Turkish Journal of Computer and Mathematics Education *Vol.12 No.7 (2021), 2305-2325*
ResearchArticle

Personalized E-Learning on Social Web with Machine Learning

[1]Vedavathi Assistant Professor [2]Dr Anil Kumar K M Associate Professor
[1]vedavathiresearch@gmail.com, [2]anilkm@sjce.ac.in
[1]NIE Institute of Technology, Mysuru Department of Computer Science., [2]Sri Jayachamarajendra College of Engineering, Mysuru Department of Computer Science

Article History: Received: 11 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online: 16 April 2021

ABSTRACT
Engaged User has been identified as a potential source of knowledge about personal interests, preferences, goals, and other attributes known from user models. As users are free to like the posts available on facebook such as photo and videos, status and links of their own interest, machine learning techniques are applied to analyze and predict that how many users engaged mostly with photos, videos, status and links. The main objective of this study is to analyze the Interest of engaged users on social web, Analyze and predict the engaged users on a particular post over the lifetime and analyzing and predicting facebook users personalized learning. In this work, we have proposed three machine learning algorithms to describe the users likes, shares and comments on posts of the facebook. In this article, we are proposing a new technology for classifying the users engaged in facebook based on Particle Swarm Optimization (PSO). The proposed machine learning algorithms shows that the performance of content-based engaged users is more reliable semantically for applied on personalized social web.

Index Terms: E- Learning Personalized Recommendation, Machine Learning, Engaged Users

1. Introduction

As the Social Web is growing rapidly, users are generally no longer able to obtain most of the info online due to overloading of information. As a result, customized recommended systems have evolved on the Social Web that filter irrelevant online information automatically, and provide customized user recommendations. The recommended systems are typically built on the basis of the user's personal preferences or interests for building user profiles. The user profiles are used for personalized reclassification in which objects are reclassified according to the individual user profile to higher rankings that represent the personal interests of the user. In specific, user interests or preferences may be derived by authorized information which is either implicit or explicit . Recommendation systems are using explicit data which allow users to make additional effort to fill out preferential details or to give negative or positive feedback on the outcome to the evaluation. Consequently, user profile is highly dependent on the ability to provide adequate and suitable preference data. The outcomes of these guidelines can be extremely unpredictable. An alternative of "effortless" user interaction data including query history, browsing history , current user tasks or intentions, as well as search monitoring is used in recommendation systems on modern websites and apps such as Facebook, Amazon, Google and YouTube; as complementary information about users. However, these data are generally not available for outside parties due to commercial or legal reasons, ethical, privacy, etc. Fortunately, users are now able to freely like the post, write the comments and post social annotations via the social network online. Such social annotations for custom social web recommendations are suitable knowledge. It is thus important to build customized systems for user and public data that are easily

accessible on the social network. However social annotations are perfect user data for custom social network recommendations for the following reasons:

- Social annotations are usually available online and can typically be easily accessed.

In this study we are using machine learning algorithms to resolve the issues of the engaged user's and achieve good recommendation based on social network interactions. In order to achieve good efficiency of recommendations and solutions for machine learning algorithms, we have to be train the system with large amounts of data and this method should be used regularly to capture the complex Web changes. Consequently, machine learning techniques can be prohibited with very few customers in lightweight recommendation systems.

2. Related Work

(Sunny Sharma et al. 2020) Analyze that the personalization of web searches is a process to provide the user with personalized search results. This paper includes a model of relevance for personalizing search results based on personalizing the query. If the similarity is found between the initial user query and the user profile a linear preference space combination is created in the runtime to more accurately evaluate which pages are actually the most relevant in relation to the updated query. To retain the user profile based on the ongoing actions a heuristic algorithm is used. And tests show that search results can be retrieved based on query change to provide the user with customized results [12].

(Amal Al-Abri et al. 2019) Analyze that learners from the same class can carry out the task with different applications through discussions and conversations. This article provides a collection model of machine learning techniques for the conversation data obtained from different applications in social media. Based on the attributes required to improve personalization services, this aggregation is built. The defined attributes were used for the development of a unified chat interface using various social network applications. The similitude technique using ontological model was implemented to accomplish this aggregating function. The positive results of the matching method indicate the utility of the model [13].

(GinaGeorge et al. 2019) Analyze that huge open-line courses and apprenticeship management systems have expanded the amount of learning opportunities accessible online. Under this sense, the recommendation for tailored services has become an even greater challenge to increase work in this direction. Recommend systems use ontology and artificial intelligence to render individualized suggestions and machine learning techniques. Ontology is a way, among other things, of modeling learning resources and of finding information. The comprehensive survey in this article offers an overview of ongoing work using ontology in order to individualize structures in the area of e-learning [14].

(Po-Sen Huang et al. 2019) Analyze that Semantic short text similarity is a key technology for natural language search. It is commonly used to find unknown information in social network analyzes and opinion mine with machine learning techniques. These measures typically weigh 10-20 words in short texts. Compared to spoken words, short texts do not always obey formal rules of grammar. Similarity steps are made difficult by minimal knowledge in short texts and their syntactic and semantic versatility. Therefore, a part-of - speech similarity algorithm was developed and tested in this study to solve these problems. This study analyzes the consequences of different parts

of the voice. With the word measurements corresponding to different sections of the expression, the algorithm given has achieved the best output [15].

(Jingyun Wang et al. 2019) Analyze that two ontology-driven learning frameworks aimed at creating a realistic learning environment with machine learning techniques. These include a personalized language-learning platform (CLLSS) and an e-book user support system for visualization learning (VSSE). CLLSS was designed to provide an interface to the arrangement of learning objects, showing the visual representation and relation of information. In other words, VSSE offers two learning modes: where all relationship knowledge is shielded from the first committed user of learning and in the second committed user learner is encouraged to actively establish this knowledge [16].

(Mohammed E. Ibrahim et al. 2018) Analyze that a novel approach that customizes course suggestions to suit users' individual needs. This way students can gain a detailed understanding from courses based on their relevance and can associate student profiles and profiles with work profiles through dynamic ontological mapping with machine learning techniques. Results show that a hierarchically linked-up filtering algorithm delivers better results than a filtering process that only takes into consideration keyword similarity. The method is versatile and can be applied to various domains using dynamic ontology mapping. The system proposed can be used to screen post-graduate and other domain items [17].

(John K. Tarus et al. 2018) Analyze that to order to assist learners to seeking useful and appropriate materials for their learning needs, the e-learning program is important to suggest with machine learning techniques. Customized intelligent agents and recommendation systems have been generally recognized as approaches to the problems of knowledge processing by information overload learners. Here, review and classify journal articles in the field of ontology-based e-learning guidelines published between 2005 and 2014. This study shows that ontology will boost the consistency of the recommendations in e-learning systems to facilitate information representation [18].

(Francisco García-Sánchez et al. 2018) analyze that at the beginning of the Web 2.0 era, a new source of huge user data is available. Recommendation advertising systems can also be used to enhance user's awareness of the needs and desires of such applications with use of machine learning techniques. But new problems are posed by the need to work with diverse data from different sources. Semantic techniques in general and ontology in particular have proven to be effective for information management and data integration. In this article, a recommendation system for ontology-based publicity is proposed that leverages users ' data in social networking sites [19].

