Article

User-Engagement Score and SLIs/SLOs/SLAs Measurements Correlation of E-Business Projects Through Big Data Analysis

Solomiia Fedushko ^{1,*} , Taras Ustyianovych ¹ , Yuriy Syerov ¹ and Tomas Peracek ²

¹ Social Communication and Information Activity Department, Lviv Polytechnic National University, 79000 Lviv, Ukraine; taras.ustyianovych.dk.2017@lpnu.ua (T.U.); yurii.o.sierov@lpnu.ua (Y.S.)

² Faculty of Management, Comenius University in Bratislava, 814 99 Bratislava, Slovakia; tomas.peracek@fm.uniba.sk

* Correspondence: solomiia.s.fedushko@lpnu.ua

Received: 25 November 2020; Accepted: 17 December 2020; Published: 20 December 2020

Abstract: The Covid-19 crisis lockdown caused rapid transformation to remote working/learning modes and the need for e-commerce-, web-education-related projects development, and maintenance. However, an increase in internet traffic has a direct impact on infrastructure and software performance. We study the problem of accurate and quick web-project infrastructure issues/bottleneck/overload identification. The research aims to achieve and ensure the reliability and availability of a commerce/educational web project by providing system observability and Site Reliability Engineering (SRE) methods. In this research, we propose methods for technical condition assessment by applying the correlation of user-engagement score and Service Level Indicators (SLIs)/Service Level Objectives (SLOs)/Service Level Agreements (SLAs) measurements to identify user satisfaction types along with the infrastructure state. Our solution helps to improve content quality and, mainly, detect abnormal system behavior and poor infrastructure conditions. A straightforward interpretation of potential performance bottlenecks and vulnerabilities is achieved with the developed contingency table and correlation matrix for that purpose. We identify big data and system logs and metrics as the central sources that have performance issues during web-project usage. Throughout the analysis of an educational platform dataset, we found the main features of web-project content that have high user-engagement and provide value to services' customers. According to our study, the usage and correlation of SLOs/SLAs with other critical metrics, such as user satisfaction or engagement improves early indication of potential system issues and avoids having users face them. These findings correspond to the concepts of SRE that focus on maintaining high service availability.

Keywords: educational web-projects; real-time data analysis; big data; system performance; Application Performance Monitoring (APM); business-plan; strategy; Service Level Objectives (SLO); Service Level Agreement (SLA); Service Level Indicator (SLI)

1. Introduction

Hardware and software technical assessment of an educational web project in the face of the increased need for their use not only creates many challenges, but also requires fast and objective data-driven operations and decisions. This need is especially relevant during the Covid-19 crisis, as it forces educational and business institutions to make the necessary migration to the online mode. Educational institutions have faced the need to provide teachers with a flexible IT infrastructure that empowers efficient deployment of educational materials and courses both in regular times and in a state of emergency [1]. Educational institutions' activities become almost impossible and less valuable

without the availability of electronic educational web projects, in comparison to modern competitive organizations that provide similar services. Many universities in Ukraine and around the world are already effectively using online learning as one of the leading strategies for building and developing educational institution services. In addition, it was proven that the use of web-based resources for online learning is more effective compared to traditional learning methods [2].

The transition to the online mode allows efficient and user-friendly use of e-learning technologies, which are defined as effective multimedia learning using e-educational technology [3]. Thus, we state that e-learning is a component of educational technology, the further development of which is currently underway by many technical giants of the business world. In particular, the article by R. Lakshminarayanan, B. Kumar, and M. Raju [4] considered how companies offering cloud services and technologies, in particular, Amazon Web Services (AWS), Microsoft, and Google allow educational institutions to take advantage of certain products. In the study, a comparative and features analysis of certain product usage was made.

The course of events related to the Covid-19 pandemic outbreak has led to a closer look at the need for digitalization and the establishment of so-called e-learning within certain institutions and organizations. We display in Figure 1, which was generated using the Google Trends service [5], that with the onset of the Covid-19 pandemic, the increased rates of searches related to online learning projects vary from 15% to 43%. The values show the popularity of a search term relative to the highest point on the graph for a particular region and time period. 100 is the peak popularity of a search term, while 50 means that the popularity of the term is half as popular during the specified period.

Figure 1. Google Trends statistics of online learning services search and usage.

However, many educational institutions have not been sufficiently prepared for the switching to online task mode due to various reasons:

1. lack of internal infrastructure or subscriptions to external online projects to provide educational services;
2. insufficient reliability of institutions' infrastructural and technical support;
3. complete or partial lack of the necessary teaching materials and resources for online classes;
4. the lack of the university's strategy for web projects implementation to support students' e-learning needs.

In this article, we decided to solve the following problem: web project technical control and evaluation of an educational institution or e-commerce, which faces the problem of high load on hardware and web-software during remote learning and work; identify the cause-and-effect relationships of particular infrastructure type issues and improve site reliability methods usage. The need for clear external or internal root-cause problem identification is essential because this will allow the formation of qualitative Service Level Objectives (SLOs).

The study objectives are as follows: (1) increase infrastructure visibility through correlation and use of a user-engagement score with Service Level Objectives (SLOs)/Service Level Agreement (SLA)/Service Level Indicator (SLI) as the main indicators of the web project technical equipment quality; (2) improve methods for data analysis of virtual environment performance metrics; (3) interactively monitor various web service processes and evaluate the technical characteristics of servers, applications, etc. based on indicators and goals; (4) provide statistical methods for meaningful review of web project goals based on metrics coming from the data sources in order to improve all the processes described above; (5) develop methods to increase availability and reliability of educational web projects; (6) use metrics in real-time to indicate system resource shortages and bottlenecks instead of critical user responses; (7) increase the efficiency of a university's online educational service and objectively create the requirements for scaling, creating effective solutions and architectures through monitoring; (8) improve decision-making processes regarding the architecture and IT operations of an online educational project. Consequently, a project designed for e-learning, as well as the particular university or institution's educational technology must follow data-driven decision making.

