

E-Learning Recommendation System for Big Data Based on Cloud Computing

<https://doi.org/10.3991/ijet.v16i21.25191>

Mounia Rahhali^(✉), Lahcen Oughdir, Youssef Jedidi
Sidi Mohamed Ben Abdellah University, Fez, Morocco
mounia.rahhali@usmba.ac.ma

Abstract—In educational institutions, E-learning has been known as a successful technology for enhancing performance, concentration, and thus providing higher academic success. Nevertheless, the conventional system for executing research work and selecting courses is a time-consuming and unexciting practice, that not only directly impacts the students' academic achievement, but also impacts the learning experience of students. In addition to that, there is an enormous number of various kinds of data in the E-Learning domain both structured and unstructured, and the academic establishments attempt to manage and understand big complicated data sets. To fix this problem, this paper proposes a model of an E-learning recommendation system that will suggest and encourage the learner in choosing the courses according to their needs. This system used big data tools such as Hadoop and Spark in order to enhance data collection, storage, analysis, processing, optimization, and visualization, furthermore based on cloud computing infrastructure and especially Google cloud services.

Keywords—Recommender System, E-learning, Big Data technologies, Cloud Computing, Hadoop, Spark, Online learning

1 Introduction

E-learning represents a meaningful role in helping large educational institutions meet their learning and training requirements. Presently education is increasingly correlated with ICT (Information and communication technology), so educational institutions have a high interest in servers, storage, and software [1]. Furthermore, they need access to the necessary resources to enhance the performance and reduced the cost of production education for their learners [2]. In addition, most educational institutions utilize big amounts of information in both structured and unstructured types. These large amounts of data cannot be treated by traditional learning management systems. Therefore, the processing and use of these data utilizing big data technology are necessary.

Big data is defined as databases that are very large in volume. It can be defined in terms of data processing problems that cannot be handled with conventional databases owing to increasing volume, velocity, and variety of data. Data mining is one of the methods in the pipeline of Big Data. Data mining is a set of techniques that extract

information from massive datasets [3]. They must help to make decisions. In recently, Educational Data Mining (EDM) is an autonomous research area with a set of machine and behavioral methods and analysis techniques to help understand learners and their teaching methods [4]; EDM develops strategies and applies techniques from machine learning, statistics, and data processing [5] to improve the quality of education. To enhance the level of learning, EDM builds solutions and implements approaches from machine learning and data analysis.

Recommender systems are increasingly common as a research and application field, with a range of relevant application areas including e-commerce, movies, and several others. The recommendation system has recently received attention in the educational sector, as well as generating different types of suggestions for learners, instructors, universities, etc.

This research is concentrated on the recommendation system approaches for big data, produced from educational institutions. The approach considered in this section utilizes Hadoop and Spark big data techniques, to produce suggestions for learners to select courses, and based on cloud computing that can be merged to enhance the efficiency of any e-learning system, facilitate content access, and customize students.

The rest of the paper is structured as follows: Section 2 outlines the associated work; Section 3 gives an overview of recommendation systems and techniques of recommendation. Section 4 introduces the relevant big data tools and techniques. The architecture of the proposed model is explained in section 5. Conclusion and future work are presented in section 6.

2 Related works

Surabhi et al. [6] used collaborative filtering-based proposal approaches to propose facultative courses to understudies, contingent on their evaluation focuses obtained in various topics. Similarity Log-likelihood is applied to find patterns between grades and topics. Within this study, they specified the importance of the recommender system for the enormous quantity of informative information.

Simović [7] suggested a big data smart library that can enhance the continuing learning operation by recommendations and increase user service. He also created a method for gathering, analyzing and handling big data from different sources.

Dahdouh et al. [8] applied association rules mining in the recommendation system to resolve student activities. The Parallel FP-growth algorithm of the MLlib machine learning library was applied to create a recommendation system. For effectiveness, the system can process massive of data and scalable compute capacity.

Otoo-Arthur et al. [9] proposed a framework that employed both batch and streaming dataset of learners online activities on moodle LMSs. Their method utilizes a distributed computing ecosystem and gives a flexible conventional system to enhance the data acquisition, storage, processing and analysis, for e-learning systems. BiDel system execution presents advanced data integration and data governance.

Most of the relevant literature discussed for this research does not present an intelligible model for the correspondence and combination of big data in the recommenda-

tions system. Because the volume of information produced by the new educational institutions is massive, we will use big data technologies to provide learners recommendations to enhance data collection, storage, analysis, processing, optimization, and visualization, furthermore based on cloud computing infrastructure and especially Google cloud services.

3 Recommender system

Recommender Systems (RSs) are automated tools and strategies that allow recommendations to users about items that may be of importance to them. For Schein et al. [10] Recommender systems indicate users' items of interest depending on their direct and indirect interests, other users' interests, and user and item features. Any of the practical implementations that use such devices may involve (recommending books, products, videos, jobs, music).

3.1 Recommender system approaches

Recommendation approaches are widely studied and largely divided into three sections: content-based, collaborative filtering (CF)-based, and knowledge-based.

Content-Based: in which items identical to those that the learner has previously chosen are suggested. In most cases, the items are described by a common set of attributes. The interests of the student are estimated by looking at the correlation of the item ratings with the corresponding item features. Consequently, without support from other students, students may obtain useful suggestions.