(Ozge Sürer et al. 2017) Analyze that the representation of structural data is important in order to capture the pattern between characteristics with machine learning techniques. Besides the standard variables, there is information on interrelationships between variables. In this study, ontology information in systems can be used to suggest predictions that make it more effective. They suggest two alternative tests for similarity, namely structural data representation. Experiments have shown that ontology approach increases classification precision with rising dimensions [20].

(Zixian Zhanget al. 2019) Analyze that the meaning of another word exists and both words may be said to be connected. A graph-based ontology is founded on this form of information. Nodes are terms in the graph, and if there is one word in the context of the other, then there is one line in the graph with machine learning techniques. In order to measure word similarity, the line or degree as it is called. The word similarity is then used to measure the similarity of the sentence. In the text, the word material is used in the context of machine sentence comparisons, such as substantives, verbs, adjectives and adverbs. This measured sentence similitude is efficient for the processing of natural language tasks such as question response, data extraction, etc [21].

3. Problem Statement

Our paper focuses on personalization of socially active users, but we can easily apply our methods to other social annotations, such as remarks and blog posts. Usually, social users have been commonly used to make personalized social network recommendations for active users based on content-driven filters or shared filters. To this end, content-based filters are especially relevant and increasingly attract researchers' interest as it allows systems to provide interested users additional information on contents in order to enhance recommendations. With respect to the content-based personalized users, a similarity measure is necessary in order to estimate the pertinence of a user's preferences and to estimate the social summary of documentation where both profiles can be seen in vector space models as weighted vectors of engaged users. Precisely by aggregating all the socially committed users allocated by the user to the online records, a users Profile is usually achieved; it is represented as a weighted vector of committed users, whereby each dimension is the same as the committed user used by this user, and the value of each dimension is the weight of the actual dedicated user, depending on the number of times.

4. Objective

- Analyze the Interest of engaged user on social web
- Analyze and predict the engaged users on a particular post over the lifetime.
- Analyze positive relationships explicitly between Lifetime Engaged Users and likes/shares and comments.
- Analyzing and Predicting Facebook Users' Personalized Learning.

5. Methodology

In order to determine the personalized engaged users on social web with machine learning, we analyze how the selected matching term, chosen by a specific committed user assignment strategy, pertains semantically to the profile context of such committed users so that, the more applicable the chosen matching term is for the profile context. For the selected matching concept the considered semantically relevant context of the profile for a certain committed user on facebook pages are provided that this matching concept has average semantic distance to the corresponding concepts of all other engaged users and more semantically matching concept is relevant to the post.

Therefore, a corresponding term is selected by an engaged user allocation strategy for each user in the specified user's or document profile. In addition, since the first step to quantify the personalized E-Learning is to analyze the important features of the dataset and to improve the output by using the PSO algorithm, the engaged users' allocation is very frequent in ontology-based use. The outcomes of the two approaches are evaluated on the basis of a learning algorithm. The goal in this work is actually to propose the top-down user management strategy in the present strategy [22] to resolve the high computational complexity problem. This section addresses the dedicated user distribution, the design and implementation process of Facebook app data sets for the machine learning classification model [23]

5.1. Machine Learning

Machine learning (ML) is a computer algorithm analysis that automatically develops by means of practice. The artificial intelligence is viewed as a subset of it. In order to make predictions or decisions without being programmed directly, machine learning algorithms construct a mathematical model based on sampling data known as "workout data.' In a variety of applications machine learning algorithms such as email filtering and computer vision are used, where traditional algorithms are difficult or unworkable to carry out necessary tasks [24].

Machine learning is closely related to the use of machine statistics to make predictions. Methods, theory and implementations for the machine-learning field are studied on mathematical optimization. Data mining is a related field of study which focuses on analyzing data through unmonitored learning. Machine learning is also known as predictive analytics in its application to market problems [25].

This section has discussed the software packages, libraries, environments, and hardware requirements that are used as a part of this research work. Python is an effective programming language for the creation of a deep recurrent model of neural network. It supports In addition, a version of Python 3.7.1 used for this study is used. In the code creation process, Anaconda jupyter notebook python IDE is used; it is intended for operations in data analysis. A variety of science libraries, including Pandas, Numpy, Scipy, Miplotlib, sklearn and more, are available at Jupyter Notebook. It also provides advanced analysis, debugging, and editing capabilities in many apps, ("Jupyter Notebook: Anaconda Cloud"). There are type of machine learning algorithm used here such as

5.1.1. Random Forest

Random Forest is an algorithm for machine learning using a bagging method that generates a bunch of decision-making areas with a random database. One model has been trained on a random data set sample many times to achieve strong forest algorithm prediction performance [26]. The output of all the random forest decision-makers is combined to generate the final prediction by this ensemble method of learning. The final forecast of the random forest algorithm is extracted from the results of any decision-tab or from a prediction most commonly found in the decision-making bodies [25].

5.1.2. Lasso Algorithm

Lasso is a tool for regression analytics (the least absolute shrinking and selection operator is also Lasso or LASSO), in order to improve the prediction accuracy and interpretability of the statistical model it produces [28], by performing both variable selections and regularizations in statistics and machine learning. Initially introduced to geophysics literature in 1986, Robert Tibshirani, who coined up the word and offered additional insights into the results observed, rediscovered it and later became independently widely used in 1996 [29].

Lasso was implemented initially in the least squares sense and it may be instructive to take this case into consideration first, because it shows many of the characteristics of lasso in a simple setting (30).

5.1.3. Support Vector Machine

"Support Vector Machine" (SVM) is a supervised machine learning algorithm that can be used both for classification and regression challenges. This is primarily used for the problems of classification. In the SVM algorithm, each data object is drawn as a point in n-dimensional spatial space, with each characteristic as its value. We then distinguish by finding the very well separated hyper plane between the two groups [31]. Support vectors are data points to the hyper plane and impact the hyper plane direction and orientation. We maximize the margin of the classifier with these support vectors. The removal of support vectors would shift the hyper plane location. These are the points which help us to improve our SVM [32].

5.2. Particle Swarm Optimization

In computational science, particle swarm optimization (PSO) is a computational method that optimizes the issue by iteratively trying to improve the candidate's solution with quality measurement [34]. This solves the issue by providing a number of candidates solution, known here as particle, which are pushed in the search room by simply

using mathematical formulas over the position and velocity of the particles [35]. The motion of and particle is influenced by its well-known local location but is also directed towards the best-known search space positions, which are modified in a way that other particles find better positions. It will push the swarm to the best solutions [36].

