SCRF: Strategic Course Recommendation Framework

[1]Mohit Patel, [2]Gerard Deepak
[1]Department of Information and Communication Technology
[1] Pandit Deendayal Energy University, Gujarat
[2]Manipal Institute of Technology Bengaluru, Manipal Academy of Higher Education, India
[2]gerard.deepak.christuni@gmail.com

Abstract: The need for course recommendation in the specialised domain is salient and therefore, this paper proposes a Strategic Course Recommendation Framework (SCRF) for recommending courses pertaining to finance and economics as a prospective domain. This paper encompasses a semantic-oriented learning and reasoning model for achieving accurate course recommendations in the era of Web 3.0. The model uses a knowledge stack comprising eBooks, Blogs, and Web pages relating to finance and economic domains. This knowledge stack is actually used to enhance the auxiliary knowledge between the data-enriched terms and the query-obtained terminologies. The semantic indexing and bag of words add to the auxiliary knowledge of the query to contextualize the query words that are pre-processed. Semantics-oriented learning is achieved through a strong deep learning classifier employed to classify the financial course dataset. The usage of Second Order Co-occurrence Pointwise Mutual Information (SOC-PMI), Normalized Pointwise Mutual Information (NPMI), and Pianka Index ensures strong relevance computation mechanisms at three distinct stages in the model to allow the convergence of knowledge with all of the data set enriched terms and the query enriched terms in order to facilitate the best-in-class foundations of courses in a highly cohesive web 3.0 environment. The proposed model, which is the best-in-class model when compared to all the other strategic models, has a total precision of 94.45%, a recall of 96.81%, and a False Detection Rate (FDR) of 0.06.

***Keywords*:** Semantic Similarity, Semantic Web, SOC-PMI, NPMI

1 Introduction

Systems for recommending courses are becoming more and more important in today's educational environment for a number of reasons. These programs use artificial intelligence and data-driven algorithms to aid students in making educated choices in their academic careers. It has become difficult for students to sort through all the possibilities in a specialized subject as a result of the growth of online education platforms and the availability of a wide range of courses at traditional universities. A course recommendation system can streamline the process by suggesting suitable courses based on a student's interests or domain, academic background, and professional aspirations. It guarantees that students select the best courses, eliminating pointless detours and reducing the possibility of squandering time and money on pointless courses. A thoughtful course recommendation system takes into account a student's career goals and ambitions before recommending classes that will give them the skills and information they need for their chosen profession. Their prospects of finding employment in their preferred field can be considerably increased by this alignment. There is a need for a strategic semantics-oriented hybrid intelligence-driven framework for recommending courses specifically for finance and economics as a domain. There are not many frameworks to recommend courses specifically focused on financial or economics as a domain of choice and the lack of methods that encompass learning, hyphen reasoning, inferencing through fact-based reasoning, and convergence of auxiliary knowledge into the model is absent. Henceforth, a novel framework has been proposed.

Implementation of strategic inclusion of knowledge to that of the query words using structural topic modelling and the bag of words model, ensuring a high degree of enrichment of the query words and contextualizing it, which is a novel approach. Subsequently, the dataset is also enriched by encompassing the Wiki data API and also classified using the strong deep learning LSTM model, which makes the dataset more permeable into the model, which is also a novel way to handle the dataset. Another innovative contribution involves the use of three distinct similarity measures, namely the SOC PMI, NPMI, and Pianka Index, in three different stages to transform knowledge from the knowledge stack, enriched keywords from query words, and the course dataset. The usage of a static knowledge stack- free books, blogs, and web pages related to the domain of finance and economics ensures a strong level of semantics and a strong density of auxiliary knowledge which inculcates diversity and as well as relevance into the model. The suggested model has improved overall precision, recall, and accuracy, making it the best in its class when contrasted to all other standard models.