Collaborative Filtering: in which the user is recommender courses that learner with equivalent interests and behaviors desired in the past. Although it is the oldest recommendation approach, this approach is also very helpful and does not need to include the description of the entity that the systems can understand quickly.

If item 1 read by both learners if learner A read item 2. Therefore, there is a high chance that learner B may like item 2 because, from the first two statements, we knew that learner A and learner B read item 1.

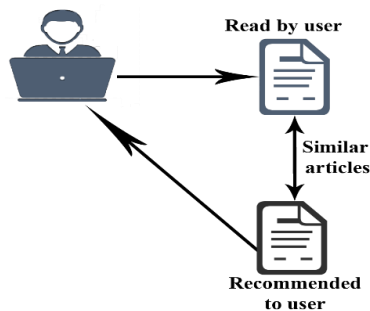


Fig. 1. Content-based filtering

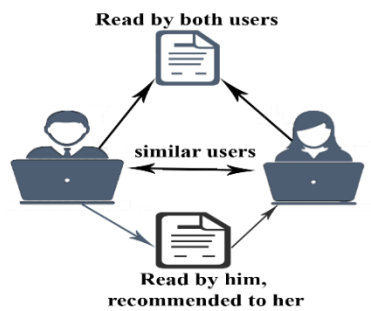


Fig. 2. Collaborative filtering

The secret to effective collaborative recommendations is the capacity to establish significant interactions between learners and their items choices to support the end learner in prospective dealings.

Although there are numerous collaborative filtering techniques, they can essentially be divided into two main categories:

- a) **Memory-based Approaches:** The memory-based approach uses user-rating data to calculate the resemblance between learners or items. Such techniques have two main steps: calculation of similarities between user and items using rating information and predicting the unknown rating and thus providing either a single value or a list of top n items that the person may additionally like.
- b) **Model-Based Approaches:** In this methodology, models are created utilizing various data mining, Machine-learning calculations to foresee users' evaluating of un-rated items.

Knowledge-Based: Knowledge recommender systems are particular types of recommender systems based on clear knowledge of the range of items, consumer interests, and parameters for a recommendation. In these approaches, the similarity function measures the degree to which the user's requirements (problem description) complement the recommendations (solutions of the problem). Here it is important to clearly view the similarity score as the effectiveness of the consumer recommendation [11].

Hybrid Recommender system: A hybrid recommendation that blends the attributes of two or more different types of recommendation to profit from the advantages of each method and enhance efficiency [12]. The hybrid recommendation approach is really helpful and much of the difficulties identified by the independent recommendation methods can be solved [13].

There are various forms of hybridization [14]:

1. Implementing CF and CB separately and combine their predictions.
2. Integrating some content-based characteristics into a collaborative approach.
3. Integrating some collaborative characteristics into a content-based approach.
4. Creation of a general unifying paradigm that combines both content-based and collaborative characteristics.

3.2 Recommendation system applied in education

E-Learning is a general term that represents learning conducted on a computer, usually linked to a network allowing us the ability to learn almost anywhere, anytime [15]. E-learning is also an innovative means of delivering long-term education, in contrast with conventional face-to-face teaching and learning [16]. Besides, the e-learning system enables learners to manage time and offer additional information to facilitate student learning. However, the conventional method for doing study work and selecting courses is time-consuming and uninteresting, affecting not only the academic success of students but also the learning environment of learners. In reality, students could have various preferences; although sharing similar interests, they may

have varying degrees of competence and thus cannot be handled uniformly. It is essential that a personalized system is developed that can automatically respond to students' preferences and levels.

Systems that regain and filter the data by content and comparable profiles are known as recommendation systems (RS) [17]. These systems are very common in the research and application field, with several good application areas such as e-commerce, game, social media and others. Nevertheless, its use has expanded to include domains such as movie and music recommendations and education. Today, the research about recommender systems in an educational context has significantly advanced [18].

According to [17], various educational fields covered by RS:

- Supply advice to learners on their educational options. This includes suggesting learners a place to study that could be a college, school, and institute.
- Promoting learning by RS. Studies like [19] suggested a recommender system implemented in E-learning in order to be smart to present personalized courses.
- Applied RS to improve educational achievement.
- Employ recommender systems to recommend online courses from various vendors.
- Develop a suggested design for a recommendation system for scholarship's purpose [20].

3.3 Advantages of e-learning during pandemic situation

The pandemic of COVID-19 made various institutions and universities remain temporarily locked. This closing tested the preparation of colleges to deal with a crisis that needs the assistance of advanced technology to allow efficient E-learning. This closing spurred the increase of E-learning activities so that there would be no delay to study. E-Learning provides learners with the opportunity to study at any place and at any time, including their requirements excellently. It presents online programs and learning resources following different classes such as Data Science, Design and HealthCare. For example, edX is an online learning platform. It hosts and offers valuable free online university-level courses and resources from around the world.

Faced with several courses and resources, it is tricky for students in the e-learning environment to determine associated learning resources. Besides, due to the diversity in the background; various students have various learning demands [2]. Accordingly, it is essential to improve recommender systems to assist learners in choosing courses, resources, or learning materials in E-learning [21]. E-learning has earned popularity due to the COVID-19 pandemic. It also encourages its advancement to deal with difficulties and gives them a more agreeable user experience.

