

Data Analysis and Visualization of Student Learning Patterns over web Materials

Pasula Kaushik Reddy, Marella Viswa Chaitanya, Veluru suhasnadh Reddy, Ramavath
Mohanlal Naik, Rajasekaran.R
School of Computer Science and Engineering
Vellore Institute of Technology

ABSTRACT

In this computerized time, world is honored with the greater part of the offices on a solitary snap and learning propensities are the same. Lately, we have seen that understudies are looking into concentrating from the learning materials accessible on the web. This transformation in study designs has made instructive establishments and colleges to consider the future showing procedures in training. In this paper, we are doing examination on the investigation example of understudies. Our spotlight has been more on the mechanism of concentrating, for example, YouTube recordings, Class notes, digital books and Textbooks. We have utilized information mining model with bunching method to sum up what level of understudies like web materials and homeroom condition to consider. This examination will foresee the fate of training mode in next 5 years. It likewise features the connection between study examples and understudies' scholarly accomplishments. This investigation can be utilized by the training organizations for their business methodologies in coming years.

Keywords: Kmode Clustering, Data Mining, Learning Patterns, Online Learning

INTRODUCTION

Learning is a word that characterizes the cycle by which an individual is growing new aptitudes and qualities by picking up information or altering the current one. As we are moving towards advanced strategies for carrying out each responsibility in our everyday life, so learning is no place lacking behind in it. Organizations are centering to comprehend the learning examples of their understudies so as to improve their learning framework. Learning designs relies on a person's getting a handle on aptitudes, for example, perusing, composing or viewing. There are significant four classes on which the students get generally arranged, for example, Visual in which students are quicker to learn through viewing the recordings or any pictorial portrayals of the data. Second classification is Aural in which students are quicker to learn through listening such meeting to generate new ideas. Third classification is Reading in which students are quicker to learn through understanding books, ppts or any content and composing the notes. Fourth class is Kinesthetic in which students are quicker to learn through doing commonsense things or accomplishing live work. Every one of these classifications assume significant parts in distinguishing the learning examples of understudies.

Instruction foundations are anticipating give the best learning framework to their understudies just as numerous new chances to their instructors for growing better educating abilities. To comprehend these learning designs individuals have utilized part of factual strategies yet in this paper, we have utilized AI methods. Measurable strategy for perceiving the example is a dull work and tedious also. AI procedures will utilize the information mining measure and order the example based on their favored methods. We have utilized solo learning calculation that is Kmode calculation in which we have arranged the learning designs in groups. Regularly alluded to as Emotion AI, Opinion mining has been turning into an unmistakable aspect of our new digitalized life. To characterize it in proper words it can expressed as the utilization of regular language handling, text investigation, computational etymology, and biometrics to deliberately distinguish, extricate, evaluate, and study emotional states and abstract data. To talk when all is said in done, sentiment mining intends to decide the demeanor of a speaker, essayist, or other subject regarding some theme or the general relevant extremity or passionate response to a report, collaboration, or occasion. The air may be a judgment or evaluation (see assessment theory), brimming with feeling state (as such, the energetic state of the maker or speaker), or the normal eager correspondence (at the end of the day, the excited effect arranged by the maker or examiner).

This method of extricating human feelings from advanced logs is as a rule generally applied to voice of the client materials, for example, audits and review reactions, on the web and web-based media, and medical services materials for applications that run from showcasing to client assistance to clinical medication. In our undertaking we have ventured forward to utilizing this method to separate the feeling of an understudy concerning an inquiry paper he/she endeavored by examining the remark segment of a review reaction that understudy filled to express his origination of trouble level of that paper as 'Simple' or 'Medium' or 'Extreme'.

RELATED WORK

Chengjiu Yin et.al (2017) [1] has explored different avenues regarding an advanced course reading framework to dissect the understudy's examination designs. They requested that understudies and educators read the computerized course book accessible on the application. Logs were getting caught in foundation of each movement done by the understudies, for example, "going to next page", "going to past page", "underlining the substance", "featuring the substance", "Putting bookmarker" and so on. They zeroed in on the recurrence of these exercises. They performed slack consecutive investigation on the information. Based on this investigation, they found that understudies like computerized sources all the more these days as correlation with traditional books for examines. Educators can utilize this framework to improve the examination materials and give more compelling learning system based on the understudy's investigation designs investigation.

