IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 3 March 2024, accepted 15 April 2024, date of publication 25 April 2024, date of current version 22 May 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3393423

RESEARCH ARTICLE

Improving Educators' Search Engine Experience: A Quantitative Analysis of Search Terms

JAVIER LEUNG, (Member, IEEE)
Network for Educator Effectiveness, University of Missouri, Columbia, MO 65201, USA
e-mail: leungj@umsystem.edu

ABSTRACT K-12 educators within the University of Missouri's Network for Educator Effectiveness (NEE) utilize the EdHub Library for professional development (PD), hosting over 500 self-paced activities since 2014. Despite the common use of click-through rates (CTR) to gauge user engagement in search engines, this metric needs to be more accurate due to their lack of ordinality in analyzing search terms. Also, CTR calculations of search terms ranging from 0% to 100% do not represent user engagement with search results that satisfy educators' search goals. This study proposes a model predicting CTR, using *SelectKBest* for feature selection and ranking search terms by the Chi-square statistic. Out of 1,317 search terms, 296 (22.5%) were relevant. After outlier removal of extreme CTR values and one-hot encoding of relevant search terms, two regression models achieved over 0.98 accuracy, categorizing 47 search terms into four groups representing the NEE teacher evaluation system components, PD resources for teachers, Teacher Standards, and PD resources for school administrators. The study suggests that 36 of the most searched terms need optimization to reduce users' cognitive load in search results. Notably, the most searched terms had around 40 search results, oscillating between 0.04 and 0.87 CTR, revealing educators' search priorities and tolerance for browsing search results. Overall, this research contributes ordinality to the search engine dataset, shedding light on educators' preferences and guiding improvements in search result relevance and usability.

INDEX TERMS Educational data mining, information retrieval, teacher professional development.

I. INTRODUCTION

The EdHub Library has been part of Missouri's Network of Educator Effectiveness (NEE) since the fall of 2014. NEE provides teacher evaluation solutions and professional development to school districts in the U.S. states of Missouri, Kansas, and Nebraska, as well as internationally. School district members can access over 500 asynchronous online professional development (PD) organized by Teacher Standards. NEE school districts have access to a video library of best practices in classroom teaching, examples for scoring classroom observations, a catalog of self-paced online modules, copyrighted assessment instruments, journal reflection activities, and annual calibration training sessions for school administrators [1]. The EdHub Library is embedded within the existing classroom observation data collection and reporting tools. The library underwent a significant interface redesign in 2017 to provide educators with improved mechanisms for searching and browsing PD materials by Teacher Standards [2].

The EdHub Library is divided into two clearly defined sections, as shown in Figure 1. In the first section, bookmarks of important content are at the top of the interface to bring users' attention to getting started with EdHub. The second section provides K-12 educators with the ability to search for PD materials using the topic directory, Teacher Standards sitemaps, and the search engine. The EdHub homepage lists all the topics alphabetically with brief descriptions and respective Teacher Standards. The Teacher Standards sitemaps contain all modules listed by topic categories and Teacher Standards that can be filtered by keyword and indicator. The search engine component provides users with a flexible way to search for PD materials across multiple topic categories and Teacher Standards.

SiteSearch360 is the search engine service that organizes all available PD and allows for custom curation of resources that mirror the topic structure of the EdHub homepage. It also allows for the custom creation of search term dictionaries

The associate editor coordinating the review of this manuscript and approving it for publication was Qiang Yang.

FIGURE 1. The EdHub library homepage interface.

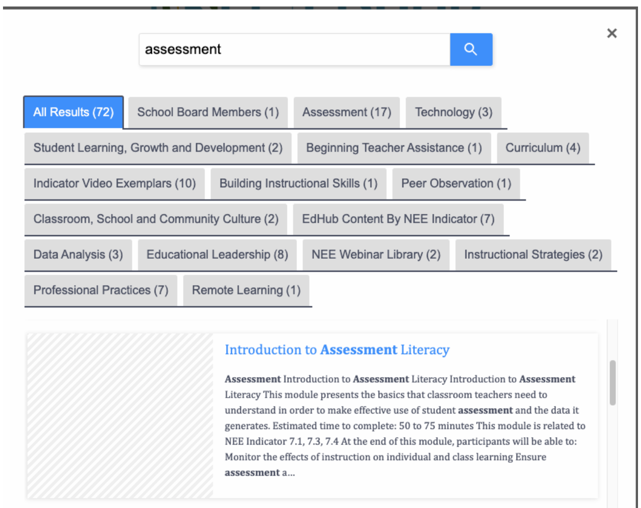


FIGURE 2. The EdHub library search engine interface.

to provide users with more targeted search results across topic categories. Users can perform query searches based on keywords. As shown in Figure 2, for instance, educators can search for PD modules related to assessment, and topic categories organize search results. Users can either review the search results or filter by topic category. Individual search results provide snippets of information about the module to help educators find the PD of interest aligned with a particular Teacher Standard from the search engine interface.