Organization: This paper is structured as follows. Section 2 depicts the related works, section 3 presents the proposed system architecture. The results are depicted in section 4. Section 5 is the conclusion of the paper.

2 Related Works

Joy et al., [1] suggested an ontology-based paradigm for the purpose of content recommendation. A more effective CRF has been produced using RDF structure and adaptive learning profile analysis. Zhang et al., [2] have suggested using a hierarchical reinforcement learning method to update user profiles and optimize the course recommendation model based on the updated profiles, even when the user has numerous interests in several courses. Hao et al., [3] proposed a Meta-Relationship Course Recommendation (MRCRec) that improves the representation of relationship information. Meta Relationship (MR) is created by focusing on the complicated semantic content of multi-entity relationships and entity associations. Pang et al., [4] proposed that It has been demonstrated that Collaborative Filtering (CF) is efficient in recommending items to users who share those items' interests. An approach known as Multi-Layer Bucketing Recommendation (MLBR) is suggested to recommend courses on MOOCs in light of the efficacy and efficiency of CF. By transforming student vectors into the same length dimension and distributing them into buckets containing comparable learners who have courses similarity. Lin et al., [5] Dynamic Attention and Hierarchical Reinforcement Learning (DARL), a unique course suggestion framework, was put out to enhance the model's degree of flexibility. Gong et al., [6] recommended an end-to-end graph neural network-based technique for knowledge idea recommendation in MOOCs called Attentional Heterogeneous Graph Convolutional Deep Knowledge Recommender (ACKRec). Zhang et al., [7] proposed MCRS, a distributed association rules mining algorithm based on the enhanced aPriori algorithm is used in a distributed computation paradigm. Zhang et al., [8] have suggested the Knowledge Grouping Aggregation Network (KGAN), a potent course recommendation algorithm that automatically predicts learners' possible interests and iteratively uses the course graph, a heterogeneous graph that depicts the links between courses and facts. Tian et al., [9] proposed that a time-effectiveness hypothesis be used to elicit the implicit reaction on a following course in MOOC recommendation models using Multidimensional Item Reaction Theory (MIRT). Wang et al., [10] created a method for utilizing attention-based convolutional neural networks (CNN) to gather user profiles, forecast user ratings, and suggest the best courses. Learner behavior and learning histories are shown as feature vectors. Ye et al., [11] proposed CERec-ME, or community-enhanced course concept recommendation, is a suggestion containing several elements. Deepak et al., [12-17] have proposed various semantics oriented frameworks for recommendation in various domains.

3 Proposed System Architecture

The design of the recommended course recommendation system for finance and economics is depicted in Fig.1. The query- or subject-initiated model's first processing stages, which include tokenization, lemmatization, stop-word removal, and Named Entity Recognition (NER), are applied to the user-inputted question or the chosen topic. Once the input query or the user-selected subject has been pre-processed, the individual query keywords or query words are retrieved. They require the addition of supplementary knowledge, for which Latent Semantic Indexing (LSI) is used for topic modeling because they are less informative in nature. By using the web as a reference corpus and applying mathematical models to LSI to produce the terms that are pertinent to the query, topic modeling is a paradigm where topics are relevantly concealed and undiscovered. The topics discovered through LSI are still less informative as they are limited. Hence, the bag of words model is applied in order to discover many more terms and add to the auxiliary knowledge. A popular and simple method for information retrieval and natural language processing (NLP) is the ‘Bag of Words’ (BoW) model. It depicts text documents as numerical feature vectors, ignoring word order and grammar in favor of just looking at word presence and frequency. Tokenization,

Fi

vocabulary development, and feature vectorization are the three primary phases of the BoW model. Since the data collection is categorical, it goes through a category extraction process before being submitted to the wiki data API to add supplemental information. All of the cases obtained through the wiki data API are given when the categories are provided into the API. Wikidata is a knowledge store repository that houses information, which is community-contributed community verified, and crowdsourced in the form of open-linked data. Instances that are linked from the wiki data API is yielded and also subsequently from the financial document data set. It is then subjected to classification using the LSTM model (Long Short-term Memory model), which is a strong deep-learning model. Experimentation and tuning are necessary for LSTM network parameterization. The hyperparameter space can be explored and the best configurations can be found through random search, grid search, or automated methods like Bayesian optimization.