5.2.1. Pseudocode

Let S be the number of particles, each with in the swarm $xi \in \mathbb{R}n$ in the search-space and a velocity $vi \in \mathbb{R}n$. Let pi be the best known particle position i and g would be the best known swarm position. There is a simple PSO algorithm:

for each particle $i=1,...,S$ do

Initialize the position of the particle with such a uniform random vector: $xi \sim (blo, bup)$

Preprocess the best known position of the particle to its original position: $pi \leftarrow xi$

if $f(pi) < f(g)$ then

update the swarm's best known position: $g \leftarrow pi$

Initialize the particle's velocity: $vi \sim (-|bup-blo|, |bup-blo|)$

When a termination condition is satisfied:

for each particle $i=1,...,S$ do

for each dimension $d=1,...,n$ do

Pick random numbers: $rp, \sim U(0,1)$

Update the particle's velocity: $vi, d \leftarrow \omega vi, d + (pi,d - xi,d) + \varphi grg(gd - xi,d)$

Update the particle's position: $xi \leftarrow xi + vi$

if $f(xi) < f(pi)$ then

Update the particle's best known position: $pi \leftarrow xi$

if $(pi) < f(g)$ then

update the swarm's best known position: $g \leftarrow pi$ [37]

The blo and bup values are the bottom and top limits of the search-space [38]. The criteria for termination may be the number of iterations performed or a solution for the right objective function value. The practitioner chooses parameters ±, ćp and μg and controls the behavior and effectiveness of the PSO method [39]

The function that will be used in used in this notebook is:

$(x) = (sen(10\pi x)) + 1$

Where $[-1<=x<=2]$.

We try to maximize the (x), where x is limited between 2 and −1 [40].

5.3. Dataset and Features

The data set and experimental context for calculate the relationships most explicitly between Lifetime Engaged Users and likes/shares, and less so for comments. More precisely, and provided that each modality used is correlated with a facebook users would be based personalized learning. Finally, here analyze each experiment's findings and present a qualitative results review. Here we use facebook user data for personalized learning. In the data variables available are page total likes, type, category, post weakly, post monthly, post hourly, paid, like, share, comment, total iterations are around 500 as shown in Figure.1. Since we must introduce a character-level model, all the lines of our dataset will be divided into character lists. The lower the frequency value, the higher the char (among the data set) the more frequently. Here need to remove null values, to get the best output. Main aim is to predict total interactions based on features such as comment, like, and share and instead focused on Total Interactions, which is for modeling. An outlier is easily visible for total interactions, at around 6000.