In this part, we present some benefits of E-learning in the pandemic Covid-19 situation:

- E-learning aided assure distance learning, it was flexible, and learners could handily arrive teachers and educational materials.
- E-learning is also adaptable for individual learner's demands and level of knowledge.

- Enabling e-learning could support introverted students. Through e-learning, they could be more sure to share their opinions and engage in their classroom debate [22].

4 Big data platforms

4.1 Definition and features of big data

Big data is a vast and complicated selection of data sets that could not be perceived, acquired, handled, and analyzed using conventional data analysis applications and on-hand information systems software within a tolerable time. Besides the mass of data, several other characteristics define the distinction between "massive data" and "very big data".

In reality, a selection of meanings for big data is available in the literature:

SAS Company describes Big Data as follows: "Big data is a concept that represents the vast amount of data that floods a company on a day-to-day basis, including structured and unstructured. However, it is not the quantity of data that is essential. It is what companies are doing with the data that counts" [23].

In 2011, McKinsey Global Institute described big data as "Big data refers to data which volume is outside the capacity of traditional database analysis tools to collect, process, organize, and evaluate" [24]. Two meanings are included in this definition: first, the sizes of the database that correspond to the big data norm are developing and can increase with time or through technological progress; Second, the sizes of the database that correspond to the big data norm vary from each other in varied uses [25].

In order to understand Big Data, we are trying to discuss 4Vs. The Big Data solution should handle the "4Vs" of the big data:

1. Volume represents the size of the Terabytes (TB) to Petabytes (PB) data and associated large data structures, include documents, transfers, folders, and tables.
2. Velocity relates to techniques of transmitting big data, like batch, near time real-time, and stream data. Velocity also requires the pacing and delay features of data processing. Data can be analyzed, interpreted, saved, and handled rapidly.
3. Variety of big data extends to and incorporates various data types, including structured, unstructured, semi-structured data. The data format can be in the form of records, e-mails, text messages, audio, photographs, video, graphics, etc.
4. Value relates to advantages/value earned.

4.2 Big data preprocessing

Owing to the large range of data sources, the datasets obtained differ widely in noise, redundancy, accuracy, etc., which is definitely a waste to archive insignificant data. In contrast, some analytical methods include specific data coherence criteria [25]. Data preprocessing will then adjust the data to the criteria within each DM algo-

rithm, allowing the handling that would otherwise be unworkable [26]. Data preprocessing aims to cleaning and improving input data so that an ML process may be later implemented easier and more effectively.

Difference between Batch Processing and Stream Processing:

- Batch processing:
 - Processing of data blocks already stored by a given period.
 - This data contains millions of records for each day that can be stored as text files (CSV), or records stored in HDFS, SQL DBMS, NoSQL, etc.
- Stream processing:

Contrary to batch processing where data is linked with a start and an end in a process that ends after finished data processing, stream processing is for processing endless data streams arriving in real-time continuously for days, months, years, and forever. It allows data to be introduced into analysis tools as soon as they are generated and to obtain instant analysis results.

Two approaches to setting up a streaming framework:

- Native Streaming (Real-Time processing): Each incoming record is processed as it arrives, without waiting for the others. For example, Storm, Flink, Kafka streams, Samza.
- Micro Batch Processing (Micro Batching): Incoming records every few seconds are batched, then processed in a single mini-batch with a delay of a few seconds. For example: Spark Streaming, Storm-Trident.

Data mining is simply part of the mechanism of knowledge discovery in Fig. 3. The concept of data mining is now used in a vast range of applications of all forms in the medical fields and security, financial data collection, media, and in particular, education [4]. For example, in the education field, the establishment can have the information required to take decisions before certain learners may drop out, or accurately select resources with a reliable estimation of how many learners may attend a specific course.

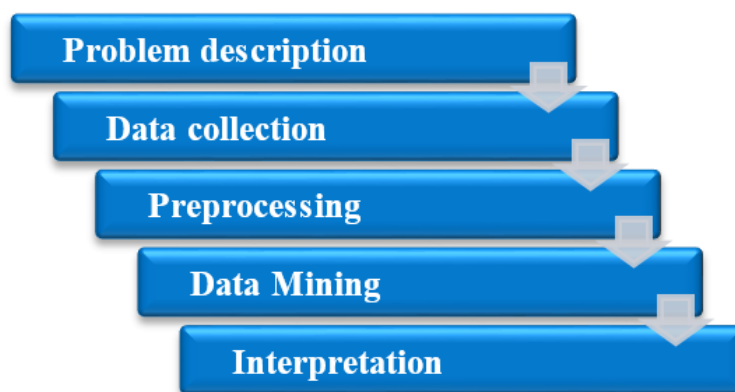


Fig. 3. Knowledge Discovery in Databases Process[26]

The most critical point to note is that no single technique or collection of techniques is uniformly available. For any given query, the design of the data would have an impact on the technique we select. As a result, we would need a number of techniques and technology to identify the best available model.

In this section, we provide a summary of the major data mining methods used in the recommender systems field.