H Awang et.al (2017) [2] has done an examination to discover the connection between learning styles and scholastic accomplishment of understudies. VARK learning model was utilized for the investigation of the information. V represents visual, A represents sound, R represents read-compose, K represents sensation. They needed to which method of study out of VARK is more ideal by the understudies. They contrasted the CGPA of the understudies with their inclination given to the method of study among VARK. They performed chi-square test on the information for the investigation. Albeit, no significant connection was found after the investigation between the scholastic accomplishments and best learning styles yet this examination can be used by the educators to improve the instructing and learning experience. Instructors can allude to the information and discover which mode is more ideal among the understudies and instruct as needed to build up their enthusiasm for the homeroom and course. R. Sridhar Anand and Dr. M. Rajendraprasad (2016) [3] has done a study in an elementary school to discover what is the ideal method of study style. The investigation depended on visual, sound and sensation learning styles. In their investigation, they analyze the effect of sex, school region and training of guardians on the learning styles. In their review, they found based on mean scores that the majority of the understudies favor visual examination style in the study hall and furthermore have better scholastic outcomes. They discovered huge effect of sexual orientation, school region and instruction of guardians on the educating and learning encounters of instructors and understudies. John Martin An and Maria Dominic M (2019) [4] has proposed a framework that catches the logs of student's exercises and adjusts their styles of learning. According to their ideal style, anticipate and suggest the suitable learning object. They have 3 modules in the framework, for example, Learner model in which student profile will be dealt with their ideal learning styles, Domain model depicts how the substance of the course is planned and student proposal model to order utilizing irregular woods calculation to prepare and discover the learning way for the forecast. This framework will likewise suggest the appropriate learning materials

Muhammad Said Hasibuan .et.al (2018) [5] proposed a framework in which the examination of learning styles depended on earlier information in which they have separated the degrees of earlier information into four, for example, information on certainty (zero level) in which understudy can just perceive the substance and name it, information on Meaning (first level) in which understudy perceives just as clarify it, Integration of information (second level) in which understudy can alter, reproduce and play out the substance, (last level) Application of information in which understudy can make something out of the information gain from the

substance. Idle Semantic Indexing and Artificial Neural Network was utilized to foresee the learning style of the student. Brahim HMEDNA et.al (2017) [6] investigates about the arrangement of ad libbing MOOCs (Massive Open Online Courses) adequacy. Utilizing the materials accessible on the web, Learners handle full substance information with superior exhibitions and less time. They utilized multi-layer neural system to distinguish and follow students learning styles and furthermore suggest them with reasonable assets and materials. Student's taking in styles is acquired from the student profile, at that point learning styles can be altered and followed by watching student's activities they act in a MOOC domain. It is done as Data Collection – Pre-preparing - Feature Extraction - Classification - Learner Profile-Adaptation (suggestion). Annabel Latham et.al (2014) [7] proposes another methodology for profiling understudy learning technique for a conversational smart coaching framework (CITS) which utilizes a Multilayers Perceptron Artificial Neural Network (MLPANN). All through a mechanized conversational instructional exercise with a CITS, parts of understudy conduct are progressively caught and contribution to a Learning Styles Predictor specialist to profile a person's learning style. CITS empowers students to cooperate utilizing normal language by its further degree of insight. A MLP-ANN is consolidated to join a lot of conduct qualities which are uncovered from the coaching discussion so as to improve the exactness of the learning styles expectation. The paper portrays the test directed with genuine understudies in a live educating/learning condition for profiling two Felder and Silverman learning styles measurements. The outcomes show that MLP-ANNs can foresee learning styles with an exactness of 84-89%.