In the facebook user dataset, need to remove null values and then identify the number of users for likes and share and comment. To train and test our models, we used a publicly available Kaggle for facebook user dataset. The data set is for facebook user and consists of well over 500 examples with 19 features categorized as follows:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 500 entries, 0 to 499
Data columns (total 19 columns):
Page total likes                                                  500 non-null int64
Type                                                              500 non-null object
Category                                                          500 non-null int64
Post Month                                                        500 non-null int64
Post Weekday                                                      500 non-null int64
Post Hour                                                         500 non-null int64
Paid                                                              499 non-null float64
Lifetime Post Total Reach                                         500 non-null int64
Lifetime Post Total Impressions                                   500 non-null int64
Lifetime Engaged Users                                            500 non-null int64
Lifetime Post Consumers                                           500 non-null int64
Lifetime Post Consumptions                                        500 non-null int64
Lifetime Post Impressions by people who have liked your Page      500 non-null int64
Lifetime Post reach by people who like your Page                  500 non-null int64
Lifetime People who have liked your Page and engaged with your post  500 non-null int64
comment                                                           500 non-null int64
like                                                              499 non-null float64
share                                                             496 non-null float64
Total Interactions                                                500 non-null int64
dtypes: float64(3), int64(15), object(1)
memory usage: 74.3+ KB
```

Figure.1. Data Features

The first move is to test the dataset. While most of the datasets were complete, some data were missing. Features such as like, share and comment were measured in case of scheduled and actual check-out and check-out through monthly, hourly and weakly. The missing values were difficult to quantify for features such as comment and like and thus examples are omitted from our data collection for missing values.

5.4. Data Pre-processing

The program will read in a single text file that contains all the features such as like, comment and share. We need to fill the values that are blank with 0. Main aim is to predict Total Interactions based on features. I have excluded comment, like, and share and instead focused on Total Interactions, which is what we will be modeling for. An outlier is easily visible for total interactions, at around 6000. The next step is to preprocess the dataset.

After getting all necessary parameters from the facebook, we have done some preprocessing operations on the data to make them compatible for the sentiment analysis. They were also used for pre-processing to a certain degree in most of the research papers. A small number of scientists used hopping, deletion of words and spelling. A growing number of the uses of stemming, elimination of words, indexing of text, reduced dimensionality and weighting of expressions. For pre-processing sentiment analysis alone, these approaches cannot be used. For example, the accuracy of the sentiment analysis will decrease if we stop deleting words. Few scientists used tokenization. To make a sentiment study, we cannot tokenize the sentences. What we should do, then, is to isolate the sentences in one paragraph as much as possible.

6. Experimental results and analysis

In this evaluation, we divide the entire feature set by Facebook activity type into different categories. It is intended to assess the number of users who apply for shares or comments between shares and subscribers. Every experiment was performed on the same computer with 1.6 GHz processor and 8 G memory configurations. A 10-fold cross validation was conducted on the same dataset as previously used. Evaluation was performed for all study classifiers and three algorithms were employed for training and testing or model assessment. Various studies have been performed in the fields of like, share and comment data collection. The parameters of the dataset are shown in Figure.1.

6.1. Summary of findings