Classification: In the classification problem, the features are classified into two parts: a multiplicity of feature space and a label space [27], where the features reflect the characteristics of the items to be categorized and the labels reflect the categories [28].

Clustering: In the clustering challenge, we organize related files together in a large database of multidimensional files. This produces data fragments that are identical inside a set of points. Relying upon the application, per of these features may be handled differently [27].

Association Rule Mining: The most useful question when interacting with more than one data kind is what is the sense of the relationship between the two data types: is it powerful, feeble, or is there no correlation at all? The aim is to determine constraints that will indicate the probability of an element based on the occurrences of other elements in a transaction [28].

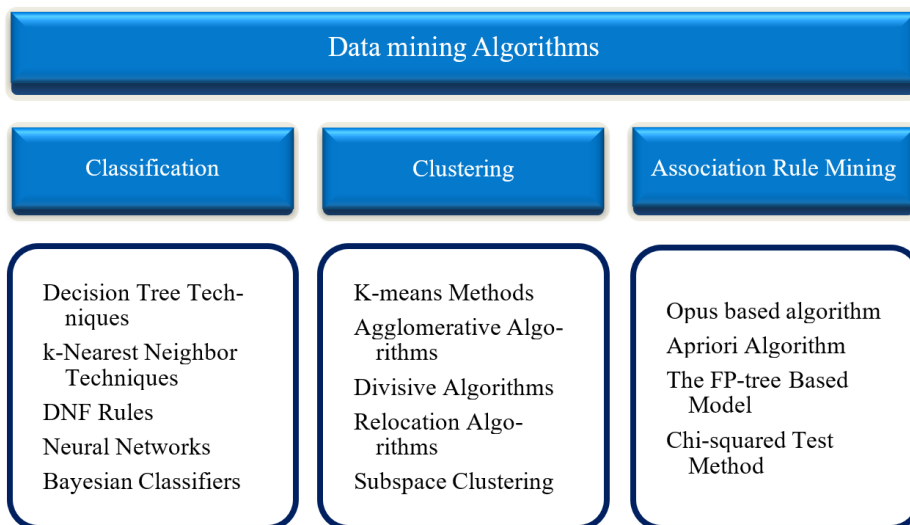


Fig. 4. Data mining algorithms

4.3 Big data technologies

Big Data technologies are a modern evolution of technologies and architectures developed to allow high-speed collection, exploration, and analysis to efficiently derive value from extremely large quantities of a wide range of data. Technologies for big data involve:

- a) Batch type treatments
- b) Real-time processing (streaming)
- c) NoSQL databases
- d) Data mining tools and techniques
- e) Cloud computing platforms

4.4 Big data layered architecture

As suggested in [29], a big data structure can be described using a layered structure.

The layered structure can be divided into three layers, i.e. the infrastructure layer, the computing layer, and the application layer.

The infrastructure layer is composed of a collection of processing and storage resources that can be managed into a Cloud computing infrastructure and made accessible through virtualization technologies. They would satisfy the need for big data in terms of optimizing device usage and resource capacity.

The computing layer produced numerous data techniques into a middleware layer that works via raw ICT resources. Typical technologies provide data integration, data management, and programming models:

- a) Data integration involves the collection of data from various sources and the incorporation of the data set into a single collection with the requisite data pre-processing activities.
- b) Data management relates to processes and techniques that offer continuous data storage and highly effective control, including distributed file systems and SQL or NoSQL data storage.

The programming model applies conceptual application logic and enables data analysis software.

The application layer provides programming model interfaces for the implementation of different data analysis operations, such as statistical analysis, clustering, classification, data mining, etc., and for the development of different big data applications.

4.5 Big data software

On various platforms, it is possible to manage big data. Two widely used frameworks are Hadoop and Spark. Apache Spark is commonly used to handle large volumes of data and to provide control for real-time analytics.

- a) Hadoop Platform: Apache Hadoop is a software application utilized for distributed file system and the processing of big data. It handles datasets of large data in the context of the MapReduce framework. The following are the key constituents of Hadoop: [6]
- b) Hadoop HDFS - Hadoop Distributed File System (HDFS) is the storage unit of Hadoop.
- c) Hadoop YARN is Hadoop's Resource Control Module.
- d) Hadoop MapReduce is based on YARN system, is the Hadoop treatment unit.
- e) Hadoop Common supports the other Hadoop modules.

The Hadoop Distributed File System (HDFS): is a distributed file system built to operate on devices. HDFS is extremely fault-tolerant and is built to be used on low-cost devices. HDFS offers superior throughput access to server data and is ideal for systems with large amounts of data. HDFS is now a sub-project of Apache Hadoop.

Spark Platform: is a unified analytics engine for large-scale data processing. It is a computational platform, which provides superior efficiency for both batch and interactive handling [30]. Spark enables a collection of libraries like SQL and DataFrames, MLlib for machine learning, GraphX, and Spark Streaming. In the same framework, you can merge several libraries perfectly. Spark Core is the most primary building block of Spark as presented in Fig. 5. Spark Core facilitates the in-memory computations that drive the parallel and distributed processing of data [31].

Spark Streaming: This element deals with processing the real-time streaming data in a scalable and fault-tolerant way. It utilizes micro batching to read and treat incoming streams of data.