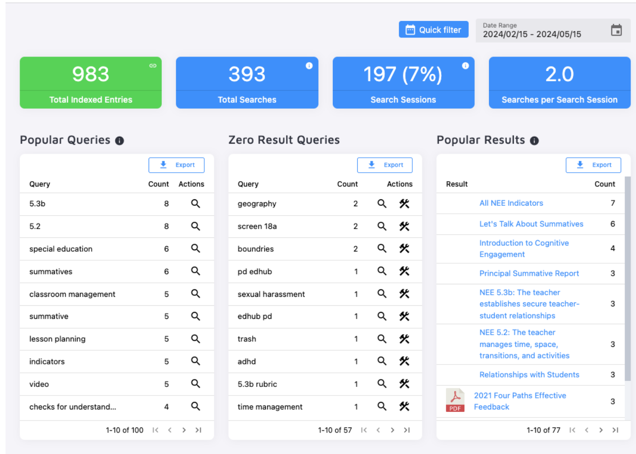


FIGURE 3. Search engine administration interface.

A. SEARCH ENGINE ADMINISTRATION

SiteSearch360 tracks users' queries, total searches performed, the number of times the queries were performed, and click-through rates (CTR), as shown in Figure 3. These variables can be downloaded from the administrative search console and are described in detail in Section III.

B. CLICK-THROUGH RATE AS A KEY METRIC OF USER ENGAGEMENT

In identifying the most used search terms in the SiteSearch360 dashboard, CTR is a standard evaluation metric expressed as a percentage describing how many users have clicked through the search results. CTR values are calculated automatically from the administrative dashboard. However, CTR values are misleading toward misspelled and longer queries. In the first instance of misspelled search terms, these terms are generally reported with a CTR of 0 or 0% user engagement, meaning that users have not engaged with the results provided. The second instance of longer queries is when search terms are shown with 1 or 100% user engagement CTR. Search terms with a CTR of 1 present end users with many results that cause increased browsing without satisfying the user's search goal. Thus, search terms with extreme CTR values present users with either no results due to misspellings in the query (i.e., CTR is 0) or many results when queries are too long (i.e., CTR is 1). The second issue in analyzing this data type is that search terms are nominal features whose values represent categorical descriptions with no particular meaning about their order that are mutually exclusive classes [3].

C. PURPOSE AND RESEARCH QUESTIONS

This study aims to extract and rank relevant search terms used in an online teacher professional development platform. By removing the extreme values of CTR, users' search terms accurately describe users' intent to search for PD materials. This study proposes a method for building a model to predict

CTR rates by selecting essential features in the search engine and users' search terms with one-hot encoding and feature selection techniques. Feature selection techniques generate a list of features ranked from strongest to lowest predictors of the target variable (i.e., CTR).

The study's motivations were to provide a quantitative evaluation method for search terms in educational settings, establish a culture of continuous improvement by periodically evaluating search terms used by educators, and mitigate potential cognitive overload in search results that may impact user experience with PD. The significance of the study is that it allows educational administrators in higher education and K-12 to perform evaluations of teacher PD programs. Program managers can understand educators' search preferences for PD materials on EdHub by understanding search terms' ranking through feature selection. The study's contributions involve a proposed methodology for extracting relevant search terms in a large online community of practice (CoP) delivering targeted search engine results, and identifying patterns in search behavior. The study's results inform program managers about the search terms that require customization of specific search queries that lead to a reduction of users' cognitive load in search results. The study explores the following research questions:

RQ1: What are the characteristics of search terms utilized in CTR and keyword frequencies in the online K-12 environment?

RQ2: What are the top and bottom search terms used in the online K-12 environment?

RQ3: Which features strongly predict click-through rate (CTR) in the online K-12 environment?

The contributions of this proposed methodology involve the following:

- Identify the search terms that require further optimization to reduce users' cognitive load in search results.
- Determine the characteristics of the most search terms based on CTR rates.
- Describe educators' tolerance level for browsing search results.
- Rank search terms to establish importance or ordinality to educators' search priorities through quantitative measures.
- Showcase a practical application of the methodology for cleaning, encoding, and preparing search engine data for algorithmic uses.

The rest of the article is organized as follows. Section II provides a review of the literature and related concepts. Section III describes the proposed methodology for extracting relevant search terms, including exploratory data analysis, outlier removal, one-hot encoding, and feature selection. Section IV provides the results of the analysis organized by the research question. Section V contains the results, optimization of the identified search terms from the analysis, and implications for research and practice. Section VI concludes the article and provides future research directions.

II. LITERATURE REVIEW

The following literature review describes the characteristics of online teacher professional development, origins of the search engine and metrics, and methodologies for analyzing search terms.

A. ONLINE TEACHER PROFESSIONAL DEVELOPMENT

1) BENEFITS OF ONLINE TEACHER PROFESSIONAL DEVELOPMENT

Numerous studies have explored the role of online teacher professional development as a critical factor that impacts student achievement [4], [5], [6]. The core features of effective teacher professional development involve five aspects, including content focus, active learning, coherence, duration, and collective participation, when professional development activities are aligned with teachers' subject matter and the school district's goals and policies [7]. Teacher PD can also influence teachers' career progression and workplace retention [8], [9], [10]. Teachers are more likely to stay in the profession when engaged and motivated in professional development activities and can validate the knowledge that improves their teaching practices.