This study shows that facebook users mainly prefers to like the posts such as photos, videos, links and status in comparison of comments and share. Facebook is mainly an internet site and only a platform to post certain posts that is charged for secondary, free for others and mainly users prefer free content. In particular, certain features such as likes, comments and sections differ in correlation coefficients. For example, we have found lots of difference of engaged user that like a post instead of comment and share and some user does not prefer paid post. These findings suggest that users are much satisfied or users prefer to like the post instead of comment and share. We also analyzed how users correlated similar objects, such as paying or free, between observational data (i.e. images, videos, shares and links). It is important to note that the position of the user in a picture, video, share and connect is different. The disparity between images, shares, videos and links in which posts are included in photographs, sharing, video and links purposely and accidentally created incredible results. It is natural to believe that only devoted users will view images, share, videos, or links in which the post is unintended. However, the research found that good images, upload, videos and links unintentionally posted are linked to paid or free postings. The application of machine learning techniques produced good results with 0.58 to 100 % accuracy.

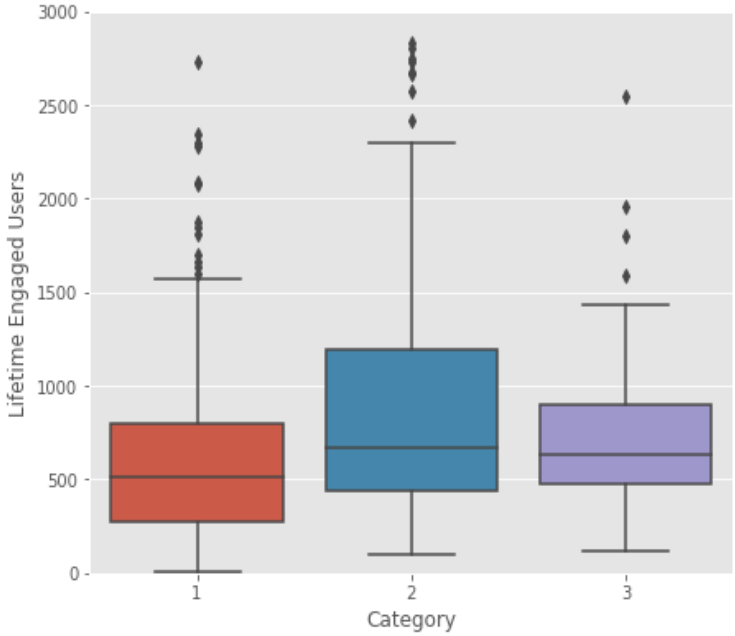

Figure.2.Category

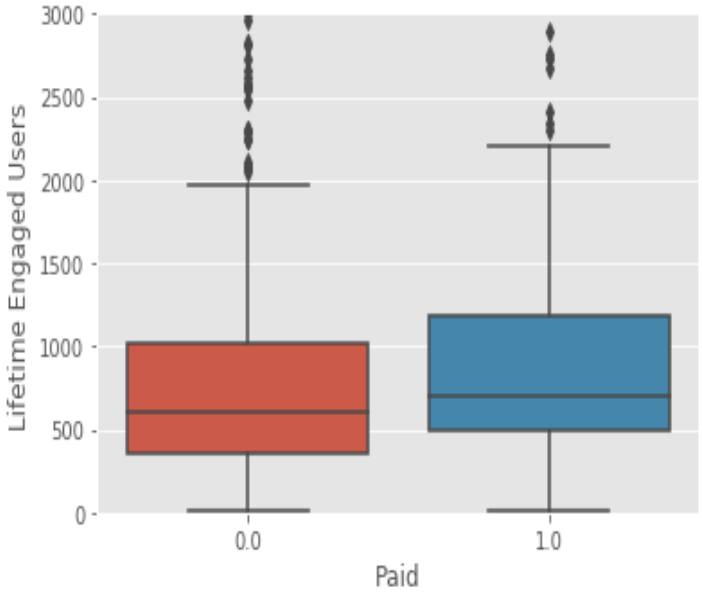

Figure.3.Paid

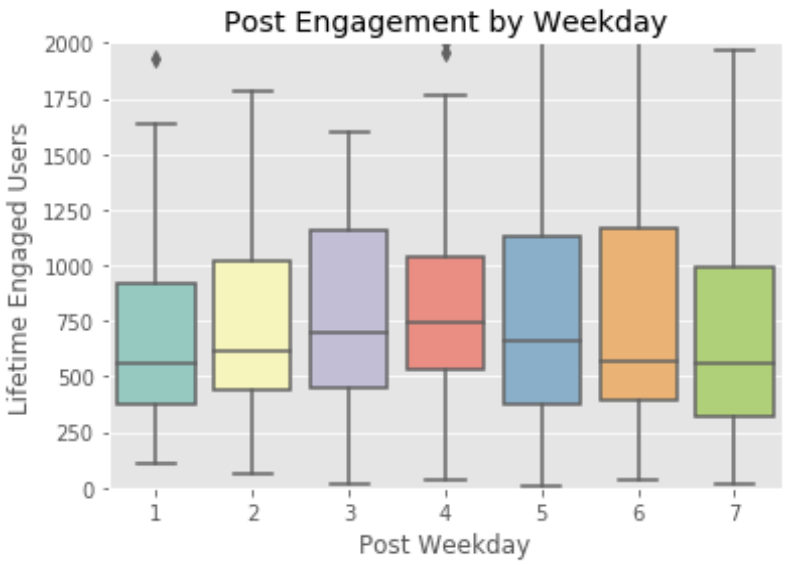

Figure.4.Post Engagement By Weekday

Fig.2 analyze the "Category" variable from the facebook data, Fig.3 analyze "Paid" variable from facebook data and finally Fig.4 analyze the "Weekdays" variable from the data, influence of the remaining four least relevant traits that hide 32 percent of the model's knowledge is shown. With regard to the "Category," it is noteworthy impact that "Activity" has composed of the remaining two traits. This category of 'Actions,' which clearly takes greater care than "products" or non-explosive branded content, stands for special offers and competitions. This type of "action" is used by the engaged users to increase the involvement of social media and therefore to enhance the postal services. The "Hour" influence graphics show that, while some peaks can be observed, no trends related to the hour of publication are found. The "Weekday" shows that the local impact is highly significant on "Monday," and that the local impact decreases until "Friday," when the total impact is highest. The anticipated greater effect on weekends tends to be available more in this period considering the users. This is an interesting result to examine when users interact with posts with additional data in future studies. The results for payment are expected: a post for which the committed user chooses, either for publicity paid or not paid, to have a higher effect. This is however one of the most significant input characteristics for the given model with a relevance of just 7 percent.

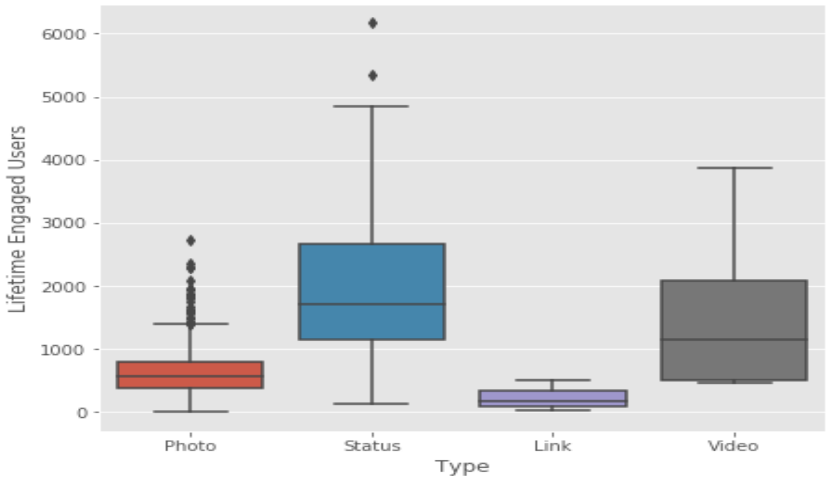

Figure.5.Influence of "Type" on "Lifetime Post Consumers

Fig.5 shows that status posts have clearly the greatest effect on the post results, more than two times the values for "Picture" and "Connection," and 60% more than "Video." This result is in line with findings that the most comments were received from the "Rank," the most preferred from "Videos," and the less interactive from "Images" and "Links." In addition, even though a similar conclusion has been reached for 'Status' postings, 'Photos' have come to another conclusion that they receive more likes and comments than 'Links' and 'Videos'. The user's expectations when they open Facebook were explicitly based here. Users mostly prefer to like images, photos, status and connections instead of status. "Lifetime engaged consumer" is the second most important function to be considerably less significant than "form" even though it has still an impact of 17 percent. This input feature refers to the page of the user that the post was published on at the time the post was written.

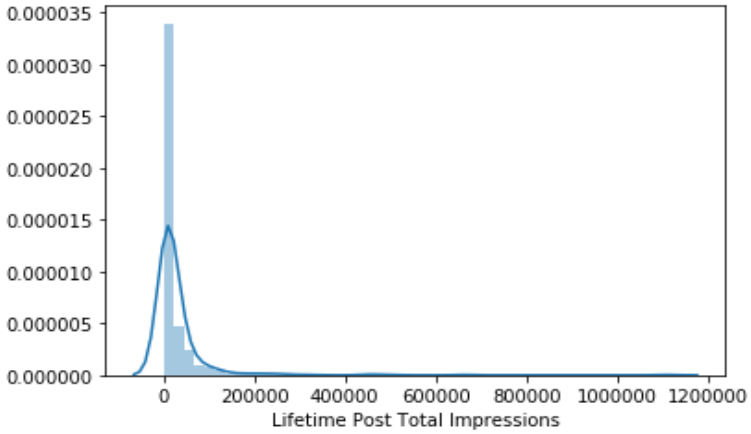

Figure.6.Lifetime Post Total Impressions

As the number of posts published on the page they previously liked increases overtime, more users receive feedback. A statement from Fig.6 shows, however, that after lifetime total printing the customer likes to decrease, that is to say that users are reluctant to engage in publishing posts. This issue may reveal some degradation of the facebook page because users display, but do not interact with the published content.

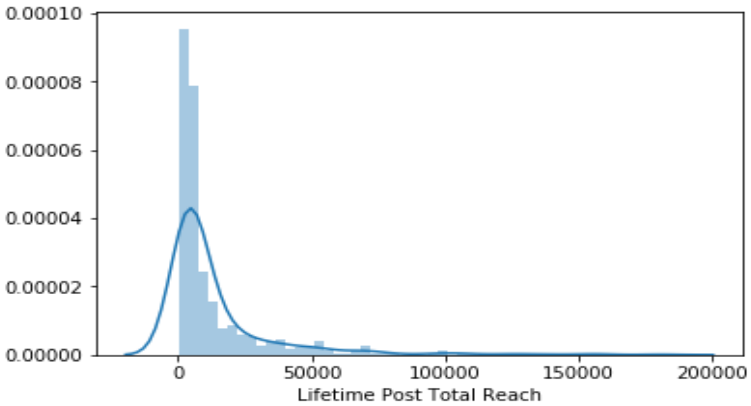

Figure.7.Lifetime Post Total Reach

However, Observation of Fig.7 showed that the cumulative lifetime span of the post for Facebook users is just 9.52%. It was decided to have more importance to the total lifetime of Post hour and website, form, month post and post weekday than paid ads. These findings indicate that paid advertising contributes little to the maximum life cycle, and after an hour it is the most significant factor in reaching various publications. Users are not involved in posting articles. This issue may reveal some degradation of the facebook page because users display, but do not interact with the published content.

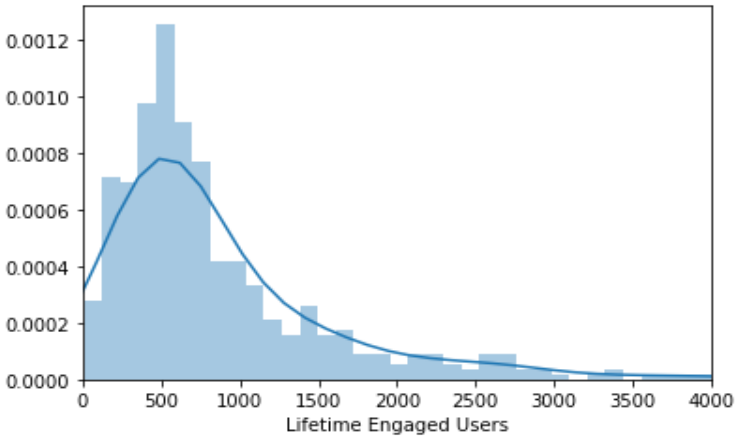

Figure.8.Lifetime Engaged Users

Fig.8. analyze that the Number of posts on the previously liked page increases overtime, more users receive feedback. An observation of Fig.8 shows that although the page is more popular, it was expected that postal users will decrease, that is to say, users are not keen to publish postings. Such a problem can reveal a degradation of the facebook page of the user, because users see, but do not interact with it, the content released.