Spark MLlib: is the fundamental Machine Learning API for Spark. The conventional method of building ML models utilizing Python’s scikit learn library fronts much of difficulties when data size is enormous whereas MLlib is created in a way that gives feature engineering and machine learning at scale [31]. ML Algorithms constitute the center of MLlib. These involve common learning algorithms such as regression, clustering, recommendation system, and natural language processing.

Spark GraphX: This element surpasses in graph analytics and graph parallel execution.

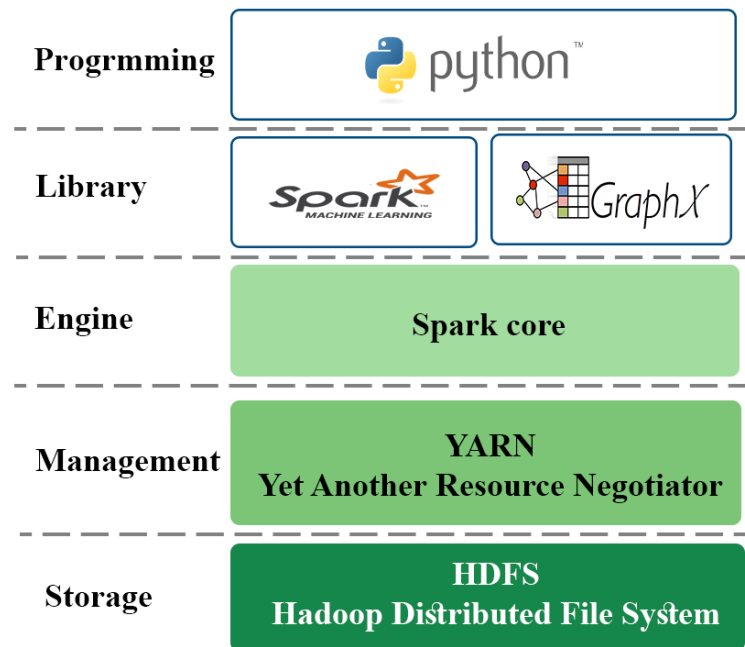


Fig. 5. Apache spark Framework

4.6 Cloud computing importance in big data

Dependable hardware infrastructures are essential to ensuring reliable computing in the big data model. The hardware architecture requires a mass of Flexible Distributed Information and Communication Technology (ICT) tools. Cloud computing refers to the distribution and utilization of the IT infrastructure, i.e. the procurement of the required services on-demand or in an adjustable manner over the Internet [25]. These services may be linked to the web and applications, or other services.

Big data is intimately associated with cloud computing. Big data is the purpose of the process of computation and demands a cloud server's processing power and computing ability. The key goal of cloud computing is to use immense computational resources and processing capabilities into focused control in order to supply granularity for information storage applications and to provide computing ability for big data technologies [25].

5 The Design of Recommender System using Big Data technologies based on Cloud computing

5.1 The recommendation process

First, from the diverse, multi-source, and irrelevant data, we collect student information, item information, and interest information to create the model of students and items through data ETL. Then create multiple separate recommendation systems by employing various algorithms. The method can be described in Fig. 6.

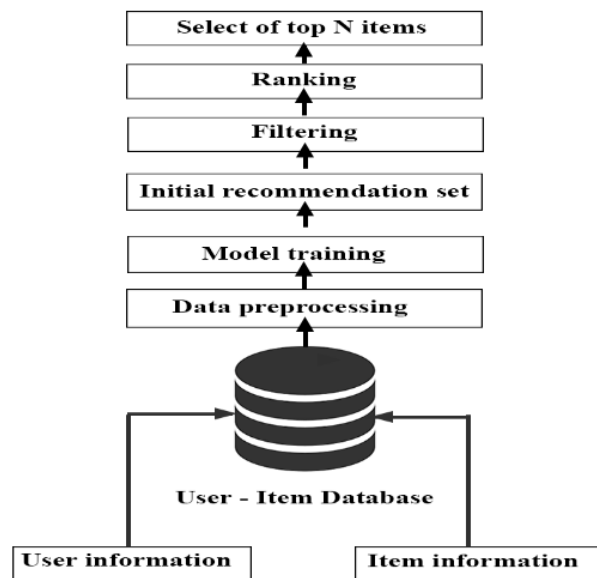


Fig. 6. Recommendation process

5.2 The proposed architecture of big data for e-learning recommendation system based on cloud

This part explains the overall architecture of our E-learning recommendation system using big data tools based on Google Cloud Platform depicted in Fig. 7, which can be separated into six-module: Data sources, Data storage, Data preprocessing layer, Data query layer, Data analytics, the Application layer.