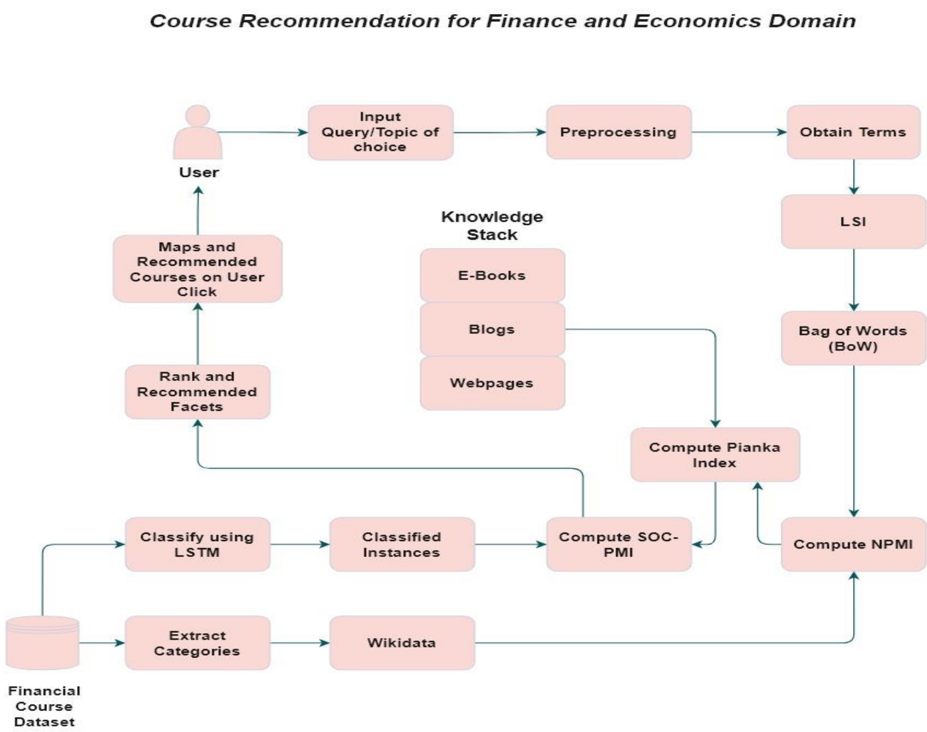

Fig.1. Proposed System Architecture

LSTM is preferred as the feature selection is implicit and auto-handcrafted and is viable for large-scale data sets. Therefore, on classifying the financial code data set with the LSTM, the classified instances are yielded and the yielded classified instances that come out of the LSTM are used for the model. A knowledge stack comprising of eBooks repositories pertaining to finance and economics as the domain, blogs put into finance and economics as a domain of choice, and web pages regarding finance and economics as a domain are gathered and housed in the knowledge stack. From these eBooks, the index terms and the glossaries are extracted and the keywords from the blogs and web pages are extracted and delivered for the computation of the Pianka index as depicted in Equation (1).

$$\sum_i \frac{x_i y_i}{[(\sum_i x_i^2)(\sum_i y_i^2)]^{\frac{1}{2}}} \qquad (1)$$

Pointwise mutual information (PMI) is a metric used to evaluate the association between two events. It assesses the probability of two events happening together in comparison to their probability if they were unrelated. PMI is considered a crucial concept in Natural Language Processing (NLP) and is related to the notion that the most effective approach is measuring the connection between two words is to determine how much more frequently they co-occur in a corpus compared to what would be expected by chance alone.Assuming that two discrete random variables, X and Y, are independent, the PMI (Pointwise Mutual Information) between a pair of outcomes, x and y, measures the chance of their occurrence based on their joint distribution and individual distributions. Mathematically PMI is depicted as Equation (2).