2. Literature Review

Many methodologies and frameworks have already been developed for technical assessments and monitoring. For example, M. Bashirov's study used mathematical modeling and an electromagnetic-acoustic effect to determine the defect and reliability of pipelines. The predictive model usage at an early stage is determined to increase the probability of defect identification [6]. A certain percentage of web-project technical equipment (hardware) or software in e-learning or e-commerce is outdated, in which case its monitoring and reliability assessment is difficult and limited. The main reason is that there are difficulties in establishing connection with the legacy devices themselves; fewer metrics and logs are generated, and data processing and transfers are slower compared to modern IT solutions. A study by Bednarski et al. showed the results of the historical structure's technical condition assessment. Physical quantity measurements were important for the crack state of a historic church in Jangrot, the assessment allowed identifying the kind of components which needed to be replaced. The study mainly emphasized the need for environmental data collection and its correlation with information about a particular equipment unit's condition and reliability. The need for advances in technical measurement data collection and the development of new means for quick data ingestion was emphasized as well [7–9]. The most focus is on the problem of infrastructure insecurity and unreliability, which can lead to web project instability and the inability to provide educational services by institutions in real-time. The educational web projects' reliability and availability is a prerequisite for providing qualitative services in higher education institutions and schools. It plays a particularly important role in isolation and quarantine conditions, for example, Covid-19, as educational institutions are transitioning to online teaching and e-learning. An equally important problem faced by educational institutions during the period of abnormal load on the online project technology infrastructure is the lack of visibility and real-time monitoring, which makes it impossible to make objective decisions regarding system scaling and troubleshooting. Accordingly, if we take into account the technological needs and scale of a particular online project, the need for monitoring, site reliability engineering, and operational intelligence methods becomes increasingly clear [10]. Belforte et al. provide an example distributed across more than 60 computing centers worldwide, with CMS management and monitoring using custom and traditional machine reliability metrics. In addition, an algorithm to automate the performance of distributed resources is described, which is very valuable for an online

system that uses load-balancing between clusters [11]. The implementation of this algorithm can facilitate and improve monitoring processes.

A correct SLA definition allows a strong user/customer understanding. Operational intelligence and real-time monitoring techniques involve the implementation of user behavior analytics. This process, in turn, allows not only understanding the users and monitoring their behavior in real-time, but also to determine certain performance indicators for an individual user, to set objective SLAs. In the article by Alfian et al., big data methods were used to collect browsing history and transaction data for real-time analysis of user behavior interacting with e-services in different locations. This study allows web projects to improve service quality and to establish the optimal service level agreement, which will be beneficial for all parties involved. Equally essential is the real-time monitoring of personalized diabetic patients' health. The study is valuable because it reflects the use of the Bluetooth Low Energy method to reduce the cost of data collection, as well as to provide high-quality advice to patients. Machine learning predicts the likelihood of detecting diabetes in patients based on the collected data. This technical solution can be used not only for medical data collection, but also for the equipment and infrastructure data transmission to a centralized logging environment [12,13]. Web-analytics tools are used for real-time mouse tracking, which helps to collect data about users and their interaction with an e-service. Accordingly, the data correlate with transactions and queries' completion rates, which are necessary to monitor availability and reliability. The article by Cegan and Filip describes how user behavior monitoring allows detecting bottlenecks in the web environment, as well as in technical equipment and infrastructure components. The authors propose a new method for collecting mouse-clicking data based on real-time data transformation to convert discrete position data to system functions to optimize compression and analysis [14]. The usage of functional tests to assess the infrastructure and site reliability as a way to validate site operations and optimization techniques is described in Elmsheuser et al.'s study [15].

Machine learning and artificial intelligence are also widely applied for technical condition assessment. A recent study conducted by Kaminski et al. showed the usage of artificial neural networks, namely the multilayer perceptron, to assess the technical condition of a water supply system. The results proved that the use of artificial intelligence in such tasks can increase the efficiency of detecting defects in pipes with a distributed water supply chain and is an example of human-machine interaction [16]. An improved support function machine model indicated that the pattern recognition method based on an improved kernel function support vector machine is efficient for validating technical conditions of the recoil mechanism [17]. The embedded system's usage greatly simplifies real-time monitoring, as it allows more metrics to collect than are collected during traditional system monitoring. Studies by Bosse and Lehmhus provide a model for data collection using a structural monitoring and tactile sensing system to obtain data from the lowest system levels. It allows an IT team to assess and monitor the technical condition of the equipment with great accuracy [18]. This research is crucial for our study, because in order to form new metrics and find infrastructure issues' causal relationships, and create an optimal SLO/SLA definition, it is necessary to collect data from all possible levels of web project applications.

Despite the growing need for e-learning, many research and educational institutions are not ready to move to a full-fledged online mode due to insufficiently reliable technical equipment and/or infrastructure. This leads to the issue that many information-educational web projects during the Covid-19 crisis are not able to maintain stability when the load of materials, users and downtimes occur; in addition, the transactions might not be processed in a proper way. The problem remains unresolved, as only a small percentage of Ukrainian universities were ready to move to remote teaching and learning courses when the lockdown started. The main signals of an e-learning or e-business project problem should be data and metrics that show unsatisfactory performance indicators, but not negative user feedback and/or open incidents for the e-project support team.