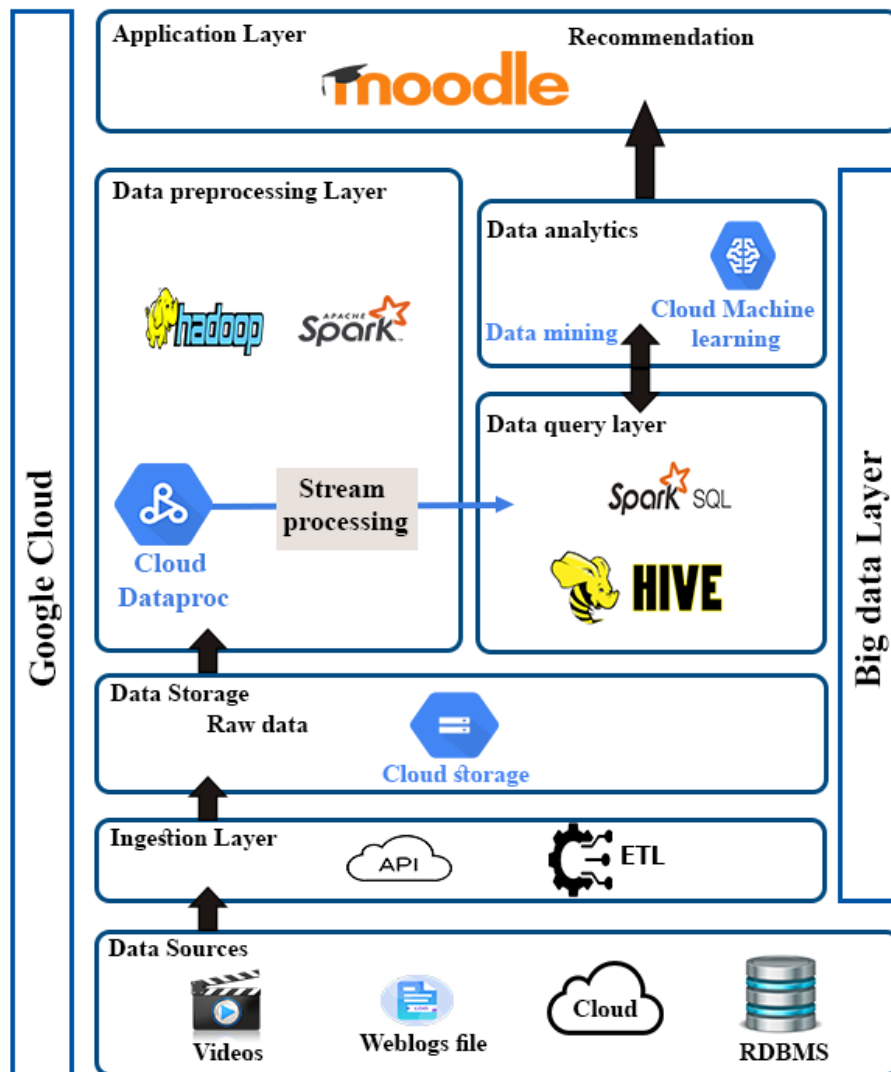


Fig. 7. Proposed architecture for e-learning recommender system

Data sources include all those fabulous sources from which the data extraction process is made, and thus this can be considered the initial step for the big data process. Datastores of platforms such as relational databases. Files are generated by a variety of platforms and are mostly part of static file systems.

Ingestion Layer: This layer is the first phase for the data arriving from variable sources to begin its travel. This layer is accountable for merging the unstructured, multi-structured, and multi-sources data, and collecting users' features and items' features through API or other such as Apache Flume (unstructured data), Apache Sqoop (structured data), which can aid immensely to enhance the efficacy of the recommender system.

Data storage: This layer concentrates on "where to keep such big data effectively". Raw data storage system stocks available information fully, safely, and continuously for potential data mining and analysis purposes

Data preprocessing layer: In this preliminary layer, a computing system is used to process large raw data with the aim to extract useful information. We can say that the information we obtained in the preceding layer is being handled in this layer. We can use a Dataproc service: it makes it possible to process Open Source data analysis (Apache Hadoop, Apache Spark, Kafka etc.) in the cloud quickly, easily, and more securely within the Google Cloud Platform ecosystem. It is a service that is completely devoted to processing and improving data in stream mode.

Data query layer: This is the layer where the analytical handling is involved. The preliminary objective here is to collect the value of the data to make it more usable for the following layer.

Data analytics: In this module, we begin with a data-mining algorithm or machine learning in an attempt to discover an appropriate model for data.

Application layer: This layer is an interface between students and the device, like setup, control, feedback, and display. It enables the modification of variables and indicators of hybrid recommendation systems to ensure the efficiency of the framework.

5.3 Setting up the Environment Configuration

a) Launching Instance

In order to give a proposal for implementing our design displayed in Fig.7 we create a Linux virtual machine instance in Compute Engine using the Google Cloud Console.

- Operating system: Ubuntu 20.04 LTS.
- 2 vCPU With an allocation of 4GB main memory

b) Data Science Environment Set Up

When we launch a Google Compute Engine instance, it is started from the Google Compute Engine page. This console is used to set up the data science environment, using Dataproc, we create Hadoop and Spark cluster. For cluster type, we choose a standard multi-node cluster. The configuration could yet, be scaled up from single machines to multiple machines. We created our environment big data

utilizing for storage: Hadoop Distributed File System (HDFS) Apache Hadoop 3.2.2 version, and for processing data, we utilize Apache Spark 3.0.1 version.

This form is a trial implementation for combining big data tools and techniques in E-learning. Our future works will present more details on the processes of implementation, algorithms, and results.