2) IMPROVING AND UNDERSTANDING EDHUB USERS

Prior studies have used machine learning to improve educators' experiences by analyzing web analytics data from the EdHub Library. Prior to the interface redesign of the library in 2021, the first generation of EdHub posed significant navigation challenges for locating and aligning PD materials to Teacher Standards. Additionally, the analysis of the first generation of EdHub using a clustering algorithm revealed the general access patterns to the site and its resources [1]. The result of the 2018 study showed that principals performed essential activities related to supporting the PD of teachers, including reviewing teachers' journal reflections and accessing examples of teaching practices and teaching videos after reviewing collected data from classroom observation reports. The redesigned the interface of the professional library was designed to help with two critical information tasks to support educators regardless of PD experience: (1) aligning PD materials to Teacher Standards regardless of how educators browse for materials and (2) facilitating such alignment through Teacher Standards sitemaps, homepage directory, and search engine [2].

Leung [11] investigated educators' self-regulated learning behaviors of frequent users and all users using a process mining algorithm before and during the COVID-19 pandemic. Before the pandemic, frequent library users relied more on browsing PD materials through the Teacher Standards sitemaps and less on the search engine. Principals also tended to open browser tabs and leave activities idle while performing principal duties throughout the day. At the onset of the pandemic, library users accessed four topic areas out of the 27 topic categories related to the Introduction to NEE, Beginning Teacher Support, Social-Emotional Learning, and

Cognitive Engagement. The events from all users across the above categories showed how educators accessed PD materials on EdHub to support remote teaching efforts while providing students with social-emotional support and cognitively engaging them via conferencing tools.

B. SEARCH ENGINE ORIGINS AND EVALUATION

1) ORIGINS OF THE SEARCH ENGINE

Search engines are web-based information retrieval systems that search several indexed websites and their databases to match query terms [12]. While several types of search engines exist, search engines rely on three methods for searching text: keyword searching, concept searching, and meta-search engines. Most search engines perform text query and retrieval functions. The second method involves concept searching, where the search engine attempts to determine the meaning behind the queries. The third method relies on multiple search engines to aggregate search results.

Archie was the first search engine created by Adam Emtage, a student at McGill University in Montreal, in 1990, using UNIX commands to look for file directories from hundreds of systems [12]. Though several keyword-based search engines came after Archie in the early 1990s, web crawling became the primary means for search engines to index massive amounts of information on websites in the late 1900s. In 1998, Google became the most successful search engine that relied on the PageRank algorithm to determine the importance of a resource among similar websites. Web pages are ranked based on importance by calculating the probability of random clicks on a link and the quality of links to a page [13].

2) EVALUATING SEARCH ENGINES

Prior studies have investigated how users interact with search engine services in the early days of the search engine using server web logs [14], [15], [16], [17]. Although early search engine keyword analysis relied on the server's log data in earlier studies that required extensive data cleaning, search engines can now automatically calculate performance metrics of individual search queries through the CTR metric that measures how many users clicked on a result after looking up a query. CTR is the total number of clicks on results divided by the number of times the results were shown to the user. CTR data is interpreted from 0 to 1, with 1 being the highest probability of 100% that users were engaged with their query results. Studies pointed out that CTR data from search engines is a valuable feature for measuring the relevance of search terms, users' satisfaction, frustration, and intent with search queries, and predicting user engagement through predictive modeling of CTR [18], [19], [20], [21]. However, relying solely on CTR has limitations in evaluating user satisfaction and relevance. Also, CTR doesn't account for users who don't click any links from search results. The specificity and quality of search queries can significantly influence the CTR metric.

Furthermore, Lewandowski [22] stated that other approaches to evaluating search engines' performance are user satisfaction, click signals, and credibility assessment of search results. User satisfaction stems from user-oriented evaluations of the utility of the information retrieval capabilities of search engines while satisfying users' information needs. Although it is challenging to generalize the search behaviors of users due to the significant variability of information needs, user satisfaction with search engines can be estimated with a click signal analysis in which search queries are analyzed in terms of the first click, the rank of a click, and the fraction of abandoned search queries, including queries without any click. Regarding users' assessment of search results, Kammerer and Gerjets [23] reported that users often rely on the search engine's ability to rank search results without paying attention to the credibility of sources.

C. QUANTITATIVE METHODOLOGIES FOR SEARCH ENGINE ANALYSIS

Although user studies and methodologies in search engines are valuable for understanding search queries and user behaviors, these can be impractical to conduct on a large scale because of the cost of recruiting participants, limited scalability of interventions, intrusiveness, and lack of privacy. For these reasons, quantitative methodologies that use log data and machine learning approaches are better suited to analyze vast amounts of data for the whole population of interest while ensuring data privacy and validity of results.

The standard search engine keyword analysis methodologies are exploratory data analysis, one-hot encoding, and feature selection. These methodologies rely on search engine logs to understand search intent and the effectiveness of search queries.

1) EXPLORATORY DATA ANALYSIS

The goal of exploratory data analysis involves visually and statistically exploring the data to gain insights, identify patterns, and detect anomalies through summary statistics, data visualization, term co-occurrence analysis, temporal analysis, and sentiment analysis [24], [25].