$$pmi(x;y) = \log_2 \frac{p(x,y)}{p(x)p(y)} = log_2 \frac{p(x/y)}{p(x)} = log_2 \frac{p(y/x)}{p(y)} \qquad (2)$$

Using the Bayes theorem, the last two expressions are equal to the first. The expected value of the PMI (across all conceivable outcomes) is the mutual information of the random variables X and Y.In order to overcome a drawback of Pointwise Mutual Information (PMI), Normalised Pointwise Mutual Information (NPMI) normalizes the values to lie between -1 and 1. NPMI is depicted as Equation (3). h(x,y) (joint self-information) $-\log_2 p(x,y)$.

$$npmi(x;y) = \frac{pmi(x;y)}{h(x,y)} \qquad (3)$$

The resulting entities of wiki data and the LSI bag of words pipeline is subject to computation of Normalized Pointwise Mutual Information measure (NPMI) that has been set to the median threshold of 0.5 considering the positive values between 0 and 1 for the NPMI model. So, the matching instances, the output of NPMI is subject to compute semantics with that of the entities from the knowledge stack using a Pianka index, which is set to be assumed at step deviance of 0.25, because of the large number of entities which have to be anchored at this stage.The entities that come out of the Pianka Index are subjected to computation of second-order co-occurrence PMI (SOC PMI) with that of the classified instances from LSTM. SOC PMI is set to a threshold of 0.75 and the motive for setting at this specific value is because of the

Fi

higher relevancy of the entities. Hence, SOC PMI although being a very strong measure is set to have a threshold of 0.75. The SOC PMI measure's output entities are rated and suggested to the user as its aspects in ascending order. Based on the user clicks on these facets, the courses automatically get mapped from the financial course data set based on the categories. If the query is satisfied the process is terminated. Otherwise, the user clicks are recorded and it is fed as the query pre-processed terms, and this process is continued recursively until no further user clicks are ever logged, which explains that the user is satisfied with the results produced.

4 Results

Experimentations are conducted on three distinct datasets integrated as a single large dataset. The three subset datasets are the Kothari proposed business courses – Udemy (10k courses) dataset,]the integrated intellectual property office finance dataset, and the economics dataset by port4siro(2022). The literal data is not used, rather the terms in the dataset and the annotations are used to crawl courses from several other course-offering sites like Udemy, Coursera, Edx, etc. The essential data on the finance, as well as economics courses, were crawled and added to this dataset. Customer annotations were also executed and were integrated into a single large dataset. Depending upon the annotations they were rearranged and experimentation was conducted.The implementation of the proposed work is achieved using Python 3 with Google Collaboratory as a preferred IDE. The LSTM is configured through the Keras framework. SOCPMI, Pianka index, and NPMI are computed using a customized agent design using JADE. LSI and Bag of Words is also implemented using Python and the knowledge stack was formulated by crawling eBooks, blogs, and web pages using a customized crawler. The tasks involving open-action language processing were carried out using the NLTK model.

The suggested system architecture of the SCRF, a strategic course recommendation framework for economics and finance as a prospective area, is shown in Table 1. The identical set of data and the same number of queries were used to test the baseline models as well, and their results were summarized. 1141 questions were used in the experiments to test the suggested framework. The suggested SCRF is evaluated using the False Discovery Rate (FDR) as an adjunct parameter in addition to the Precision, Recall, Accuracy, and F-measure percentages. Expressed as percentages, Precision, Recall, and Accuracy allow us to assess the performance of all models since they indicate the importance of the findings and False Discovery Rate calculates the model's error rate.