A study by Feldmann et al. found that internet traffic grew by about 15–20% within one week of the Covid-19 crisis due to the increased use of online resources, namely: web conferencing,

VPN, e-commerce, e-learning, and gaming. These findings are similar to the insights shown by mobility reports published by Google and prove the increased digital demand during Covid-19 [19,20]. In addition, ensuring the infrastructure reliability requires the implementation of system quality monitoring methods, setting certain goals to warrant highly reliable and uninterrupted IT operation, as well as scalability, which might previously have been lacking. The scarcity of real-time data analytics deprives a project of observability and does not allow accurate estimation of the actual educational web project needs for handling end-to-end operations and measuring the load on them.

The research by Canizo et al describes the implementation of a monitoring solution on a real industrial use case that includes several industrial press machines. The effectiveness and its scalability factors are proved. A big data architecture for industrial Cyber-Physical System (CPS) monitoring is proposed, considering four main data factors during the implementation, namely: data acquisition, processing, persistence and server availability. The data collection process is implemented using programmable logic controllers. Message streaming and parallel processing tools are used to transfer and transform the data. The research is valuable because of Signe and multiple data anomaly detections that are applied as calculation frameworks to detect issues and anomalies. The anomaly detection algorithms are mainly based on checking previous and current system states. Nevertheless, this implementation addresses all the main issues that a CPS faces. The proposed solution has increased the overall equipment effectiveness [21].

The state-of-the-art real-time big data processing technologies that are used for anomaly detection, abnormal system behavior and vital machine learning algorithm features are studied by Habeeb et al. In the research they describe frameworks to handle big data processing in real-time in order to identify system issues and security vulnerabilities; a survey of big data techniques was conducted. The research also provides comprehensive big data techniques to monitor network data [22].

The definition of Service Level Objectives and usage is necessary for reliability monitoring, resource utilization reduction, and performing computationally inexpensive calculations. The article describes performance modeling with profiling to ensure low system performance usage so that the resources used for e-commerce can be decreased by three times [23]. This framework can be effectively automated and applied to universities' e-learning projects in order to ensure high reliability and to reduce costs of infrastructure maintenance. In addition, it proves that monitoring and control of Service Level Objectives can increase project efficiency, and therefore their usage within EdTech remains necessary.

3. Materials and Methods

EdTech and e-business project availability and reliability are important to ensure qualitative service delivery and product distribution. This need is especially necessary and noticeable during the period of remote work or study. Service unavailability might cause financial losses and also does not allow for a quality educational process. Research by Melo et al. [24] helps to estimate how much money can be saved by increasing system availability using SLA; it also presents analysis of various system architectures to ensure a beneficial cost-benefit relationship. Research from Fortune 1000 companies shows the downtime value for business-critical metrics. For example, on average the total cost of an unplanned system downtime per year is about \$1.25 billion and up to \$2.5 billion [25]. Additionally, making objective data-driven decisions [26–28] about a particular service [29,30], user [31], and technical condition [32] is a key prerequisite for ensuring equipment quality and reliability.

The modern monitoring tools usage will simplify the task of the equipment's technical assessment and control. According to the Gartner Share Analysis Report for 2019, digital products focus is increasingly on end-user experience monitoring, which is very important because it provides a clear user-to-web project interaction understanding. This has a critical impact on business income and the user's desire to continue using the service in the future. Especially notable is the end-user experience in the period of the increased need to use online education and business web projects, such as in the period of COVID-19. Also, the ITOM (IT Operations Management) performance analysis software

market grew by around 11% compared to 2018. The AI (Artificial intelligence) usage, namely AIOps (Artificial Intelligence Operations), ITIM (IT Infrastructure Monitoring) and other monitoring solutions hold around 45% of the performance market demand, whereas the APM and network monitoring segments [33] have decreased to demands of 34% and 21%, respectively.

Downtime and minimalizations in spending are essential to address, as well as insufficient hardware and software viability/availability. The SLO definition and usage is not capable of solving the system technical condition assessment problem without the help of SLI/SLO/SLA data and metadata.

The use of SLOs is a popular trend today; different-sized enterprises that run electronic services are increasingly using it. SLOs and the SLA serve as motivating factors that provide a goal setting process. It encourages an organization to achieve the goals, increase the threshold and overcome it again; measure user satisfaction level with and without an Apdex score correlation; accept limiting the threshold of system availability; and help understand at what infrastructure improvement stage a web project can improve performance in the future. There are many tools for SLO computations and monitoring on the market, and they can be easily applied to solve various problems with very complex demands.

We used the following methods to conduct and obtain the research results: data engineering; data collection and logging; mathematical and statistical methods to calculate KPIs, user-engagement, and correlation analysis; site reliability engineering methods; exploratory and descriptive methods for prescriptive data analysis; incidents management analysis; data visualization; business-plan and long/short-term strategy for infrastructure improvement formation.

To obtain the required data for the algorithm implementation and SLO definition, it is necessary to monitor the infrastructure of the electronic environment from the log files stored in the system and collect web project user behavior and activity statistics.

Moreover, SLO/SLA adherence metadata should be correlated with other important metrics for e-education and e-business, namely: Customer Profitability Score (CPS), Net Profit Margin, Conversion Rate, Net Promoter Score (NPS), and relative market share. These and other metrics need to be monitored on interactive panels, in dashboards for operational intelligence monitoring.

Before determining the service level objectives, the following questions should be answered: what percentage of web project performance increase should be met; whether an increase in availability will affect a raise in profits, and if yes, then how much; what interdependence level between user-engagement score and service level indicators (SLI) is observed.

4. Results

The short- and long-term SLO definition is fundamental in monitoring. The former is vital for systems engineers, while the latter is necessary for the strategy development and management departments of a particular web project. However, in terms of infrastructure technical condition assessment and investigating web environment impact on hardware, both SLO types are useful and need to be analyzed.