Chengjiu YINa*, Hwang (2006) [8] has completed a few trials to gather the instructive information to improve and give viable learning condition just as the understudies conduct. By applying these numerous viable methodologies in social occasion the information, they need give the correct direction, proposals and supports from their picking up behavior. Talking about the commitment Hwang likewise proposed the viable utilization of Meta Analyzer that guided the educators how to anticipate the conduct of the understudies by web-looking through lead. The live models were done in Kyushu University where numerous numbers of sensors were fitted in the grounds so as to assemble and to dissect the student's conduct and furthermore to amass the greatest informational collections where every understudy are asked and urge understudies to bring their own PCs. They are accepting the static information as well as the understudy's movement logs caught camera just as video sensors. They likewise acquainted the advanced reading material with the freshers. OERs represents open instructive assets in like

manner there are such other OCW and Moocs which expect the student to get to the learning materials whenever and anyplace. There has been likewise the examination between the conventional training System and OERs which Clearly show the viability learning and empowering the gathering learning. There is finished clear forecast that by 2020, (MEXT) known as Ministry of training and innovation is supplanting all the course reading materials to computerized assets in Japan.

Lahcen Oughdir (2019) [9] plans to discover and anticipate the student's learning style immediately. His intention is to give successful and improved method of learning methodology. The learning style give client numerous choices to pick and along these lines they learn best. It was additionally seen that in customary framework has parcel more restrictions as it was requiring some investment to fill the inquiry paper and also due to the ignorance of the understudy's decision of inclination they don't get precise yield. Along these lines, in this paper they have proposed programmed approach for recommending the able learning styles for understudies utilizing Felder and Silverman style of learning Model by utilizing Machine learning. They dole out the student to those style of adapting way where they serenely learn. Their methodology comprises of two strategy: In first they are gathering the dataset utilizing information mining of the understudy dependent on their inclinations and from that log record they are breaking down the student's groupings. At that point they club them in sixteen bunches utilizing unaided K-mean calculation dependent on FSLSM and the further they anticipated the learning grouping of new forthcoming students by utilizing Supervised Naïve Bayes calculation. To concoct end and perception the informational index was all around extricated and examined from understudy's log documents. So as to assess the exhibition disarray framework strategy has been utilized. So as to accomplish well exact outcome, the model ought to have PPV, NPV for better execution. The Confusion grid utilized above utilizing R has given high worth measurements. Thus, we can say that the parts and all the calculation utilized has very much picked and has accomplished better execution. Later on scope the calculation utilized Naïve Bayes will be contrasted with its presentation with other AI techniques for instance Neural Network and other Decision tree. Sucheta V. Kolekar et.al. (2017) [10] proposed a strategy for the computerized acknowledgment and depiction of student learning styles utilizing Web Log Analysis approach. The web utilization information gathered is pre-handled and traded to XML organization to group and students' particular arrangements according to their meetings. Such arrangements are planned into eight Felder-Silverman Learning Style Model gatherings dependent on the learning objects recognized in every classification. The planning was finished by applying the Fuzzy C Means calculation to the

arrangements and gathering them into eight FLSM classes characterized as Active, Sensing, Reflective, Intuitive, Verbal, Sequential, Visual, and Global. A few arrangements are given various marks as indicated by the FLSM learning objects esteems. The proposed arrangement calculation is approved with the utilization of k-cross overlay function. Ouafae EL Aissaoui et al. (2019) [11] proposed a computerized approach for characterizing the learning examples of the students utilizing information mining strategies and ML calculations. The log record of the E-learning stage was pre-handled utilizing web-use mining procedure to recover arrangements of the students. Such groupings were planned to mixes of learning styles utilizing the bunching calculation K-modes and dependent on the FLSM. The named groupings have been utilized as preparing set to prepare the gullible Bayes classifier and conjecture another understudy's learning style blend. As the subsequent stage, the disarray network technique was utilized to assess execution of the classifier. Liang-Yi Li, and Chin-Chung Tsai (2017) [12] has done a nitty gritty investigation of framework logs delivered by a learning the board framework in which 59 software engineering understudies partook in a blended adapting course to learn cell phone programming. The outcomes uncovered various significant perceptions. Initially, the understudies analyzed the learning materials for more and more frequently than other learning materials related with their study hall addresses. Second, despite the fact that the understudies invested a great deal of energy getting to the learning materials on the web, the dominant part didn't utilize explanation assets. Third, the watching propensities for the understudies indicated incredible decent variety and were ordered into three personal conduct standards: "predictable use understudies" who utilized seriously all the learning materials, "slide concentrated use understudies" who utilized seriously the talk slides and "scarcely any utilization understudies" who utilized scarcely any learning material. Such different examples of activity were likewise identified with their inspiration and learning success. Sherica Lavinia Menezes and Geeta Varkey [13] proposed a structure which predicts the missing things taking into account the previous information and prescribes the equivalent to the customers. To fulfill this endeavor the structure used gathering frameworks going before the conjecture strategy. The upside of using portrayal/gathering is that the desire is done at a more raised measure of consultation and the expense of oversee age in connection direct mining is restricted. Out of the various decisions available, Naive Bayes classifier is chosen for request since this classifier will work splendidly for broad enlightening records and in modestly simple to complete. Different leveled gathering instrument is chosen for batching. The data extraction is done using a robotized data extraction instrument web Harvest. The data structure chose for graphs is a Hash List which is a blend of the hash table and associated once-over and is gave off an impression of being a