Table 1. Comparison of Performance of the proposed SCRF with alternative approaches

Model	**Average Precision %**	**Average Recall %**	**Average Accuracy %**	**Average F-Measure**	**FDR**

				%	
OCRPL [1]	89.92	90.34	90.13	90.13	0.11
HRLCR [2]	91.12	92.45	91.78	91.78	0.09
MRCR [3]	91.89	93.07	92.48	92.47	0.09
CFCR [4]	92.63	94.08	93.35	93.35	0.08
Proposed SCRF [5]	94.45	96.81	95.63	95.61	0.06

Table 1 depicts that the suggested SCRF has yielded the highest average precision of 94.45%, the highest average recall of 96.81%, the highest average accuracy of 95.63%, the highest average F-measure of 95.6154%, and along with the FDR of 0.06. The Ontology model for content recommendation in the personalized learning environment (OCRPL), Hierarchical Reinforcement Learning for Course Recommendation (HRLCR), Meta-relationship for course recommendation (MRCR), and Collaborative filtering for Course Recommendation (CFCR) are the baselines of four different course recommendation frameworks with which the proposed SCRF is compared in order to assess its performance. From table 1 it is depicted that OCRPL yields 89.92% average precision, 90.34% average recall, 90.13% average accuracy, and 90.1295% average F-measure with the FDR of 0.11. The HRLCR has shown 91.12% average precision, 92.45% average recall, 91.785% average accuracy, and 91.7802% average f-measure with an FDR of 0.09. The MRCR model has yielded 91.89% average precision, 93.07% average recall, 92.48% average accuracy, 92.4762% average F-measure, and FDR of 0.09. The CFCR model furnished 92.63% average precision, 94.08% average recall, 93.355% average accuracy, 93.3494% average F-measure, and an FDR measurement of 0.08.

Because it uses tactical models like LSI and bag of words to enhance the preprocessed query terms, the suggested SCRF model—which is knowledge-centric and semantically inclined—has achieved the greatest accuracy, precision, recall, and F-measure percentage as well as the lowest value of FDR. By using the World Wide Web as a reference corpus, LSI and Bag of Words both collect information from their existing architecture. Apart from this, a knowledge stack of finance and economics domain comprising of eBooks, blogs and web pages discovered from the current structure of the web specifically pertaining to the documents related to the fields of finance and economics is used from which the entities are extracted on a large scale. Also, the classification of the financial course data set using the LSTM, which represents a strong deep-learning model that operates on the basis of autonomous feature selection is very compliant and has excellent performance when dealing with very big datasets. The usage of SOC-PMI, NPMI, and Pianka index for relevance computation of semantic similarity measure and step deviance measure guarantees that all baseline models are outperformed by the suggested SCRF model. As SCRF has strong knowledge aggregation ecosystems, as well as strong semantic similarity measures,

Fi

the best-in-class model is served by the suggested SCRF framework, which performs better than all baseline models.

The ontology model for content recommendation in a personalized learning environment is one that is based on ontologies and is used in customized learning environments (OCRPL). In this model, lightweight ontologies are directly used which incorporates the learner profile and then object attributes. So, learner personalization is the target of this model. A suitable knowledge model is provided by the ontology, and tools built on the Resource Description Framework (RDF) are used to describe the data. However, the relevance computation is quite weak. There are no strong models for a little computation which leads to loss of semantics. Because there isn't enough diversity and heterogeneity in the amount of information produced, the OC-RPL model doesn't work as intended.

The Hierarchical Reinforcement Learning for Course Recommendation (HRLCR) model also doesn't perform as expected. For MOOC course selection, is a hierarchical reinforcement learning model. The historical course data is one of the main attributes which it considers and apart from this, the hierarchical reinforcement learning algorithm provides a strong learning ecosystem, but there is no substrate to act in the form of attenuated knowledge. It only acts on the traditional data, which makes it quite a deficit of knowledge and also there is no strong reasoning mechanism as compared to the proposed framework.