4.1. User-Engagement Calculations to Assess Educational Web Project Parts Interaction

Both reliability and availability metrics are valuable in the application of performance management and monitoring. However, both are different from each other, because a piece of technical equipment may be available but not reliable. For example, we could consider a case where we suppose we do have equipment X, which has a frequent connection loss for 6 min every hour. That means 90% availability, but less than 1 h of reliability, which is a poor indicator value for e-commerce.

Self-education e-resources have been in demand especially since the start of the Covid-19 crisis. For instance, registration on the popular educational platform Coursera, which underwent a partnership with various universities during the Covid-19 quarantine period, is up by 173% in March 2020, while course enrollment increased by 145% compared to February the same year [34].

In the study, we analyzed the 2017 business and finance courses data from an educational platform Udemy in order to obtain the user-engagement score and perform association tests for the obtained values with the subscription and review count, prices, and course duration. This allowed us to validate the user-engagement score as an unbiased metric for web project assessment and correlation with technical condition data. We investigated how the user-engagement score might affect the increase in efficiency of careful Service Level Objectives definition and computation. Also, we evaluated various indicators and strategies to assess infrastructure technical conditions based on the correlation between SLO/SLA and user-engagement scores. The developed framework is to help identify possible system issues and improve regular downtimes management. According to Google Trends statistics, the peak of popularity of the platform from 2017 to 2020 inclusive was reached in late March–early April 2020 and was 100%. Thus, the losses due to the unavailability of the service can be much greater than in the normal period. Therefore, the implementation of technical condition assessment using site reliability and SLO methods is necessary for the full services and information product provision.

We found that in 2017 Udemy business and finance domain courses data, there are 1195 observations out of which 94.85% entities have a label “paid”, although, the user enrollment in paid courses is 75.58%. The average cost of a studied online course was \$120, and the median and the mode were \$124 and \$200, respectively. In the majority of cases, online classes that were published earlier in the year had more students signed up by the end of the year. In addition, 0.03% of free courses had a higher than average number of subscribers and even exceeded subscriptions to paid courses. This indicates that if the web project training is free and the user believes that its material is well-structured enough, it is more likely that a person will register for the class, and not for a paid one. In turn, this will create a load on the infrastructure, induce more course enrollments, but can be less profitable to an e-learning service provider.

The paid course duration takes longer compared to free ones. We determined the total sum of hours required for each registered participant to complete the course. The number of subscribers is multiplied by the course duration to calculate this. The total number of subscribers was 668,938, so the entire time for all users’ course completions was 3,754,806 h.

On average, one hour of paid content costs a user \$18.6. Assume that a 1% loss of availability can block users from enrolling or attending the selected class, a loss in profit will be about \$123,700 per hour, according to our data.

A detailed data examination revealed that the most expensive courses are at the expert level, and in terms of duration they are in the interval of 1–100 h, while classes with a longer span (~200–300 h) are only beginner level and cost less by about \$20 than those in less long courses. Thus, the shorter the course duration (Figure 2) —the less load on the infrastructure is created, and more profit is made not only due to the higher e-training cost, but also lower costs associated with the service disposal. This is important to consider.

User-engagement calculation and monitoring along with correlation of the SLAs compliance levels will allow us to assess the technical equipment, infrastructure condition, as well as to determine which categories of users face particular online project bottlenecks. An equation to determine an online project user-engagement score with respect to n number of content entities has been developed. Our formula offers the possibility of its application both to a web project separate element and the service entirely; input parameters are the most important metrics for understanding the web project and its participants’ interactions; the weights are calculated on a comparative basis of all project components. Below is Equation (1) for specific content i .

$$U_i = \frac{\max(R_i, C)}{\max(V_i, C)} W_i, \quad (1)$$

where U_i is the user-engagement value. Its range depends on the minimum and maximum values available in the dataset for the arguments V_i and R_i , V_i is the number of views/visitors/subscriptions on a certain online project topic. R_i is the number of reactions on a certain online project topic, where 0

$\leq R_i \leq V_i$. C is a constant used to avoid getting zero during the division. In our research, we set it to 0.5 because it is an artificially created value that does not count as a real user review R_i value. W_i is the weight parameter, which is different for each topic and its user-engagement U_i value.

Figure 2. Business and finance course price and duration grouped by competency levels.

Max functions are used to obtain the maximum value between V_i or R_i and the constant. We found that more than 9% of courses do not have any user reviews, but the number of subscribers ranged from 0 to 1600. This is quite unclear and at the same time captivating because it means that there was insufficient user-engagement, probably due to a lack of a user- and content-understanding. Apart from that, we consider that users might have faced technical issues when viewing the content of these courses. This finding needs further study to determine what leads to this kind of user-behavior. Table 1 shows the number of courses according to the total number of subscribers (X-axis).

Table 1. Total subscribers count interval per Udemy business and finance courses.

Subscribers Count Interval	Number of Courses
0–5000	1119
5000–10,000	52
10,000–15,000	14
15,000–20,000	6
20,000–30,000	5
50,000–70,000	3

We propose the following Equation (2) to calculate the weight value:

$$W_i = \sum_{j=1}^k f_{ij}, \quad (2)$$

where f_{ij} is the relative frequency of a certain factor related to a web-resource category or topic, for instance, user interaction, total time spent for page visit, page load time, views, reviews, subscriptions, and comments left. We might also face an issue when the W_i equals zero. Then the user-engagement is equal to zero as well. This might happen in rare cases, because if the U_i has a high value without multiplying it by the weight coefficient, it is very likely that W_i will be greater than 0.

As mentioned above, the output range of the user-engagement score is dependent on the arguments V_i and R_i ; in order to standardize it, adjusted to the range from 0 to 1, and receive a user-engagement score value for a topic i , we propose the following simple Equation (3) below:

$$U_{standard_i} = \frac{U_i}{\max(U_i)}, \quad (3)$$

where we get the ratio of a certain category user-engagement to the max user-engagement scores.