successful data structure for addressing a diagram. J. Peruse [14] used Twitter spouting data gave by Firehouse, which gave all messages from every customer persistently. They inspected with brisk gradual methodologies that were a great idea to oversee data streams: stochastic tendency dive, multinomial guiltless Bayes and the Hoeffding tree. Thusly they derived that SGD-based model, used with a sensible learning rate was the best. Ruchi Mehra, Mandeep Kaur Bedi [15] introduced investigation for supposition direct of Twitter data. The proposed work utilizes the sincere Bayes and cushy Classifier to arrange Tweets into helpful, opposite or neural lead of a particular person. They presented preliminary evaluation of dataset and portrayal comes about which exhibited that joined proposed strategy is more compelling similar to Accuracy, Precision and Recall.

Theresa Wilson, Janyee Wiebe et al., [16] proposed another approach to manage state level evaluation examination can be that at first chooses if an enunciation is fair or polar and thereafter disambiguates the limit of the polar explanations. With this methodology, the structure can therefore perceive the important furthest point for a broad subset of feeling verbalizations, achieving comes about that are generally better than pattern. Simranjeet kour Bindra, Akshay Girdhar et al., [17] examined that the inquiry paper age is a manual methodology inciting insufficient from time to time owing to inclination, excess and security concerns. The current paper shows a modified strategy for question paper assemble which can be balanced, smoothed out, synchronized and made sure about. Every task done by this structure is modified, with the ultimate objective that taking care of room, tendency and security isn't a dread any more. Earlier, the request paper was made by concerned subject teacher truly and was incredibly repetitive, work was required and sometimes the request paper required exactness. Result Based Education (OBE) allots what understudies will know and be insightful to do, as they progress in a program.

Dominic Seyler et al., [18] proposed an Automated inquiry age for quality control in human estimation tasks. The issue of creating question things from ontologies has starting late expanded a lot of thought in the product building gathering. This is basically a direct result of the utility of the made requests in various informational and master works out, for instance, understudy assessments in eLearning systems, quality control in human computational tasks and, blackmail area in publicly supporting stages to give a few models. Tahani Alsubait et al., [19] proposed an Ontology-based numerous decision question age. A Traditionally, question age (QG) approaches have by and large focused on recouping requests from unrefined substance, data sets and other non-semantics-based data sources. Regardless, since these

sources don't get the semantics of the space of talk, the created questions can't be machine-dealt with, making them less employable in an enormous number of this current reality applications. For example, addresses that are delivered from unrefined substance are sensible only for lingo learning tasks. Using semantics-based data sources in QG has various focal points, for instance, in ontologies, we show the semantic associations between zone substances, which help in delivering noteworthy and machine-processable requests ontologies engage standard reasoning and addressing organizations over the getting the hang of, giving a framework to making addresses even more viably. An early push to recognize factors that might foresee the difficulty level was by Seyler et al., They have familiar a strategy with bunch a request as straightforward or hard by finding the features of the tantamount request components in the Linked Open Data (LOD). Feature regards for the portrayal task are gotten considering the accessibility of the request substances in the LOD.

Xinming An et al., [20] proposed an Item reaction hypothesis. This speculation was first proposed in the field of psychometrics, later, the theory was used extensively in informative exploration to alter and evaluate questions things in the general assessments, for instance, the Scholastic Aptitude Test (SAT) and Graduate Record Examination (GRE).