6 Conclusion and future works

The main aim of the current work was to design a model of a recommender system for finding obtain higher-quality resources and achieving the learning objective. The development of big educational data is similar to the development of e-learning and associated systems. Therefore, the principal contribution of this article is the recently suggested model design for handling big educational data produced by the Moodle system using big data technologies. At the same time, advanced cloud computing technologies will allow academic institutions to enhance their services to satisfy their requirements and facilitate the management of data and educational resources. Our next work will involve more details on the processes of implementation, algorithms, and results and adapting it to the education of our university.

7 References

- [1] N. V. Naik and K. Madhavi, "Cloud computing architecture for collaborative e-learning system," presented at the 2015 International Conference on Applied and Theoretical Computing and Communication Technology (iCATecT), 2015. <https://doi.org/10.1109/icatct.2015.7456855>
- [2] M. Rahhali, L. Oughdir, Y. Jedidi, Y. Lahmadi, and M. Z. El Khattabi, "E-learning Recommendation System Based on Cloud Computing," in *WITS 2020*, vol. 745, S. Bennani, Y. Lakhrissi, G. Khaissidi, A. Mansouri, and Y. Khamlichi, Eds. Springer Singapore. https://doi.org/10.1007/978-981-33-6893-4_9
- [3] Y. Jedidi, A. Ibriz, M. Benslimane, M. Tmim, and M. Rahhali, "Predicting Student's Performance based on Cloud Computing," in *WITS 2020*, vol. 745, S. Bennani, Y. Lakhrissi, G. Khaissidi, A. Mansouri, and Y. Khamlichi, Eds. Springer Singapore. https://doi.org/10.1007/978-981-33-6893-4_11
- [4] A. Hicham, A. Jeghal, A. Sabri, and H. Tairi, "A Survey on Educational Data Mining [2014-2019]," presented at the 2020 International Conference on Intelligent Systems and Computer Vision (ISCV), Jun. 2020. <https://doi.org/10.1109/iscv49265.2020.9204013>
- [5] M. Salihoun, "State of Art of Data Mining and Learning Analytics Tools in Higher Education," *Int. J. Emerg. Technol. Learn.*, vol. 15, no. 21, p. 58, Nov. 2020, <https://doi.org/10.3991/ijet.v15i21.16435>
- [6] S. Dwivedi and V. K. Roshni, "Recommender System for Big Data in Education," 2017. <https://doi.org/10.1109/eleltech.2017.8074993>
- [7] A. Simović, "A Big Data smart library recommender system for an educational institution," *LHT*, vol. 36, no. 3, pp. 498–523, Jun. 2018, <https://doi.org/10.1108/lht-06-2017-0131>

- [8] K. Dahdouh, A. Dakkak, L. Oughdir, and A. Ibriz, "Large-scale e-learning recommender system based on Spark and Hadoop," *J Big Data*, vol. 6, no. 1, p. 2, Dec. 2019, <https://doi.org/10.1186/s40537-019-0169-4>
- [9] D. Otoo-Arthur and T. L. van Zyl, "A Scalable Heterogeneous Big Data Framework for e-Learning Systems," in *2020 International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems (icABCD)*, Durban, South Africa, Aug. 2020, pp. 1–15. <https://doi.org/10.1109/icabcd49160.2020.9183863>
- [10] U. Javed, K. Shaukat, I. A. Hameed, F. Iqbal, T. Mahboob Alam, and S. Luo, "A Review of Content-Based and Context-Based Recommendation Systems," *Int. J. Emerg. Technol. Learn.*, vol. 16, no. 03, p. 274, Feb. 2021, <https://doi.org/10.3991/ijet.v16i03.18851>
- [11] F. Ricci, L. Rokach, and B. Shapira, "Recommender Systems: Introduction and Challenges," in *Recommender Systems Handbook*, F. Ricci, L. Rokach, and B. Shapira, Eds. Boston, MA: Springer US, 2015, pp. 1–34. https://doi.org/10.1007/978-1-4899-7637-6_1
- [12] K. I. Ghauth and N. A. Abdullah, "Measuring learner's performance in e-learning recommender systems," *AJET*, vol. 26, no. 6, Oct. 2010, <https://doi.org/10.14742/ajet.1041>
- [13] J. K. Tarus, Z. Niu, and D. Kalui, "A hybrid recommender system for e-learning based on context awareness and sequential pattern mining," *Soft Comput*, vol. 22, no. 8, pp. 2449–2461, Apr. 2018, <https://doi.org/10.1007/s00500-017-2720-6>
- [14] S. Meenakshi and M. Sandeep, "A Survey of Recommender Systems: Approaches and Limitations," *IJIET*, 2013.
- [15] T. Huiyi, L. Yong, and G. Junfei, "E-Learning Recommendation System," presented at the 2008 International Conference on Computer Science and Software Engineering (CSSE 2008), Hubei, Dec. 2008. <https://doi.org/10.1109/CSSE.2008.305>
- [16] R. Sikka, A. Dhankhar, and C. Rana, "A Survey Paper on E-Learning Recommender System," *IJCA*, vol. 47, no. 9, pp. 27–30, Jun. 2012, <https://doi.org/10.5120/7218-0024>
- [17] A. C. Rivera, M. Tapia-Leon, and S. Lujan-Mora, "Recommendation Systems in Education: A Systematic Mapping Study," in *Proceedings of the International Conference on Information Technology & Systems (ICITS 2018)*, vol. 721, Á. Rocha and T. Guarda, Eds. Cham: Springer International Publishing, 2018, pp. 937–947. https://doi.org/10.1007/978-3-319-73450-7_89
- [18] N. Manouselis, H. Drachler, K. Verbert, and E. Duval, *Recommender Systems for Learning*. New York, NY: Springer New York, 2013. <https://doi.org/10.1007/978-1-4614-4361-2>
- [19] K. Dahdouh, L. Oughdir, A. Dakkak, and A. Ibriz, "Smart Courses Recommender System for Online Learning Platform," in *2018 IEEE 5th International Congress on Information Science and Technology (CiSt)*, Marrakech, Oct. 2018, pp. 328–333. <https://doi.org/10.1109/cist.2018.8596516>
- [20] F. M. Pinto, M. Estefania, N. Cerón, R. Andrade, and M. Campaña, "iRecomendYou: A Design Proposal for the Development of a Pervasive Recommendation System Based on Student's Profile for Ecuador's Students' Candidature to a Scholarship," in *New Advances in Information Systems and Technologies*, vol. 445, Á. Rocha, A. M. Correia, H. Adeli, L. P. Reis, and M. Mendonça Teixeira, Eds. Cham: Springer International Publishing, 2016, pp. 537–546. https://doi.org/10.1007/978-3-319-31307-8_56
- [21] O. Mazhoud, A. Kalboussi, and A. H. Kacem, "Educational Recommender System based on Learner's Annotative Activity," *Int. J. Emerg. Technol. Learn.*, vol. 16, no. 10, p. 108, May 2021, <https://doi.org/10.3991/ijet.v16i10.19955>
- [22] A. Sufyan, D. Nuruddin Hidayat, A. Lubis, U. Kultsum, M. Defianty, and F. Suralaga, "Implementation of E-Learning During a Pandemic: Potentials and Challenges," in *2020*