4.2. SLO/SLA and the Obtained User-Engagement Score Correlation

Having the user-engagement score calculated, the service level objectives and agreements definition procedure gets improved. By calculating the appropriate user-engagement scores for educational or e-business project components, we can filter these components according to particular criteria, which will let the web project parties find the aggregated user-engagement values (mean, median, percentiles) only for those units that interest us most, and compare them with other groups. Also, based on multiple entities of the user-engagement scores, we can find the overall scores using statistical and mathematical functions and monitor them over time. Based on the values obtained from these computations and SLOs/SLAs data, we propose a correlation matrix to show how the user-engagement score can affect them and the technical equipment, and facilitate its assessment. Thus, applying technical condition assessment is necessary to increase the availability of an e-commerce or e-learning project, especially when there is a high demand for it. Using the SLO and the SLA allows us to assess the infrastructure condition and make objective decisions about the application architecture. We should take into account that the SLO setting process requires a preliminary system and potential infrastructure risks understanding as well as performing user-behavior analysis.

The following is a contingency table that helps to quantify user-engagement score intervals frequency with data about the SLO/SLAs adherence. The user-engagement score can be generic for the whole platform, as well as applying for individual web project components. Based on the contingency table data, we can find the association between a user and the online–resource interaction, how positive it has been, and what might be improved. For example, if the SLO for availability is defined, which is directly related to the technical equipment condition, it is profoundly observable how it associates with a certain web–resource user-engagement or interaction. It is thus possible to display the cases when the SLO or the SLA has been high (Table 2), even though the user-engagement has not been or vice versa.

Table 2. Contingency table for the SLO/SLA data and user-engagement scores correlation.

SLO/SLA	User-Engagement Standardized Score					Total
	0–0.25	0.25–0.5	0.5–0.75	0.75–0.95	0.95–1	
0–25%	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}	X_{1+}
25–50%	X_{21}	X_{22}	X_{23}	X_{24}	X_{25}	X_{2+}
50–75%	X_{31}	X_{32}	X_{33}	X_{34}	X_{35}	X_{3+}
75–95%	X_{41}	X_{42}	X_{43}	X_{44}	X_{45}	X_{4+}
95–100%	X_{51}	X_{52}	X_{53}	X_{54}	X_{55}	X_{5+}
Total	X_{+1}	X_{+2}	X_{+3}	X_{+4}	X_{+5}	X

The relative frequencies usage is also recommended for a clear contingency table results interpretation. We developed a 3×3 matrix based on the SLO/SLA compliance levels and user-engagement score. The levels (low, moderate, high) are custom for each web project. In our case we define the following ranges (0–0.05), (0.05–0.65), (0.65–1) as low, moderate and high, correspondingly. The SLO/SLA levels depend on their definition documents. For instance, the current published target for Google Compute Engine availability is 99.95% availability [35]. If this target is met, we define the SLO/SLA level as high (Table 3).

Table 3. Interpretation table for the SLO/SLA and the user-engagement score levels correlation.

	Low User-Engagement	Moderate User-Engagement	High User-Engagement
Low SLO/SLA	Poor technical equipment condition. Critical infrastructure problems might arise.	Users face potential issues. Web-service needs to be improved, and the logs might be analyzed.	Users face many issues during interaction with a service. The bounce rate (single page visit) is likely to be high.
Moderate SLO/SLA	A service can face potential issues and problems. Even with a low load, the SLO/SLA is not high enough to be met.	Changes and improvements to the online-resource are necessary as well as user-monitoring implementation.	A service has a high throughput and users actively interact with it. The technical conditions need to be improved.
High SLO/SLA	A service met its objectives and is reliable, however, due to low user-engagement, it is quite difficult to predict online-resource behavior, and the SLO/SLA levels themselves when the load increases.	A system infrastructure/network/technical component performs well. The user-interaction needs to be improved and its future increase outcomes must be observed.	A service works well, users actively interact with a platform. The technical condition meets the desired requirements, so features and improvements can be added to maintain high scores.

Accordingly, the higher the user-engagement score level, the more active users interact with the platform, the more they expect this platform to meet the SLAs. If the user-engagement score level is low and the SLO/SLA is not met, this is a sign of a web project with critical problems, even with hardware-related ones. If the user-engagement score level is low, but the SLO/SLA is met frequently and is determined by our table as high, the service and equipment under certain conditions are quite reliable and cope with the load, but with user flow and load increase, it is difficult to predict the web service operations performance. In this situation, we recommend testing the platform and executing increased user flow simulations.

A moderate user-engagement and the same average SLO/SLA levels indicate minor problems with technical equipment and infrastructure, as well as the need to improve the web project user interaction. With high SLO/SLA adherence stats and a top user-engagement score, the technical equipment works with stability, the development team can test and release new features for the web project and gradually add improvements, making it more attractive to users than before.

Based on the Udemy business and finance courses dataset, the user-engagement score was calculated, and obtained values are represented in a table view (Table 4).

Table 4. The standardized user-engagement score distribution for Udemy business and finance courses.

Standardized User-Engagement Score Interval	Value
(0–0.1)	1165
(0.1–0.2)	20
(0.2–0.3)	7
(0.3–0.4)	2
(0.4–0.5)	0
(0.5–0.6)	3
(0.6–0.7)	1
(0.7–0.8)	0
(0.8–0.9)	0
(0.9–1)	1

To compute it, we used such parameters as the number of subscribers, the number of users who left feedback on the course, and the number of published course materials. We can see that only a small number of Udemy courses (2.8%) crossed the standardized user-engagement score of 0.1, where the minimum and maximum values are 0 and 1, respectively. In these courses with the value above 0.1,

the duration ranges from 1 to 6 h; more than 61% are paid, the price varies between \$150–200, while the minimum price among this course category is \$20, and the mean is \$120; the expertise level for this category is mostly beginner or suitable for all levels, although some of them have the intermediate expertise level.