2) ONE-HOT ENCODING

One-hot encoding is used in search engine keyword analysis to represent categorical data, such as terms or keywords, as binary vectors that machine learning algorithms or statistical models can use. One-hot refers to transforming categorical data as binary values to represent values as "hot" (1) and the rest as "cold" (0). The advantages of one-hot encoding include simplicity of implementation and running efficiency. However, one-hot encoding consumes storage, carries little information about similarity and ordinality, tends to overfit with high variance, and fails to generalize unseen data [26]. Even though the limitations of one-hot encoding are noted, one-hot encoding is preferable for nominal over label encoding because nominal data does not carry information on ordinality among categories [27].

3) FEATURE SELECTION AND FEATURE IMPORTANCE

Feature selection involves identifying and choosing the most pertinent and meaningful attributes (or variables) from a larger pool of available features. The objective of feature selection is to enhance the accuracy, efficiency, and interpretability of models or analyses by concentrating on the most informative features and reducing the overall complexity of the data. Feature selection techniques include filter, wrapper, and embedded methods [28]. Filter methods select subsets of variables as a preprocessing step for reducing dimensionality and overfitting by eliminating redundant features and ranking variables in ascending order with top-n features using information gain, correlation measures, Chi-square, and Gini index. Chi-square is commonly used as a filter method to rank predictive attributes of one-hot encoded data and the target variable. Wrapper methods involve forward or backward elimination strategies where predictive variables are progressively incorporated into larger subsets or variables with less predictive power are eliminated. Although filter methods lack a clear cut-off point for ranking features due to the top-n features, wrapper methods can be computationally expensive for identifying predictive subsets [29]. In embedded methods, feature selection is performed during training, and internal parameters are adjusted to determine the weight of each predictive feature. Feature selection is generally done before training a model, whereas feature importance ranks and quantifies the distribution and contribution of principal features to the overall predictive model after model training [30], [31].

III. PROPOSED METHODOLOGY

Although exploratory data analysis, one-hot encoding, feature selection, and feature importance can be used individually to understand search queries and users' behavior, implementing such methodologies individually without understanding the shortcomings of CTR can lead to bias, assumptions, and limited depth of analysis. The proposed methodology aims to leverage these methodologies in a series of steps to understand the characteristics of search engine data by addressing the downsizes of CTR through outlier removal of extreme CTR values before one-hot encoding, feature selection, and feature importance. The proposed methodology follows a feature engineering approach to cleaning, encoding, and preparing the dataset for further algorithmic uses [24]. The proposed methodology aims to identify the search terms relevant to the K-12 educator community. Once ordinality is established among relevant search terms, search terms can be scrutinized individually and modified to reduce cognitive overload in search results.

In the first step, an exploratory data analysis is carried out to determine the dataset's overall distribution and identify outliers for reducing dimensionality. Once outliers are removed during the exploratory phase, search terms are one-hot encoded to create numerical representations while dropping the categorical column (i.e., query). Although converting search terms from nominal to continuous features does not carry any order or hierarchy, the next step involves running a feature selection algorithm as a filter method, using top-n (50) features in *SelectKBest* to establish ordinality among the one-hot encoded search terms and continuous variables that predict the target variable (i.e., CTR)—two attempts aim to reduce dimensionality in the exploratory and feature selection steps. Although data normalization is standard in machine learning, it is not always required because the one-hot encoded data only indicates the presence or absence of a variable [32]. Once the one-hot encoded search terms are identified and ranked by Chi-square scores, two regression tasks are performed to obtain accuracies and root mean square errors. As a final step, program managers use the search engine console to optimize the identified search terms that yield 40 or more search results. Table 1 summarizes the steps of the proposed methodology. The remainder of this section describes the details of the data collection procedures, environment, and Python packages used for exploratory data analysis, one-hot encoding, and feature selection.

TABLE 1. Feature engineering approach and python packages.

Step	Task description	Python package/outcome
1	Exploratory data analysis	Pandas profiling
2	Outlier removal of extreme CTR values	Pandas
3	Visualization of relevant terms after outlier removal	Plotly
4	One-hot encoding of search terms and drop categorical term column	Pandas.get_dummies
5	Rank one-hot encoded search terms by Chi-square to predict CTR	Univariate Feature Selection *SelectKBest*
6	Evaluate regression tasks (accuracy and root mean square error)	GaussianNB and Random Forest Regression
7	Identify search terms to be optimized	Create custom search results in administrative console

A. DATA COLLECTION

A search engine service called SiteSearch360 provides a mechanism for curating and crawling online professional development resources. This study used the dataset of 1,317 search terms from May 2018 to November 2021 to identify five variables: *Query*, *Total Searches*, *Unique Searches*, *Total Search Results*, and *CTR* [33]. The *Query* field is a categorical variable that tracks users' search terms. *Total Searches*, *Unique Searches*, *Total Search Results*, and *CTR* fields are continuous variables that track different aspects of users' search queries. The *Total Searches* variable tracks the number of times a query has been searched for, including non-unique entries. The *Unique Searches* variable collects the number of unique entries, not including the multiple instances of the query performed by the same user. The *Total Search Results* variable is the total number of results displayed to the user when a query is entered. *CTR* is calculated automatically to identify the percentage of users who clicked on a result after looking up a query.