The Meta-relationship for course recommendation (MRCR) model also doesn't perform as expected, although it uses meta-relationships to enhance the expression of relational information. To mine and fuse the latent representations of users and courses as user preferences and course features, the MRCR model employs graph-based embedding, which surely contributes to the strong relevance computation mechanism. Meta-relationship correlation measure is used by the MRCR model to obtain semantic correlation information. However, there is a lack of sufficient knowledge to act upon and an absence of a strong learning model. Therefore, the MRCR model also lacks in comparison to the proposed model. The Collaborative filtering for Course Recommendation (CFCR) model uses collaborative filtering. Collaborative filtering requires the rating matrix to be computed in which the courses have to be rated. So, it is not always necessary the courses needed are extensively well, but for the particular data set it performs well. Given that collaborative filtering lacks an auxiliary knowledge base and a learning mechanism, it is clearly not the optimal platform. As a result, in comparison to the suggested model, the CFCR also lags.

To the lacuna of all the baseline models, there is the presence of a strong learning course system in the proposed SCRF in the form of an LSTM model to classify the financial course data set. Its knowledge aggregation model repositories like wiki data and also knowledge aggregation models like LSI, Bag of words and a strong knowledge stack comprising of eBooks, blogs, and web pages pertaining to finance and economics as a domain of choice there is a strong relevance computation ecosystem in form of semantic similarity, and statistical indices like SOC-PMI, NPMI, and Pianka index for relevance computation of the semantic similarity measure and step deviance measure. For the previously mentioned reasons, the proposed model outperforms all baseline models and forms, making it the best-in-class model. Fig.2. displays the precision in relation to the total number of suggestions made by the proposed SCRF using alternative baseline models as a plot.

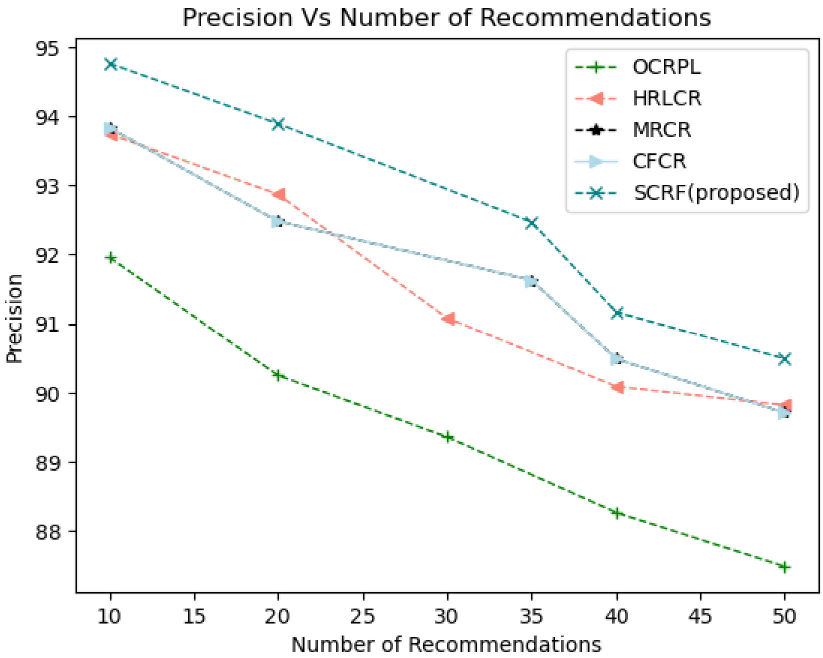

Fig. 2. Precision percentage vs. number of Recommendations of the Proposed SCRF with other baseline models

References

1. Joy, J., & Raj, N. S. (2019, December). An ontology model for content recommendation in personalized learning environment. In Proceedings of the second international conference on data science, e-learning and information systems (pp. 1-6).
2. Zhang, J., Hao, B., Chen, B., Li, C., Chen, H., & Sun, J. (2019, July). Hierarchical reinforcement learning for course recommendation in MOOCs. In Proceedings of the AAAI conference on artificial intelligence (Vol. 33, No. 01, pp. 435-442).
3. Hao, P., Li, Y., & Bai, C. (2023). Meta-relationship for course recommendation in MOOCs. Multimedia Systems, 29(1), 235-246.