Thereafter, we found that courses with a low user-engagement score (<0.1) not only have few subscribers and feedback but also focus on all levels or beginner level as well. The duration of these courses usually varies from 1 to 3 h, meaning the courses are not long-term, and most of them are paid (92.87%), the prices vary from \$20 to \$200, while the median and mean are \$102 and \$104, respectively. We used Pearson's correlation coefficient to study the association among user-engagement scores, the number of subscribers, and the number of reviews of a specific course. We obtained a very robust association between the score and the number of reviews (1), and a strong association between the score and number of subscribers (0.77). Also, a correlation between the number of subscribers and the number of reviews exists (0.79) and is strong as well (Figure 3).

Figure 3. Pearson's association coefficients and plots for the user-engagement score, number of subscribers and reviews to the Udemy Business and Finance courses.

This strong association between user-engagement score and reviews/subscribers count exists because they are used as the main arguments for the equation. We consider correlation of other dataset features with the obtained user-engagement score, so that more insights can be received. To study the dataset covariance and obtain the most critical features for dimensionality reduction, we performed Principal Component Analysis (PCA). We used the singular value decomposition method, which examines the covariances between individuals in the dataset. This statistical method allowed us to simplify the correlation observation and define features that have the highest value for web-projects user-engagement and popularity. Figure 4 shows that the first dimension (PC1), with eigenvalue and variance out of all principal components equals 35%. We observe that such features as the number of subscribers, number of reviews, weights, user-engagement, boolean value for free/paid entity, and content activeness time (how long a course is available on the platform) are included in this component.

Figure 4. PCA squared coordinates correlation plot.

The PCA biplot (Figure 5) proved again that user-engagement, the number of reviews, and the number of subscribers are positively correlated. We also found that the higher the price, the greater the number of published lectures, and the longer the course duration, which is obvious. Moreover, the time since a course is available correlates with the user-engagement score and corresponding to its values. We assume that the longer a web-project content is available, the larger the chance to obtain high user-engagement and many subscribers/reviews. A negative correlation between the course payment option (free/paid) and the number of subscribers, which has significance for the user-engagement score, is examined. Correspondingly, if a web-project content is free of charge, it has a higher probability to obtain high user-engagement and attention than the paid ones.

Figure 5. PCA biplot of business and finance course generated numerical features.

The above analyses can help to obtain new knowledge about the data and determine types of web-project content that have high user-engagement and are attractive for customers. The main features that correlate with the user-engagement score are defined. The PCA results can be applied for the development of predictive machine learning models to solve various tasks in the fields of e-learning and e-commerce.

5. Conclusions

We conclude that commerce and educational project representatives, who still do not use electronic and EdTech resources, experience many losses during the period of the urgent need for digitalization and remote working/teaching activities. The article provides a framework for improving technical equipment reliability and availability, and detection of insufficient resource allocation, which can lead to profit, users, or customers' loss, and harms business competition, especially during a crisis. We propose the application of user-engagement and Site Reliability Engineering tools with the concept of Service Level Objectives/Service Level Agreement in an efficient way using real-time monitoring, due to the fact that it allows organizations to make web-project infrastructure observable and achieve data-driven decision making. The presence of operational intelligence and performance monitoring is necessary for data research and to provide high-quality service in the remote work and learning modes. We claim that log management of a web-project facilitates efficient Service Level Objectives definition as well as the possibility of task automation in the future with intelligent methods.

In this article, we provided an equation for the user-engagement score calculation, which was applied to the Udemy business and finance educational content dataset. Our user-engagement score is valuable for determining user-behavior and learning trends from topics with their different values. The developed contingency table will simplify the study of the relationship between SLO/SLA adherence and user-engagement data. Accordingly, to calculate user-engagement, we propose to use more than one metric, as well as to use weights that are independent and reflect a specific web project unit in its total spectrum (relative frequency, percentage, ratings). We would like to pay special attention to the need for user rating score presence in the dataset, which can broaden the user-engagement study as well as its correlation with the obtained SLO/SLA. It is necessary to develop a strategy for collecting the necessary business/machine data, as well as a business plan to determine the desired SLI thresholds so that the SLO/SLA calculation can be done in an efficient manner.

The limitation of the study is the analysis of static data generated in 2017 when the demand for web-educational content was high but not so great as opposed to the 2020 period of lockdown and the remote working and learning modes. Also, we analyzed just the business and finance Udemy educational web-service content, however, the platform contains other popular educational content and user groups to study, where the customer preferences might differ, as well as being on other platforms. Our study can lead to certain kinds of social implications—to an increase in commerce/educational web-project profitability due to being available and reliable as well as to allow people to access specific content. That might also enforce web-projects to adjust the content in order to increase user-engagement and meet the SLOs/SLAs. A tendency for user-friendly web-projects and improvements in zero downtime system [36] can evolve. However, we should also consider that organizations will need to collect more data about users than before, so security and privacy concerns might arise. Correspondingly, the more data that will be collected, the more efficient techniques for data handling and storage need to be developed, and organizations need to adopt these new IT solutions.

In the article, we proposed an SLO/SLA and user-engagement levels matrix that improves the infrastructure technical condition interpretation and speeds up the above-mentioned contingency table formation. Our findings associated with educational/commerce web projects show that factors such as the web project service costs, required knowledge level, and class duration affects user-engagement. A data lake, which will contain raw data, and logs that are of significance to the e-learning and e-commerce strategies, can be developed in further research.