K-Mode Clustering

Bunching is a strategy for making gatherings of comparable articles based on the similitude between their highlights. K-mode calculation is one of the grouping techniques that is an altered form of K-mean calculation explicitly for the unmitigated information. This calculation can be utilization of enormous datasets. Modes are utilized to speak to the focal point of the groups. We update the estimations of mode according to the recurrence of the straight out qualities. This is an iterative strategy for bunching wherein modes gets update in every emphasis so as to cover each insignificant information in the set. Keeping up the effectiveness of the apparent multitude of bunches ought to be the most extreme earlier.

Steps needs to follow in this algorithm:

- ✓ Select the number of K clusters as the initial value of mode. We can take it as random.
- ✓ Calculate the minimum deviation from the mode value and update the mode accordingly. Repeat this step for the entire data till the time each data gets assigned to any cluster.
- ✓ Compare the value of new mode with older one. In case of different values, repeat step 2.



Data Collection

We created a questionnaire and circulated amongst the students to gather the information about their studies patterns. After collecting the response, we created a balanced dataset and apply the model on it to get the prediction as below:

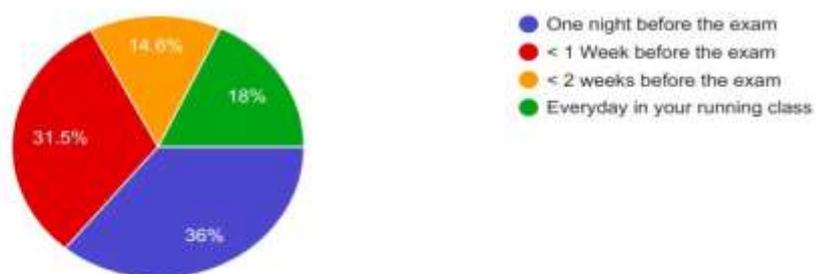
QUESTIONNAIRE

Converted the questionnaire options from text to numbers to make algorithm works simple as per below:

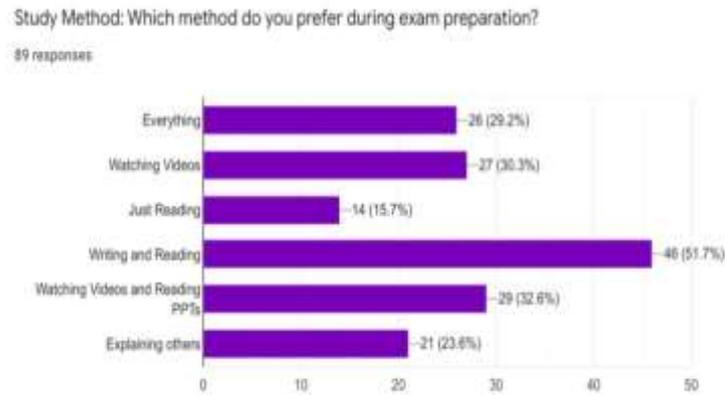
LABEL	S	A	B	C
CGPA	Greater than 9	Between 8.5 to 9	Between 8.0 and 8.5	Less than 8
Branch	Computer Science	ECE, EEE and EI	Bio Technology	Mechanical
Study hours before exams	One night before exam	2-3 days before exam	Week before exam	Every day in running class
Study Place	Classroom	Home/Hostel	Library	
Study Medium	Class notes	Books	PPTs	Youtube Videos
Study Method	Just Reading	Reading and Writing	Watching Videos/Reading PPTs	Everything
Avg Score	40-50	50-60	60-75	More than 75

How much time before do you start your exam preparation?