B. EXPLORATORY DATA ANALYSIS AND ENVIRONMENT

In the exploratory data analysis phase, the Pandas Profiling package in Python was used to explore the frequency distribution of *Query, Total Searches, Unique Searches, Total Search Results,* and *CTR* [34]. Bar charts and tables were generated to visualize the most and least searched terms against CTR rates. A total of 836 (63.5%) and 185 (14%) search terms with CTR of 0 and 1 were removed, respectively, leaving a final count of 296 (22.5%) search terms for analysis.

The analyses were performed in Jupyter notebooks using Python 3.7.7. The Pandas Profiling package was used to visualize the distribution of queries across the three continuous variables before removing outliers [35]. Once outliers were removed, Plotly generated heatmap bar charts to explore the characteristics of top and bottom search terms against *CTR* and *Total Search Results* [36]. These heatmap bar charts are particularly useful in understanding which search terms have an overwhelmingly large number of search results presented to users and the level of user engagement with search results.

C. ONE-HOT ENCODING

One-hot encoding is a technique used in search term analysis to represent categorical data, such as search terms or keywords, as binary vectors. It is a way to transform textual data into a numerical format used by machine learning algorithms. The term "one-hot" refers to the fact that each query is represented by a vector of binary values, where only one element is "hot" (1) and the rest are "cold" (0). After removing the CTR outliers, 296 search terms were one-hot encoded as continuous variables to prepare the dataset for the CTR prediction tasks. The Pandas get_dummies method converted the search terms, or categorical attribute, into a dummy continuous variable while dropping the existing categorical *Query* column. Also, data normalization is not applicable because one-hot encoded data are transformed into binary vectors, which would not affect their value. After all, numerical features are centered on the mean with a unit standard deviation [32].

D. UNIVARIATE FEATURE SELECTION

The univariate *SelectKBest* function can still perform feature selection when working with one-hot encoded data to evaluate one-hot encoded search terms based on Chi-square to identify the significant relationships to the target variable [37]. The univariate *SelectKBest* function evaluates the relationships between each feature (i.e., search term) and the target variable (i.e., CTR) and ranks the predictive features by Chi-square. Chi-square scores obtained for each one-hot encoded search term reflect the strength of the association between a search term and CTR.

TABLE 2. Descriptive statistics of relevant search terms after outlier removal.

	Total searches	Unique searches	Total search results	CTR
count	296.00	296.00	296.00	296.00
mean	9.91	8.85	73.09	0.73
std	29.97	27.20	55.05	0.17
min	2.00	1.00	0.42	0.04
25%	2.00	2.00	23.80	0.29
50%	4.00	4.00	61.75	0.50
75%	8.00	8.00	129.07	0.05
max	489.00	445.00	171.00	0.87

The Gaussian Naive Bayes (GaussianNB) and the Random Forest Regression (RFR) were utilized for regression tasks to measure the predictive accuracies and root mean square errors (RSME). The predictive scores were 0.984 and 0.998, while RSME scores were 0.008 and 0.202 for GaussianNB and RFR, respectively. Out of the 50 features, three search engine features (i.e., *Unique Searches, Total Searches,* and *Total Search Results*) and 47 search terms were strong predictors of the CTR target variable. The 47 search terms were ranked based on feature scores to identify their relevance. Unsurprisingly, the three search engine features were the top three most important features since CTR is a percentage calculation of the number of clicks a search term receives by the number of times it is shown to the user. The critical aspect of the proposed methodology is to bring ordinality to search engine data by ranking one-hot encoded search terms by Chi-square scores. The ordinality of search terms can well reflect educators' search preferences for PD materials on the platform.

IV. FINDINGS

The following findings are organized by the research question below.