For further studies, we propose to perform Natural Language Processing (NLP) of the text data in the studied dataset and correlation with our user-engagement score. In this way, it will be possible to find out whether there is a relationship between the name or description of an educational web project element and the interaction frequency. The most valuable might be the task to identify which words and phrases are key to the audience interested in educational/commerce web projects by their increase/decrease. Also, hypotheses regarding the ratio of user-engagement score, business metrics, user evaluation, and SLO/SLA adherence can be developed and confirmed or rejected.

Author Contributions: Conceptualization, T.U., S.F. and Y.S.; methodology, T.U. and S.F.; validation, T.U., S.F. and Y.S.; formal analysis, T.U., S.F. and Y.S.; investigation, T.U., S.F., T.P. and Y.S.; resources, T.U., S.F., T.P. and Y.S.; data curation, T.U., S.F. and Y.S.; writing—original draft preparation, T.U., S.F., Y.S. and T.P.; writing—review and editing, T.U., S.F., Y.S. and T.P.; visualization, T.U.; project administration, S.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research is supported by National Research Foundation of Ukraine within the project “Methods of managing the web community in terms of psychological, social and economic influences on society during the COVID-19 pandemic”, grant number 94/01-2020.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bojović, Ž.; Bojović, P.D.; Vujošević, D.; Šuh, J. Education in times of crisis: Rapid transition to distance learning. *Comput. Appl. Eng. Educ.* **2020**, *28*, 1467–1489. [\[CrossRef\]](#)
2. Benta, D.; Bologa, G.; Dzitac, S.; Dzitac, I. University Level Learning and Teaching via E-Learning Platforms. *Procedia Comput. Sci.* **2015**, *55*, 1366–1373. [\[CrossRef\]](#)
3. Moreno, R.; Mayer, R. Interactive multimodal learning environments. *Educ. Psychol. Rev.* **2007**, *19*, 309–326. [\[CrossRef\]](#)
4. Lakshminarayanan, R.; Kumar, B.; Raju, M. Cloud Computing Benefits for Educational Institutions. In Second International Conference of the Omani Society for Educational Technology. Available online: <https://arxiv.org/abs/1305.2616> (accessed on 12 May 2013).
5. Google Trends. Available online: <https://www.sciencedirect.com/science/article/pii/S1877050915015987> (accessed on 2 July 2020).
6. Bashirov, M.G.; Bashirova, E.M.; Khusnutdinova, I.G.; Luneva, N.N. The technical condition assessment and the resource of safe operation of technological pipelines using electromagnetic-acoustic effect. In *IOP Conference Series: Materials Science and Engineering*; IOP Publishing: Bristol, UK, 2020; Volume 734, p. 012191. [\[CrossRef\]](#)
7. Bednarski, L.; Sieńko, R.; Howiacki, T. Supporting historical structures technical condition assessment by monitoring of selected physical quantities. *Procedia Eng.* **2017**, *195*, 32–39. [\[CrossRef\]](#)
8. Belodedenko, S.V.; Yatsuba, A.V.; Klimenko, Y.M. Technical condition assessment and prediction of the survivability of the mill rolls. *Metall. Min. Ind.* **2015**, *7*, 85–94.
9. Błachnio, J. Analysis of technical condition assessment of gas turbine blades with non-destructive methods. *Acta Mech. Autom.* **2013**, *7*, 203–208. [\[CrossRef\]](#)
10. Fedushko, S.; Ustyianovych, T.; Gregus, M. Real-time high-load infrastructure transaction status output prediction using operational intelligence and big data technologies. *Electronics* **2020**, *9*, 668. [\[CrossRef\]](#)
11. Belforte, S.; Fisk, I.; Flix, J.; Hernández, M.; Klem, J.; Letts, J.; Magini, N.; Saiz, P.; Sciaba, A. The commissioning of CMS sites: Improving the site reliability. *J. Phys.* **2010**, *219*, 062047. [\[CrossRef\]](#)
12. Alfian, G.; Ijaz, M.F.; Syafrudin, M.; Syaekhoni, M.A.; Fitriyani, N.L.; Rhee, J. Customer behavior analysis using real-time data processing: A case study of digital signage-based online stores. *Asia Pac. J. Mark. and Logist.* **2019**, *31*, 265–290. [\[CrossRef\]](#)
13. Alfian, G.; Syafrudin, M.; Ijaz, M.F.; Syaekhoni, M.A.; Fitriyani, N.L.; Rhee, J. A personalized healthcare monitoring system for diabetic patients by utilizing BLE-based sensors and real-time data processing. *Sensors* **2018**, *18*, 2183. [\[CrossRef\]](#)
14. Cegan, L.; Filip, P. Advanced web analytics tool for mouse tracking and real-time data processing. In Proceedings of the IEEE 14th International Scientific Conference on Informatics, Informatics 2017–2018, Poprad, Slovakia, 14–16 November 2017; pp. 431–435.