```
Out[594]: <matplotlib.text.Text at 0x11b9bdd10>
```

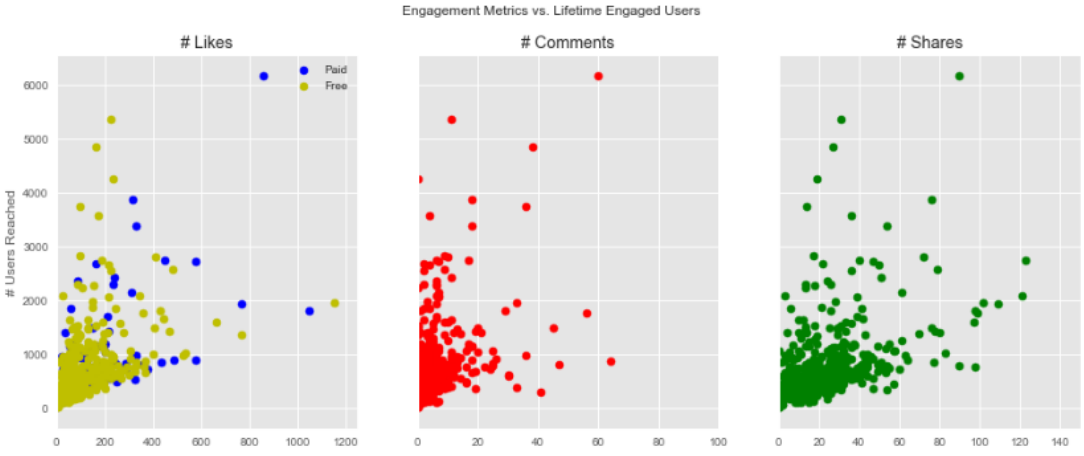

Figure.9.Relationships between Engaged Users and likes, shares and comments

Fig.9 analyzes the including three variables such as like, share and comments. The data collection contains only the missing label values, departure wait. In case of lack of labels the dataset will be removed. It is achieved in order to avoid a curse of dimensionality for the category functions assuming a greater number of distinct values or groups. Here see positive relationships most explicitly between Lifetime Engaged Users and likes/shares, and comments. Fig.9 analyze that the number of users are much interested for like in comparison of shares and comments. The following output analyze that mainly peoples are less interested for comments and share instead of like.

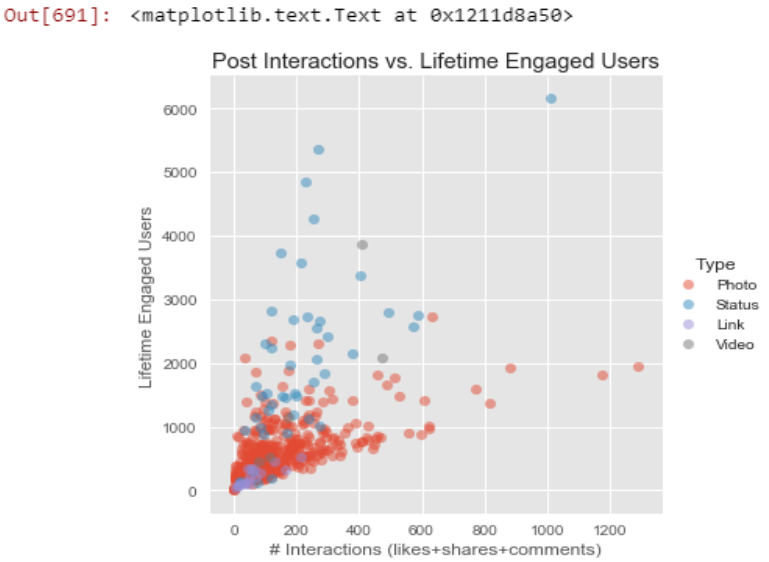

Figure.10. Interactions vs. Lifetime Engaged Users

Fig.10. analyzes interactions vs. Lifetime Engaged Users. Like, share and comments are an efficient factor of facebook data. Here x axis is defined the lifetime engaged user and y axis is defined the interactions includes like, share and comments. The output illustrate that the users are much engaged with the photo instead of status and link and video. Photo is defined by red color, status is defined by blue color, link is defined by green color and video is defined by purple color.

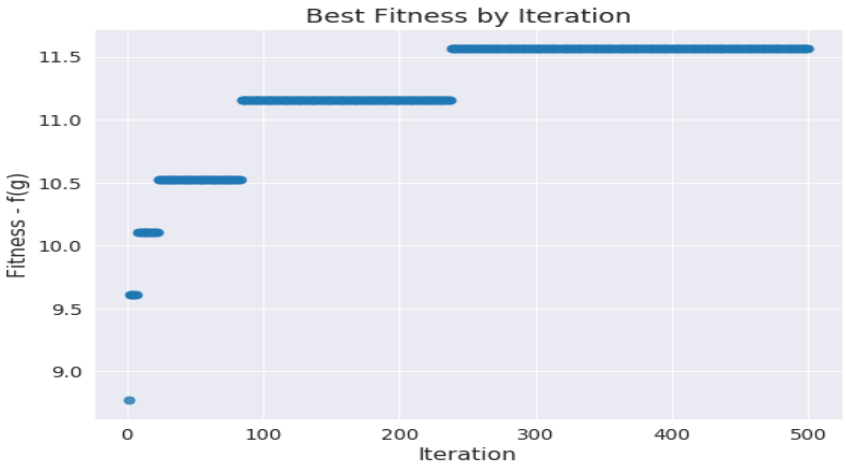

Figure.11.Best Fit Iteration

Fig.11. generate the initial population. There are some steps to analyze the best fit iteration. That algorithm based on the pseudo code, shown in the beginning of the notebook. Next, we show the results by iteration. We can notice that the g position merge to a local minima. The PSO is a modern and enhanced way of controlling maximum power point, output and fewer steady state oscillations. Both particles have fitness values to be measured by modifying the fitness function and have speeds that direct the flying of particles. Firstly, the discrete search field must be transformed into a continuous domain, a classic PSO used, and then the best solution found so far is to access the result.

6.2. Train Test Split

Training can be divided into two parts, the trained sentence context model and the trained generative model. Both training processes may be considered without manual marks as unregulated training. We divide three parameters, namely sharing and comment, for the training of the random forest model, lasso and support vector machine algorithm. This project is mainly aimed at predicting that committed users prefer lie, comment and share on a Facebook post. In the data following variables are available like page total like, type, category, post weakly, post monthly, post hourly, paid, like, share, comment and total iterations are around 500.

Data cleaning is an initial phase for the final analysis assessment of the dataset. Databases are susceptible to noisy, incomplete, and incoherent data because of the large amount of data available. This project's Facebook data are derived from kaggles that have 15 variables of various kinds and may not be consistent with the format in which Python allows the data to be used. Data Cleaning helps to erase noisy details and to eliminate incoherence. A missing value, sequentially entered, given a constant value or a mean value, may be ignored in the data cleaning process. In this case, the whole frame of the data is structured and arranged to preserve and delete the appropriate attributes. This is done to make the usage simpler and more feasible. The fill factor tells us how much room can be added on each page. The created filling factor value can generally be defined from 1 to 100 as a percentage.

Figure.12.Training and Testing By Random Forest Algorithm

Fig.12 illustrates the model performed no differently when taking the top 4 features by importance from the first model. It does not appear to have over fit (difference in train/test r^2 is now 0.014) and by all metrics is very consistent from training to test performance. The simpler model (with the top 4 features only) had higher performance than the model with the k-best 20 features. So, the final features are: Total Interactions, Status, Page total likes and Photo.