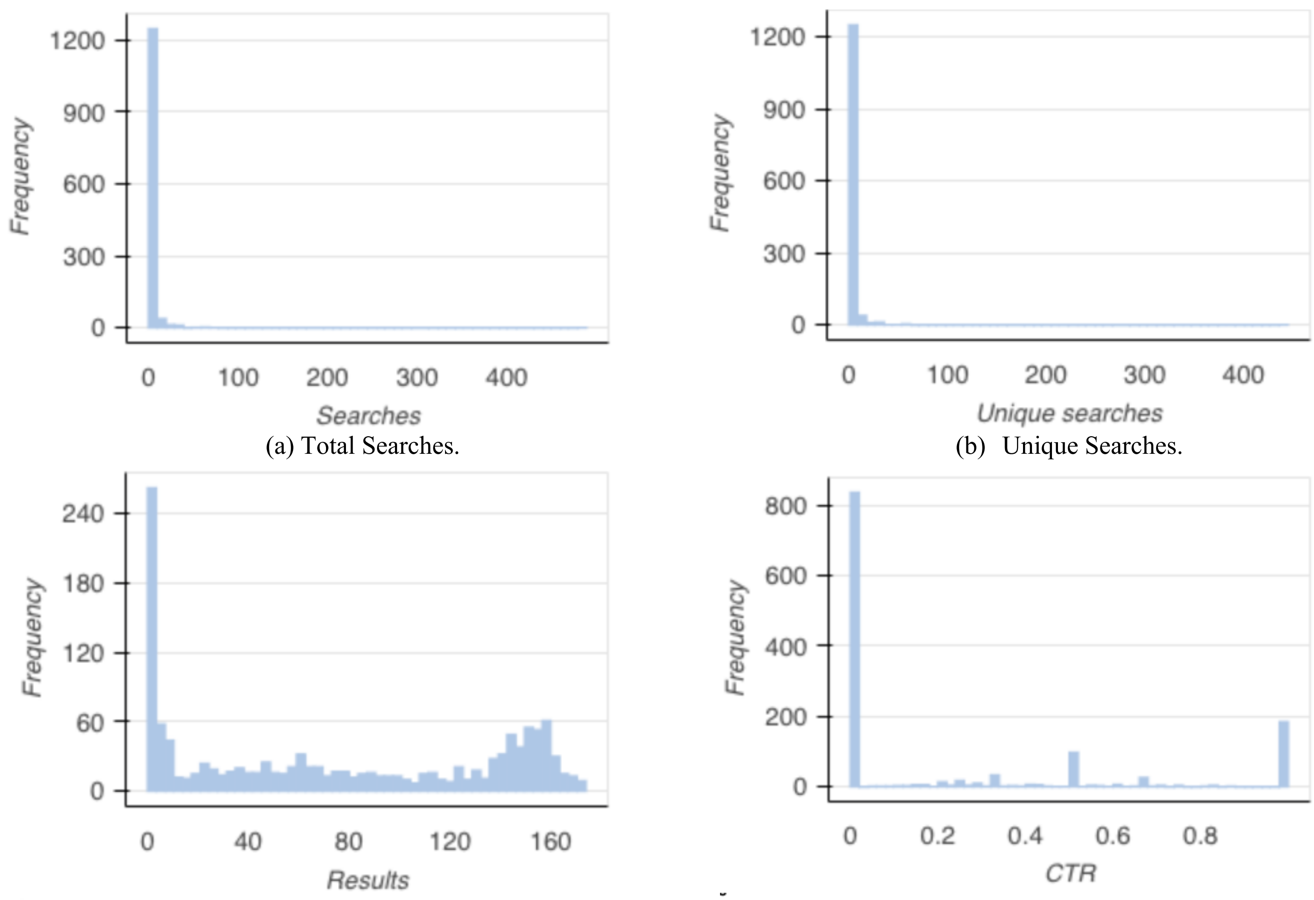
A. WHAT ARE THE CHARACTERISTICS OF SEARCH TERMS UTILIZED IN CTR AND KEYWORD FREQUENCIES IN THE ONLINE ENVIRONMENT?

A total of 1,317 search terms were extracted from the search engine console, including 836 (63.5%) search terms with a CTR of 0, 185 (14%) with a CTR of 1, and 296 (22.5%) with CTR values greater than 0 and less than 1. Figure 4 shows the search term distribution based on CTR before removing outliers.

After removing the search terms with CTR of 1 and 0, 296 search terms were filtered as relevant with an average of 9.91 Total Searches and 0.73 CTR, as shown in Table 2 and Figure 5a. The most frequent search terms were *indicator* (77), *learning* (75), *nee* (56), *scoring* (50), and *rubric* (48), as shown in the word cloud in Figure 5b.

B. WHAT ARE THE TOP AND BOTTOM SEARCH TERMS USED IN THE ONLINE K-12 ENVIRONMENT?

As shown in Figure 6, the most searched terms (e.g., *ethical, superintendent*, and *principal evaluation*) with CTR above



(a) Total Searches. (b) Unique Searches. (c) Total Search Results. (d) CTR.

FIGURE 4. Distribution of search terms prior outlier removal.

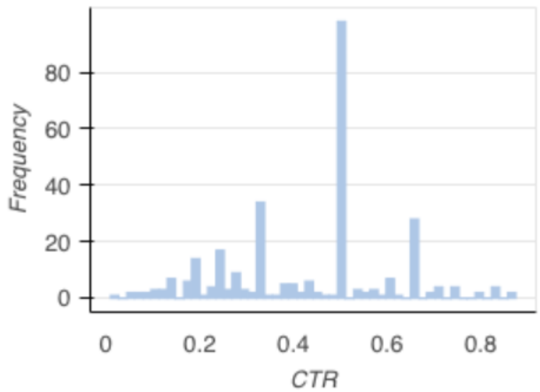
0.80 are in blue and purple shades that presented users with approximately 20-40 search results, indicating a high level of user engagement with the search results. Figure 7 shows the least searched terms (e.g., *critical thinking skills*, *indicator 4.1*, and *1.1*) with a CTR below 0.25 in an orange shade, indicating the low level of user engagement when presented with an overwhelmingly large number of search results. However, it is essential to note that search terms such as *exemplar, cooper*, and *remote* were the least searched terms but displayed search results between 20-40 without satisfying users' information needs, as shaded in blue and purple in Figure 7.