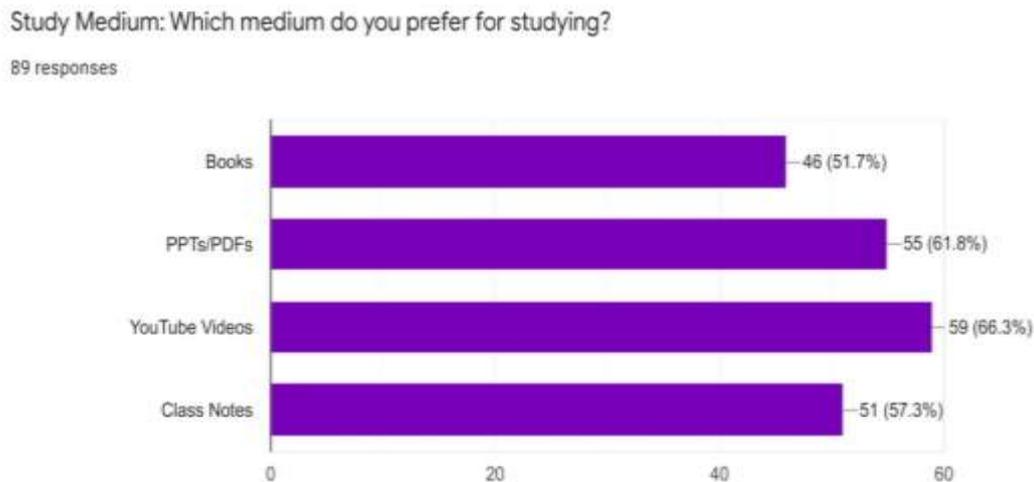
89 responses



Study Method: Which method do you prefer during exam preparation?



Study Medium: Which medium do you prefer for studying?



The majority of the understudies were found to begin the test readiness only one night before the assessment. The least understudies liked to plan in typical running classes. Contemplating over Study techniques, most understudies were found to peruse, compose and practice for their test arrangement while the least were just for simply perusing. On the perception made for study medium, the most favoured medium was by viewing YouTube recordings while the least favoured books. By this review, it is clear that most understudies want to take benefits with computerized stage than going to books.

IMPLEMENTATION AND RESULTS

Dataset has been collected through the questionnaire that was circulated among the Engineering students age between 19 – 23 year of a University.

CGPA	Branch	Internal_Exams	Study_hours_before_exam	Avg_hours_for_Midsem	Study_place	Study_medium	Study_method	Study_hours_before_final_exam
3	2	2	4	1	2	4	3	4
1	1	2	2	1	2	4	3	4
1	1	2	1	1	2	3	2	4
1	1	2	5	1	2	1	1	3
1	4	2	3	2	2	3	3	1
2	1	2	1	1	2	3	1	2
1	4	2	1	1	2	4	3	2
2	1	2	3	1	2	3	2	4
1	1	2	4	1	2	2	2	4
1	1	2	3	1	2	3	1	4

After applying Kmode clustering model on the given data, we found that Computer Science students more prefer to study through PPTs in hostel and score 9 CGPA.

CGPA	Branch	Study_place	Study_medium	Study_method	Difficulty_Level
0	1	1	2	3	3

Electrical and Electronics branch students prefer more to study from hostel/home through youtube videos and like to read and writing.

CGPA	Branch	Study_place	Study_medium	Study_method
0	1	2	2	4

Bio Science students prefer studying from hostel/home through reading books and like to have reading- writing.

CGPA	Branch	Study_place	Study_medium	Study_method
0	4	3	2	2

Mechanical Students prefer to have YouTube Videos and reading writing as preferable method and score more than 9 CGPA.

CGPA	Branch	Study_place	Study_medium	Study_method
0	1	4	2	4

CONCLUSION

A study pattern analysis was performed on a sample of 200 UG students within the campus using K-mode clustering algorithm. The performance characteristics of the proposed algorithm were quantified in terms of study time for the preparation of the examination, study medium and method of study. It was noticed that among 200, there were only 89 responses from different students. Statistics show that the 36% of students are studying for the exam overnight using writing and reading method on YouTube platform.