```
In [109]: RfPerf
```

Out[109]:

	Score	Step	metric
0	0.708000	train	r2
1	0.664000	test	r2
2	0.869696	train	Spearman
3	0.688929	test	Spearman
4	0.849800	train	Pearson
5	0.816980	test	Pearson

Figure.13. Scores By Random Forest Algorithm

Fig.13. iterating through a random forest using the most important variables led to some improvement, and the robust scoring metrics suggest that Lifetime Engaged Users can be modeled for to a reasonable degree. The model was able to predict total engaged users by the number of total interactions a post had, if the post was a status or photo, and how many likes the page had at the time of posting. A simple (4 feature) Random Forest performed better than a Lasso Regression model to a large degree, with a difference in test r^2 values of .05. However, both models demonstrated no signs of over fitting, with the test r^2 values all being close to or higher than the train r^2 value. This model can provide value by giving posters an accurate estimate of how many users they engaged by post, particularly useful for optimizing for post scheduling, and measuring post performance. Best results came from RF parameters are 500 estimators, 15 min sample split and Train/test split of 0.3. Here had solid performance in the test set, with 772 r^2 and 919 Spearman Correlation. But the model showed some signs of over fitting when exposed to the test sets are reduction of .15 in the r^2 and reduction of .10 in test Spearman Correlation. One reason that there could be over fitting is the large amount of features in the model. We can take the feature importance to get the top 15 features, then iterate through the Random Forest again, and see if over fitting persists. The error increases as the value of engaged users increases. Thus, this violates the one of the main assumptions of a support vector machine model. The support vector machine model performed solid overall. No over fitting: the R2 value rose .045 points from train to test. Moderate predictive power: .64 R2 train, .68 in test. I had to tune the test set division to 10% of the total dataset, to get the highest R2 value.

```
Train data R-2 score: train_score
Train data Spearman correlation: 0.6729813330121396
Train data Pearson correlation: 0.7457752447859023

Test data R-2 score: 0.6368296625459501
Test data Spearman correlation: 0.7126593378794965
Test data Pearson correlation: 0.800240564059518
```

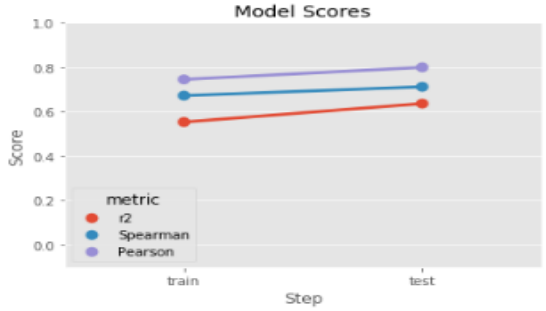

Figure.14.Training and Testing By Lasso Algorithm

In [62]:
```
lassoPerf
```

Out[62]:

	Score	Step	metric
0	0.554000	train	r2
1	0.637000	test	r2
2	0.672981	train	Spearman
3	0.712659	test	Spearman
4	0.745775	train	Pearson
5	0.800241	test	Pearson

Figure.15.Score by Lasso Algorithm

Fig.14 initiate that an analysis of a baseline model on facebook users engaged through facebook user dataset. The baseline model includes 20 different variables. However, that several L1 regularization coefficients are neutralized, that are implemented in a LASSO regression to avoid over fitting model. The fig.15 the optimal model analysis makes 64% reliable forecasts. For comparison purposes, the prediction of four characteristics of facebook user data is based on the same model. The predicted accuracy is higher. Lasso algorithm is good and it can handle load of heavy dataset. Train data spearman correlation is 87 and person is 74.