C. WHICH FEATURES STRONGLY PREDICT CLICK-THROUGH RATES IN THE ONLINE K-12 ENVIRONMENT?

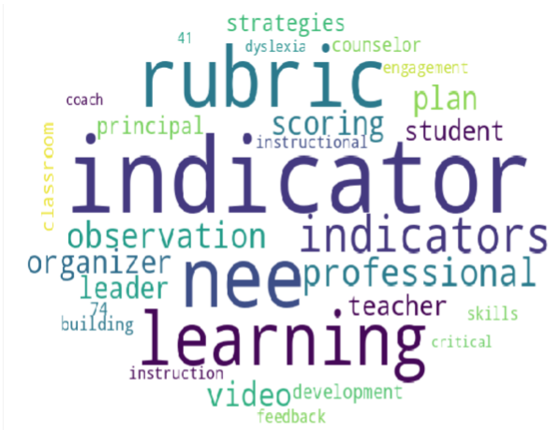
The feature selection using *SelectKBest* showed the highest model accuracies of 0.984 for GaussianNB and 0.998 for RFR using 50 features, including three search engine features and 47 one-hot encoded search terms, for predicting CTR. The three search engine features were the most relevant based on the Chi-square scores, and the prediction scores were as follows: *Unique Searches* (74,890.47), *Total Searches* (66,552.65), and *Total Search Results* (11,528.17). Not surprisingly, the search engine features were relevant since CTR is a calculation using these variables. Table 3 shows the 47 search terms in four groups based on Chi-square for predictive scores and the corresponding continuous variables from the original dataset for further optimization of search results in the administrative console.

V. DISCUSSION

The proposed methodology leverages search engine analysis in a series of steps to filter out misleading search terms with extreme CTR values, one-hot encode relevant search terms and perform univariate feature selection to rank the predictive search terms of CTR as the target variable. After observing search terms by the Chi-Square scores, patterns among search terms emerge, revealing users' search patterns. Then, the identified search terms can be mapped to the original search engine dataset to understand which queries are used the most and cause cognitive overload in the *Total Search Results* variable. By understanding the search terms that cause cognitive overload in search results, proactive steps can be taken to improve user experience with the search engine by updating



(a) CTR distribution.



(b) Most frequent search terms.

FIGURE 5. **CTR distribution and search term frequency after outlier removal.**

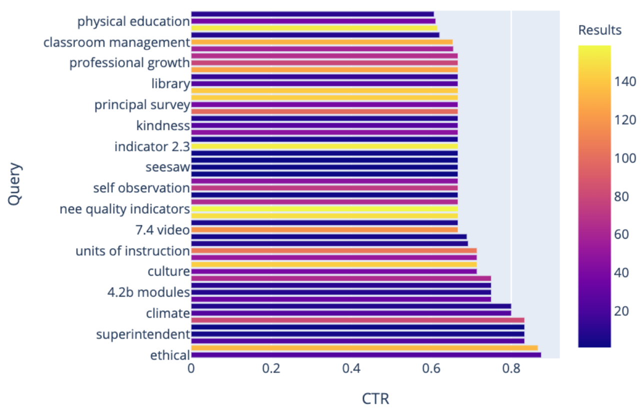


FIGURE 6. **Search terms with high user engagement.**

search term dictionaries and creating custom search results in the search engine administrative console.

In the context of online teacher PD, it is critical to remove search terms with extraneous CTR values (i.e., 0 and 1) that do not represent users' search intent. After removing the CTR outliers and performing univariate feature selection, search terms were related to the NEE teacher evaluation system components, PD resources for teachers, Teacher Standards, and PD resources for school administrators. The proposed methodology comprehensively eliminates biased

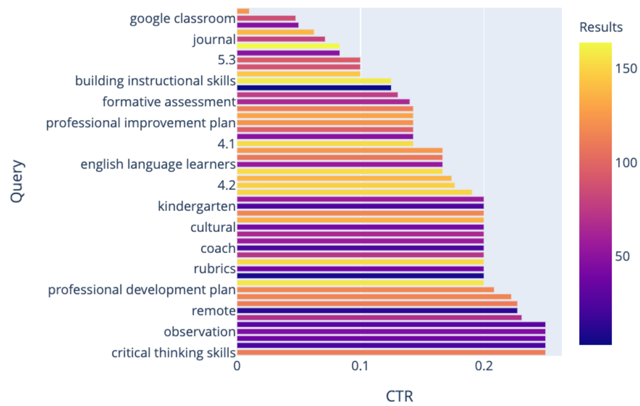


FIGURE 7. **Search terms with low user engagement.**

search terms caused by the shortcomings of the CTR metric. Also, it serves as the foundation of machine learning classification tasks that have the potential to classify search terms based on the types of resources and navigation functions in EdHub.