REFERENCES

- [1] Chengjiu YIN, Noriko UOSAKI, Hui-Chun CHU, Gwo-Jen HWANG, Gi-Zen Liu, Jau-Jian HWANG, Itsuo HATONO, Etsuko KUMAMOTO & Yoshiyuki TABATA “Learning Behavioral Pattern Analysis based on Students’ Logs in Reading Digital Books” International Conference on Computers in Education. New Zealand: Asia-Pacific Society for Computers in Education (2017)
- [2] H Awang, N Abd Samad, N S Mohd Faiz, R Roddin and J D Kankia “Relationship between the Learning Styles Preferences and Academic Achievement” International Research and Innovation Summit (IRIS2017) IOP Conf. Series: Materials Science and Engineering 226 (2017)
- [3] R.Sridhar Anand and Dr.M.Rajendraprasad “ A study on learning styles and their impact of science teaching among primary school students International Education and research journal volume 2 Issue 1 Jan 2016
- [4] John martin A, \Maria Dominic M “Adaptation using Machine Learning for Personalized eLearning Environment based on Students Preference” International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-10, August 2019
- [5] Muhammad Said Hasibuan, Lukito Edi Nugroho, Paulus Insap Santosa “MODEL DETECTING LEARNING STYLES WITH ARTIFICIAL NEURAL NETWORK” Journal of Technology and Science Education JOTSE, 2019 – 9(1): 85-95 – Online ISSN: 2013-6374 – Print ISSN: 2014-5349 November 2018
- [6] Brahim HMEDNA, Ali El Mezouary, Omar Baz, and Driss Mammass “Identifying and tracking learning styles in MOOCs: A neural networks approach “International Journal of

Innovation and Applied Studies ISSN 2028-9324 Vol. 19 No. 2 Feb. 2017, pp. 267-275 © 2017
Innovative Space of Scientific Research Journals

[7] Annabel Latham, Keeley Crockett, David Mclean “Profiling Student Learning Styles with Multilayer Perceptron Neural Networks” IEEE Xplore 27 January 2014

[8] Yin, C., Uosaki, N., Chu, H. C., Hwang, G. J., Hwang, J. J., Hatono, I., & Tabata, Y. (2017, January). Learning behavioral pattern analysis based on students' logs in reading digital books. In *Proceedings of the 25th international conference on computers in education* (pp. 549-557).

[9] Awang, H., Samad, N. A., Faiz, N. M., Roddin, R., & Kankia, J. D. (2017, August). Relationship between the Learning Styles Preferences and Academic Achievement. In *IOP Conference Series: Materials Science and Engineering* (Vol. 226, No. 1, p. 012193). IOP Publishing.

[10] Prasad, M. R., & Reddy, D. K. (2016, February). Computer Based Teaching Methodology for Outcome-Based Engineering Education. In *2016 IEEE 6th International Conference on Advanced Computing (IACC)* (pp. 809-814). IEEE.

[11] Hasibuan, M. S., Nugroho, L. E., & Santosa, P. I. (2019). Model Detecting Learning Styles with Artificial Neural Network. *Journal of Technology and Science Education*, 9(1), 85-95.

John Martin A and Maria Dominic M (2019). Adaptation using Machine Learning for Personalized Elearning Environment based on Students Preference. (2019). *International Journal of Innovative Technology and Exploring Engineering*, 8(10), 4064–4069.

[12] Kolekar, S. V., Pai, R. M., & MM, M. P. (2017). Prediction of learner's profile based on learning styles in adaptive e-learning system. *International Journal of Emerging Technologies in Learning (iJET)*, 12(06), 31-51.

[13] El Aissaoui, O., El Madani, Y. E. A., Oughdir, L., & El Alloui, Y. (2019). A fuzzy classification approach for learning style prediction based on web mining technique in e-learning environments. *Education and Information Technologies*, 24(3), 1943-1959.

[14] Li, L. Y., & Tsai, C. C. (2017). Accessing online learning material: Quantitative behavior patterns and their effects on motivation and learning performance. *Computers & Education*, 114, 286-297.

[15] Simranjeet kour Bindra, Akshay Girdhar and Indrajeet Singh Bamrah “Outcome Based Predictive Analysis of Automatic Question Paper using Data Mining”, Proceedings of the 2nd International Conference on Communication and Electronics systems ICCES 2017.

- [16] Go, R. Bhayani and L. Huang, “ Twitter Sentiment Classification Using Distant Supervision”, Stanford University, Technical paper, 2009
- [17] Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends® in Information Retrieval*, 2(1–2), 1-135.
- [18] Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, 5(1), 1-167.
- [19] Lowd, D., & Domingos, P. (2005, August). Naive Bayes models for probability estimation. In *Proceedings of the 22nd international conference on Machine learning* (pp. 529-536). ACM.
- [20] Wilson, T., Wiebe, J., & Hoffmann, P. (2005, October). Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of the conference on human language technology and empirical methods in natural language processing* (pp. 347-354). Association for Computational Linguistics.