Fi

4. Pang, Y., Jin, Y., Zhang, Y., & Zhu, T. (2017). Collaborative filtering recommendation for MOOC application. Computer Applications in Engineering Education, 25(1), 120-128.
5. Lin, Y., Feng, S., Lin, F., Zeng, W., Liu, Y., & Wu, P. (2021). Adaptive course recommendation in MOOCs. Knowledge-Based Systems, 224, 107085.
6. Gong, J., Wang, S., Wang, J., Feng, W., Peng, H., Tang, J., & Yu, P. S. (2020, July). Attentional graph convolutional networks for knowledge concept recommendation in moocs in a heterogeneous view. In Proceedings of the 43rd international ACM SIGIR conference on research and development in information retrieval (pp. 79-88).
7. Zhang, H., Huang, T., Lv, Z., Liu, S., & Zhou, Z. (2018). MCRS: A course recommendation system for MOOCs. Multimedia Tools and Applications, 77, 7051-7069.
8. Zhang, H., Shen, X., Yi, B., Wang, W., & Feng, Y. (2023). KGAN: Knowledge grouping aggregation network for course recommendation in MOOCs. Expert Systems with Applications, 211, 118344.
9. Tian, X., & Liu, F. (2021). Capacity tracing-enhanced course recommendation in MOOCs. IEEE Transactions on Learning Technologies, 14(3), 313-321.
10. Wang, J., Xie, H., Au, O. T. S., Zou, D., & Wang, F. L. (2020, August). Attention-based CNN for personalized course recommendations for MOOC learners. In 2020 International symposium on educational technology (ISET) (pp. 180-184). IEEE.
11. Ye, B., Mao, S., Hao, P., Chen, W., & Bai, C. (2021). Community enhanced course concept recommendation in MOOCs with multiple entities. In Knowledge Science, Engineering and Management: 14th International Conference, KSEM 2021, Tokyo, Japan, August 14–16, 2021, Proceedings, Part II 14 (pp. 279-293). Springer International Publishing.
12. Manoj, N., & Deepak, G. (2021). ODFWR: an ontology driven framework for web service recommendation. In Data Science and Security: Proceedings of IDSCS 2021 (pp. 150-158). Springer Singapore.
13. Kannan, B. D., & Deepak, G. (2022, December). I-DLMI: Web Image Recommendation Using Deep Learning and Machine Intelligence. In International Conference on Hybrid Intelligent Systems (pp. 270-280). Cham: Springer Nature Switzerland.
14. Yethindra, D. N., & Deepak, G. (2021, September). A semantic approach for fashion recommendation using logistic regression and ontologies. In 2021 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES) (pp. 1-6). IEEE.
15. Agrawal, D., & Deepak, G. (2022, January). HSIL: hybrid semantic infused learning approach for course recommendation. In International Conference on Digital Technologies and Applications (pp. 417-426). Cham: Springer International Publishing.
16. Adithya, V., & Deepak, G. (2021, July). HBlogRec: a hybridized cognitive knowledge scheme for blog recommendation infusing XGBoosting and semantic intelligence. In 2021 IEEE International conference on electronics, computing and communication technologies (CONECCT) (pp. 1-6). IEEE.
17. Krishnan, N., & Deepak, G. (2022, July). KnowCommerce: A Semantic Web Compliant Knowledge-driven Paradigm for Product Recommendation in E-commerce. In Congress on Intelligent Systems: Proceedings of CIS 2021, Volume 2 (pp. 37-45). Singapore: Springer Nature Singapore.