology, CIMS and computer network Technology.

Article submitted 2020-01-28. Resubmitted 2020-03-20. Final acceptance 2020-03-22. Final version published as submitted by the authors.