```
Train data R-2 score: train_score
Train data Spearman correlation: 0.44698661324382666
Train data Pearson correlation: 0.5206862737014576

Test data R-2 score: 0.13343067046094337
Test data Spearman correlation: 0.5369570742245938
Test data Pearson correlation: 0.4461070803905970 4
```

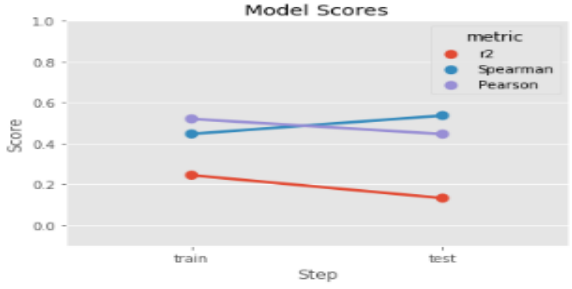

Figure.16.Training and Testing By Support Vector Machine

Fig.16 illustrate that each of the ensemble methods has predicted using machine learning algorithm and feature of data. The average performance of data has shuffled and recorded sampling used for creating samples for each of the validations. It is observed that the support vector machine algorithm method has predicted the best results using performance of R-2 score, .spearmen correlation and Pearson correlation for the process of training and testing the data.

```
In [53]: svmPerf
```

Out[53]:

	Score	Step	metric
0	0.245000	train	r2
1	0.133000	test	r2
2	0.446987	train	Spearman
3	0.536957	test	Spearman
4	0.520688	train	Pearson
5	0.446107	test	Pearson

Figure.17. Score by Support Vector Machine

Fig.17 illustrates the model performed no differently when taking the top 4 features by importance from the first model. For the accuracy calculation the R2 score is stated in a formula. In addition, time to complete the classification is also reported. For tests of various SVM assemblies, the test data set and the accuracy values are presented in Fig.17. As shown in Fig.17, the SVM labeling ensemble carries out the best 100% value accuracy measurement. This is consistent with the R2 score and spearman and Pearson shown in Fig.17, where most of the approaches have higher scored as compared to other two algorithms. The simpler model (with the top 4 features only) had higher performance than the model with the k-best 20 features. So, the final features are: Total Interactions, Status, Page total likes and Photo.

Table 1: Comparison Results of Metrices Prediction of the Machine Learning tools [41]

Methods	Comments	Shares	Likes
SVR	0.002988	0.003396	0.003817
ESN	0.002078	0.001640	0.003263
ANFIS	0.0022082	0.001967	0.002221

Table.1. have estimated the effect of a post on Facebook's social network. There are 7 characteristics known prior to post publication in the dataset, and 3 output variables that are used for the post effect. The variables for output are: comments, shares, and likes. Better results than SVR are obtained by the new proposed methods in this paper. While ANFIS seems to do better in predicting the number of likes, in the other two situations, the ESN model has better accuracy. We will continue observations with the use of the ANFIS model for many other data to include it with other second methods.

Table 2. Comparison Results of Performance vector values of SVM [42]

	SVM
Root mean squared error	5.403 ± .630
Absolute error	4.216 ± .417
Correlation	.606 ± .077

Table.2. clarifies the attributes that should be clearly discussed in clinical circumstances. In this study, however, self-report measurement instruments were used, reflecting a limitation. In addition, data was obtained using a system of sampling techniques. The sample finally had more males than females. Measurement instruments that include more reliable data instead of self-report scales may be used in subsequent studies. Furthermore, by using a random sampling process, the reliability of the analysis can be increased. In the distribution of women and men, the random sampling approach will produce a more heterogeneous sample.

We have used best algorithms in comparison of these two works. We have used four algorithm such as Particle Swarm optimization, Random Forest Algorithm, Lasso Algorithm and Support Vector Machine. These are the best algorithms in comparison with other two works. We are getting 100 percent accuracy with SVM algorithms. It is predicting best results. These algorithms are able to handle large and overfit data. These algorithms are able to predict higher accuracy in comparison of other two works. Each comparison works defines the user interest on social media with the use of machine learning. But their results are not effective in comparison of our results. We have used the best algorithms and technique to predict the user interest and accuracy.

7. Conclusion

In this research work on theoretical machine learning model has been proposed for prediction of relationship between like, share and comments for lifetime engaged users on facebook. The evaluation of the theoretical model using three algorithms such as lasso regression and random forest classifiers and support vector machine showed good performance of relationship between like, share and comment on facebook. Based on the analysis in this research work it was concluded as there is getting higher accuracy from support vector machine in comparison of other two algorithms. Here used PSO algorithm for increase the performance of the results. A simple (4 feature) support vector machine formed 100% accuracy than Lasso Regression and random forest model to a large degree, with a difference in test values of .05. PSO algorithm is used to solve the optimization problem. PSO has been mainly used to solve unregulated problems. PSO algorithm is used to increase the performance of the results as

shown in figure.11. However, both models demonstrated no signs of over fitting, with the test r^2 values all being close to or higher than the train r^2 value. This model is providing the value by giving posters an accurate estimate of how many users they engaged by post, particularly useful for optimizing for post scheduling, and measuring post performance. Therefore, a combination of two or more machine learning algorithms used for get the number of users engaged for facebook user engaged.

References

1. Troussas, C. and Virvou, M., 2020. *Advances in Social Networking-based Learning: Machine Learning-based User Modelling and Sentiment Analysis* (Vol. 181). Springer Nature.
2. Jithendran, A., Karthik, P.P., Santhosh, S. and Naren, J., 2020. Emotion Recognition on E-Learning Community to Improve the Learning Outcomes Using Machine Learning Concepts: A Pilot Study. In *Smart Systems and IoT: Innovations in Computing* (pp. 521-530). Springer, Singapore.
3. Khanal, S.S., Prasad, P.W.C., Alsadoon, A. and Maag, A., 2019. A systematic review: machine learning based recommendation systems for e-learning. *Education and Information Technologies*, pp.1-30.
4. Razis, G., Anagnostopoulos, I. and Zeadally, S., 2020. Modeling Influence with Semantics in Social Networks: A Survey. *ACM Computing Surveys (CSUR)*, *53*(1), pp.1-38.
5. Zheng, J., Wang, S., Li, D. and Zhang, B., 2019. Personalized recommendation based on hierarchical interest overlapping community. *Information Sciences*, *479*, pp.55-75.
6. Shawky, D. and Badawi, A., 2019. Towards a personalized learning experience using reinforcement learning. In *Machine learning paradigms: Theory and application* (pp. 169-187). Springer, Cham.
7. Sengupta, A. and Ghosh, A., 2020. Mining Social Network Data for Predictive Personality Modelling by Employing Machine Learning Techniques. In Computational Advancement in Communication Circuits and Systems (pp. 113-127). Springer, Singapore.
8. Nitchot, A., Wettayaprasit, W. and Gilbert, L., 2019. Personalized learning system for visualizing knowledge structures and recommending study materials links. *E-Learning and Digital Media*, *16*(1), pp.77-91.
9. Kurilovas, E., 2019. Advanced machine learning approaches to personalise learning: learning analytics and decision making. *Behaviour & Information Technology*, *38*(4), pp.410-421.
10. Srisa-An, C. and Yongsiriwit, K., 2019. Applying Machine Learning and AI on Self Automated Personalized Online Learning. *Fuzzy Systems and Data Mining V: Proceedings of FSDM 2019*, *320*, p.137.
11. Parchoma, G., Koole, M., Morrison, D., Nelson, D. and Dreaver-Charles, K., 2019. Designing for learning in the Yellow House: a comparison of instructional and learning design origins and practices. *Higher Education Research & Development*, pp.1-16.
12. Sharma, S. and Rana, V., 2020. Web Search Personalization Using Semantic Similarity Measure. In Proceedings of ICRIC 2019 (pp. 273-288). Springer, Cham.
13. Al-Abri, A., Jamoussi, Y., AlKhanjari, Z. and Kraiem, N., 2019, January. Aggregation and Mapping of Social Media Attribute Names Extracted from Chat Conversation for Personalized E-Learning. In 2019 4th MEC International Conference on Big Data and Smart City (ICBDSC) (pp. 1-9). IEEE.
14. George, G. and Lal, A.M., 2019. Review of ontology-based recommender systems in e-learning. Computers & Education, 142, p.103642.
15. Huang, P.S., Chiu, P.S., Chang, J.W., Huang, Y.M. and Lee, M.C., 2019. A study of using syntactic cues in short-text similarity measure. Journal of Internet Technology, 20(3), pp.839-850.

16. Wang, J., 2019. Ontology Technique and Meaningful Learning Support Environments. In Learning Technologies for Transforming Large-Scale Teaching, Learning, and Assessment (pp. 215-229). Springer, Cham.
17. Zhang, Z. and Liu, X., 2019, July. Ontology-Based Computing of Sentence Similarity. In The International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (pp. 953-961). Springer, Cham.
18. Ibrahim, M.E., Yang, Y., Ndzi, D.L., Yang, G. and Al-Maliki, M., 2018. Ontology-based personalized course recommendation framework. IEEE Access, 7, pp.5180-5199.
19. Tarus, J.K., Niu, Z. and Mustafa, G., 2018. Knowledge-based recommendation: a review of ontology-based recommender systems for e-learning. Artificial intelligence review, 50(1), pp.21-48.
20. Gómez-Berbís, J.M. and Valencia-García, R., 2018, July. Ontology-Based Advertisement Recommendation in Social Networks. In Distributed Computing and Artificial Intelligence, 15th International Conference (Vol. 800, p. 36). Springer.
21. Sürer, Ö., 2017, August. Improving Similarity Measures Using Ontological Data. In Proceedings of the Eleventh ACM Conference on Recommender Systems (pp. 416-420).
22. Aeiad, E. and Meziane, F., 2019. An adaptable and personalised E-learning system applied to computer science Programmes design. *Education and Information Technologies*, *24*(2), pp.1485-1509.
23. Moubayed, A., Injadat, M., Nassif, A.B., Lutfiyya, H. and Shami, A., 2018. E-learning: Challenges and research opportunities using machine learning & Data analytics. *IEEE Access*, *6*, pp.39117-39138.
24. Gavrielov-Yusim, N., Kürzinger, M.L., Nishikawa, C., Pan, C., Pouget, J., Epstein, L.B., Golant, Y., Tcherny-Lessenot, S., Lin, S., Hamelin, B. and Juhaeri, J., 2019. Comparison of text processing methods in social media–based signal detection. *Pharmacoepidemiology and Drug Safety*, *28*(10), pp.1309-1317.
25. Bourkoukou, O., El Bachari, E. and El Adnani, M., 2016. A personalized e-learning based on recommender system. *International journal of learning and teaching*, *2*(2), pp.99-103.
26. Tang, Y. and Wang, W., 2018. A literature review of personalized learning algorithm. *Open Journal of Social Sciences*, *6*(1), pp.119-127.
27. Logesh, R., Subramaniyaswamy, V. and Vijayakumar, V., 2018. A personalised travel recommender system utilising social network profile and accurate GPS data. *Electronic Government, an International Journal*, *14*(1), pp.90-113.
28. Arora, A., Bansal, S., Kandpal, C., Aswani, R. and Dwivedi, Y., 2019. Measuring social media influencer index-insights from facebook, Twitter and Instagram. *Journal of Retailing and Consumer Services*, *49*, pp.86-101.
29. Nadar, N. and Kamatchi, R., 2019. Information and Communication-Based Collaborative Learning and Behavior Modeling Using Machine Learning Algorithm. In *Social Media and Machine Learning*. IntechOpen.
30. Kristensen, J.B., Albrechtsen, T., Dahl-Nielsen, E., Jensen, M., Skovrind, M. and Bornakke, T., 2017. Parsimonious data: How a single Facebook like predicts voting behavior in multiparty systems. *PloS one*, *12*(9).
31. Marengo, D. and Settanni, M., 2019. Mining Facebook Data for Personality Prediction: An Overview. In *Digital Phenotyping and Mobile Sensing* (pp. 109-124). Springer, Cham.
32. Bogaert, M., Ballings, M., Hosten, M. and Van den Poel, D., 2017. Identifying soccer players on Facebook through predictive analytics. *Decision Analysis*, *14*(4), pp.274-297.

33. Banouar, O. and Raghay, S., 2017, March. Machine learning for personalized access to multiple data sources through ontologies. In *Proceedings of the 2nd international Conference on Big Data, Cloud and Applications* (pp. 1-6).
34. Aeiad, E., 2017. *A framework for an adaptable and personalised e-learning system based on free web resources* (Doctoral dissertation, University of Salford).
35. Srivastava, B. and Haider, M.T.U., 2017. Personalized assessment model for alphabets learning with learning objects in e-learning environment for dyslexia. *Journal of King Saud University-Computer and Information Sciences*.
36. Kristensen, J.B., Albrechtsen, T., Dahl-Nielsen, E., Jensen, M., Skovrind, M. and Bornakke, T., 2017. Parsimonious data: How a single Facebook like predicts voting behavior in multiparty systems. *PloS one*, *12*(9).
37. Mullainathan, S. and Spiess, J., 2017. Machine learning: an applied econometric approach. *Journal of Economic Perspectives*, *31*(2), pp.87-106.
38. Jando, E., Hidayanto, A.N., Prabowo, H. and Warnars, H.L.H.S., 2017, November. Personalized E-learning Model: A systematic literature review. In *2017 International Conference on Information Management and Technology (ICIMTech)* (pp. 238-243). IEEE.
39. Guenther, N. and Schonlau, M., 2016. Support vector machines. *The Stata Journal*, *16*(4), pp.917-937.
40. Preoțiuc-Pietro, D., Schwartz, H.A., Park, G., Eichstaedt, J., Kern, M., Ungar, L. and Shulman, E., 2016, June. Modelling valence and arousal in facebook posts. In *Proceedings of the 7th workshop on computational approaches to subjectivity, sentiment and social media analysis* (pp. 9-15).
41. Sam, E., Yarushev, S., Basterrech, S. and Averkin, A., 2018. Prediction of Facebook Post Metrics using Machine Learning. *arXiv preprint arXiv:1805.05579*.
42. Savci, M., Tekin, A. and Elhai, J.D., 2020. Prediction of problematic social media use (PSU) using machine learning approaches. *Current Psychology*, pp.1-10.