15. Elmsheuser, J.; Legger, F.; Medrano Llamas, R.; Sciacca, G.; Van Der Ster, D. Improving ATLAS grid site reliability with functional tests using Hammer Cloud. *J. Phys.* **2012**, *396*, 032066. [CrossRef]
16. Kamiński, K.; Kamiński, W.; Mizerski, T. Application of artificial neural networks to the technical condition assessment of water supply systems. *Ecol. Chem. Eng.* **2017**, *24*, 31–40. [CrossRef]
17. Su, Z.; Yang, Z.; Zhang, X. Tank gun recoil mechanism technical condition assessment based on improved kernel function SVM. In Proceedings of the IEEE 10th International Conference on Electronic Measurement and Instruments, ICEMI 2011, Chengdu, China, 16–19 August 2011; Volume 4, pp. 361–363. [CrossRef]
18. Bosse, S.; Lehmhus, D. Digital real-time data processing with embedded systems. *Mater. Integr. Intell. Syst. Technol. Appl.* **2016**, 281–300. [CrossRef]
19. Feldmann, A.; Gasser, O.; Lichtblau, F.; Pujol, E.; Poese, I.; Dietzel, C.; Wagner, D.; Wichtlhuber, M.; Tapidor, J.; Vallina-Rodriguez, N.; et al. The Lockdown Effect: Implications of the COVID-19 Pandemic on Internet Traffic. In Proceedings of the ACM Internet Measurement Conference, Pittsburgh, PA, USA, 27–29 October 2020.
20. Google. COVID-19 Community Mobility Report. 2020. Available online: <https://www.google.com/covid19/mobility/> (accessed on 15 December 2020).
21. Canizo, M.; Conde, A.; Charramendieta, S.; Minon, R.; Cid-Fuentes, R.G.; Onieva, E. Implementation of a large-scale platform for cyber-physical system real-time monitoring. *IEEE Access* **2019**, *7*, 52455–52466. [CrossRef]
22. Habeeb, R.A.A.; Nasaruddin, F.; Gani, A.; Hashem, I.A.T.; Ahmed, E.; Imran, M. Real-time big data processing for anomaly detection: A survey. *Int. J. Inf. Manag.* **2019**, *45*, 289–307. [CrossRef]
23. Chen, Y.; Iyer, S.; Liu, X.; Milojicic, D.; Sahai, A. SLA decomposition: Translating service level objectives to system level thresholds. In Proceedings of the 4th International Conference on Autonomic Computing (ICAC'07), Jacksonville, FL, USA, 11–15 June 2007.
24. Melo, C.; Dantas, J.; Fé, I.; Oliveira, A.; Maciel, P. Synchronization server infrastructure: A relationship between system downtime and deployment cost. In Proceedings of the International Conference on Systems, Man, and Cybernetics (SMC), Banff, AB, Canada, 5–8 October 2017; pp. 1250–1255. [CrossRef]
25. Elliot, S. *DevOps and the Cost of Downtime: Fortune 1000 Best Practice Metrics Quantified*; International Data Corporation (IDC): Framingham, MA, USA, 2014; Available online: <https://kapost-files-prod.s3.amazonaws.com/published/54ef73ef2592468e25000438/idc-devops-and-the-cost-of-downtime-fortune-1000-best-practice-metrics-quantified.pdf> (accessed on 15 December 2014).
26. Izonin, I.; Tkachenko, R.; Kryvinska, N.; Zub, K.; Mishchuk, O.; Lisovych, T. Recovery of Incomplete IoT Sensed Data using High-Performance Extended-Input Neural-Like Structure. *Procedia Comput. Sci.* **2019**, *160*, 521–526. [CrossRef]
27. Izonin, I.; Kryvinska, N.; Vitynskyi, P.; Tkachenko, R.; Zub, K. GRNN Approach Towards Missing Data Recovery Between IoT Systems. In *Proceedings of the Advances in Intelligent Networking and Collaborative Systems*; Springer: Cham, Switzerland, 2020; pp. 445–453.
28. Beshley, M.; Kryvinska, N.; Seliuchenko, M.; Beshley, H.; Shakshuki, E.M.; Yasar, A. End-to-End QoS “Smart Queue” Management Algorithms and Traffic Prioritization Mechanisms for Narrow-Band Internet of Things Services in 4G/5G Networks. *Sensors* **2020**, *20*, 2324. [CrossRef]
29. Poniszewska-Maranda, A.; Matusiak, R.; Kryvinska, N.; Yasar, A.-U.-H. A real-time service system in the cloud. *J. Ambient. Intell. Humaniz. Comput.* **2020**, *11*, 961–977. [CrossRef]
30. Kryvinska, N.; Bickel, L. Scenario-Based analysis of IT enterprises servitization as a part of digital transformation of modern economy. *Appl. Sci.* **2020**, *10*, 1076. [CrossRef]
31. Markovets, O.; Pazderska, R.; Horpyniuk, O.; Syerov, Y. Informational support of effective work of the community manager with web communities. *CEUR Workshop Proc.* **2020**, 2654, 710–722. Available online: <http://ceur-ws.org/Vol-2654/paper55.pdf> (accessed on 17 June 2020).
32. Kaminskyj, R.; Shakhovska, N.; Gregus, M.; Ladanivsky, B.; Savkiv, L. An Express Algorithm for Transient Electromagnetic Data Interpretation. *Electron.* **2020**, *9*, 354. [CrossRef]
33. Wurster, L.; Baul, S. Market Share Analysis: ITOM. *Perform. Anal. Softw. Worldw.* **2019**. Available online: <https://www.gartner.com/doc/reprints?id=1-1ZA4D838&ct=200619&st=sb> (accessed on 17 June 2020).
34. Chaudhary, V. Covid-19 & e-learning: Coursera sees massive uptake in courses. April 2020. Available online: <https://www.financialexpress.com/education-2/covid-19-e-learning-coursera-sees-massive-uptake-in-courses/1920127/> (accessed on 6 April 2020).

35. Beyer, B.; Jones, C.; Petoff, J.; Murphy, N.R. *Site Reliability Engineering: How Google Runs Production Systems*; O'Reilly Media, Inc.: Sebastopol, CA, USA, 2016; p. 552.
36. Naseer, U.; Niccolini, L.; Pant, U.; Frindell, A.; Dasineni, R.; Benson, T.A. Zero Downtime Release: Disruption-free Load Balancing of a Multi-Billion User Website. In *Proceedings of the Annual Conference of the ACM Special Interest Group on Data Communication on the Applications, Technologies, Architectures, and Protocols for Computer Communication*; Association for Computing Machinery: New York, NY, USA, 2020; pp. 529–541.

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).