- 8th International Conference on Cyber and IT Service Management (CITSM), Pangkal Pinang, Indonesia, Oct. 2020, pp. 1–5. <https://doi.org/10.1109/citsm50537.2020.9268900>
- [23] “Official Webpage of SAS Company,” 2021. https://www.sas.com/en_us/insights/big-data/what-is-big-data.html (accessed Dec. 16, 2020).
- [24] J. Manyika *et al.*, “Big data: The next frontier for innovation, competition, and productivity.” 2011. [Online]. Available: <https://www.mckinsey.com/>
- [25] M. Chen, S. Mao, Y. Zhang, and V. C. Leung, “Big data related technologies, challenges and future prospects,” *Inf Technol Tourism*, vol. 15, no. 3, pp. 283–285, Oct. 2015, <https://doi.org/10.1007/s40558-015-0027-y>
- [26] J. Luengo, D. García-Gil, S. Ramírez-Gallego, S. García, and F. Herrera, “Introduction,” in *Big Data Preprocessing*, 2020, pp. 1–14. https://doi.org/10.1007/978-3-030-39105-8_1
- [27] C. C. Aggarwal and P. S. Yu, “Data Mining Techniques for Associations, Clustering and Classification,” *Lecture Notes in Computer Science*, pp. 13–23, 1999, https://doi.org/10.1007/3-540-48912-6_4
- [28] X. Amatriain, A. Jaimes, N. Oliver, and J. M. Pujol, “Data Mining Methods for Recommender Systems,” in *Recommender Systems Handbook*, 2011, pp. 39–71. https://doi.org/10.1007/978-0-387-85820-3_2
- [29] H. Hu, Y. Wen, T. Chua, and X. Li, “Toward Scalable Systems for Big Data Analytics: A Technology Tutorial,” *IEEE Access*, vol. 2, pp. 652–687, 2014, <https://doi.org/10.1109/access.2014.2332453>
- [30] K. Dahdouh, A. Dakkak, L. Oughdir, and F. Messaoudi, “Big data for online learning systems,” *Educ Inf Technol*, vol. 23, no. 6, pp. 2783–2800, Nov. 2018, <https://doi.org/10.1007/s10639-018-9741-3>
- [31] P. Singh, *Machine Learning with PySpark*. 2019. <https://doi.org/10.1007/978-1-4842-4131-8>

8 Authors

Mounia Rahhali is graduated from ENS (higher education establishment, Tetouan) with a degree of master specialized in educational technology in 2019. She is currently a Ph.D. candidate at ISA-Laboratory/ Sidi Mohamed Ben Abdellah University of Fez under the supervision of Pr. Lahcen Oughdir.

Lahcen Oughdir is a Professor in the Department of Mathematics, Physics and Computer Science at the National School of Applied Sciences, Sidi Mohamed Ben Abdellah University (SMBAU), Fez, Morocco (Email: lahcen.oughdir@usmba.ac.ma).

Youssef Jedidi is graduated from ENS (higher education establishment, Tetouan) with a degree of master specialized in educational technology in 2019. He is currently a Ph.D. candidate at IT-Laboratory/ Sidi Mohamed Ben Abdellah University of Fez (Email: youssef.jedidi@usmba.ac.ma).

Article submitted 2021-07-01. Resubmitted 2021-08-13. Final acceptance 2021-08-14. Final version published as submitted by the authors.