The most searched terms, shaded in blue and purple, were characterized by around 40 search results with a maximum of 0.85 CTR, as shown in Figure 6. This exploratory finding of the most searched terms showed that users were more likely to browse search results when their queries yielded around 40 search results or less. This finding shows the tolerance level among educators willing to find particular PD materials that satisfied their search goals. However, a few of the most searched terms (e.g., *teacher professional development plan* and *4.2 activities*) are shaded in orange and yellow, indicating low user engagement when search results were above 100 in Figure 6. Regarding the least searched terms, as shown in Figure 7, these search terms yielded above 100 search results with a maximum of 0.25 CTR, shaded in orange and yellow, indicating a low user engagement with search results. This finding shows educators' unwillingness to browse the search results provided due to extraneous cognitive load in search results.

Based on the Chi-square scores in Table 3, search terms appeared to be ranked in four distinct groups, revealing educators' search priorities when using the search engine. In the first group, the search terms involved broad searches about Teacher Standards and rubrics, video exemplars, and PD resources for critical thinking, dyslexia, culturally responsive teaching, and technology integration. In the second group, the search terms were more specific regarding Teacher Standards (4.1, 4.2, 4.2b, 5.1, and 7.4), teacher evaluation components, units of instruction, and subject-specific PD resources related to the journaling component. In the third group, the search terms were associated with PD resources for school administrators and specific modules for building instructional skills and stimulating critical thinking. In the fourth group, the search terms were related to video examples for Teacher Standard 4.1, examples of classroom observation scoring, and formative assessment modules.

TABLE 3. Search terms in four clusters ranked by feature score.

Rank	Search Term	Total searches	Unique searches	Total results	CTR
Rank 1 1,052.00	***critical thinking***	65	62	**62.18**	0.50
	indicators	63	59	**131.58**	0.42
	pdp (professional development plan)	72	61	**116.12**	0.20
	professional development plan	67	59	1.00	0.35
	rubric	34	33	**43.76**	0.29
	scoring rubric	24	23	**49.50**	0.08
	technology	37	30	39.70	0.48
	video	31	26	**62.80**	0.41
	5.3b (secure teacher-child relationship)	21	20	**149.95**	0.19
	counselor	28	28	9.82	0.60
	crt (culturally responsive teaching)	26	13	1.30	0.46
	dyslexia	489	445	3.77	0.68
	nee indicators	19	19	**144.47**	0.31
Rank 2 525.50	***4.1*** (instructional strategies)	35	32	**152.74**	0.14
	4.2b (technology use)	35	32	**140.97**	0.25
	7.4 (effects of instruction)	52	42	**109.88**	0.26
	classroom management	52	49	**130.90**	0.65
	dyslexia and learning	36	31	**89.16**	0.38
	feedback	29	29	**59.00**	0.65
	instructional coach	30	29	**148.63**	0.40
	remote learning	36	30	**66.02**	0.33
	rubrics	30	29	39.16	0.20
	tpdp (teacher professional development plan)	29	24	8.00	0.62
	remote	22	12	12.72	0.22
	special education	23	21	**77.17**	0.13
	student engagement	22	21	**110.86**	0.22
	student survey	23	19	**136.30**	0.17
	4.2 (use of instructional resources)	17	14	**147.94**	0.17
	5.1 (affective engagement)	17	15	**160.88**	0.29
	evaluation	20	4	**47.95**	0.05
	physical education	18	17	37.33	0.61
	short term professional growth plan	20	20	**123.00**	0.01

TABLE 3. *(Continued.)* Search terms in four clusters ranked by feature score.

	webinar	18	10	11.11	0.38
	journal	14	14	**80.07**	0.33
	units of instruction	14	12	**104.85**	0.71
Rank 3 350.00	***5.2*** (management of time, space, transitions, and activities)	13	12	**153.61**	0.61
	cognitive engagement	13	12	**68.00**	0.23
	leader indicators	15	14	**135.93**	0.26
	mentor	13	13	7.69	0.69
	teach like a champion	15	15	**135.06**	0.86
	teaching critical thinking skills	15	12	**98.86**	0.40
	building instructional skills	16	12	**158.31**	0.12
	cooperative learning	16	8	**137.12**	0.06
	instructional strategies	16	13	**154.43**	0.31
Rank 4 209.60	***4.1 video*** (examples videos of instructional strategies)	35	32	**152.74**	0.14
	classroom observation scoring rubric	11	10	**75.63**	0.27
	formative assessment	9	7	**64.66**	0.14

Although the 47 search terms identified reflect the educators' search priorities, 36 search terms (Table 3 in bold) that contained higher than 40 search results need to be optimized to reduce cognitive load on users by curating specific instructional modules from the search engine administrative console and triggering a suggestion from the search results page to indicate an alternative way to narrowing the search task using the Teacher Standards sitemaps. It is important to note that there is no clear cut-off regarding the ideal number of search results. The selection of roughly 40 search results as the ideal tolerance level for browsing search results was because the top search terms exhibited increased user engagement when the CTR metric was above 0.5 as the median CTR, as shown in Figure 4(d). Interestingly, users relied on generic search terms using one to five words to perform search queries.

A. IMPLICATIONS

The study has implications for research and practitioners of standards-based PD. First, this research shows search behaviors and levels of success in finding PD materials that address educators' PD needs. This research also builds the foundation for artificially intelligent search engine tools that make

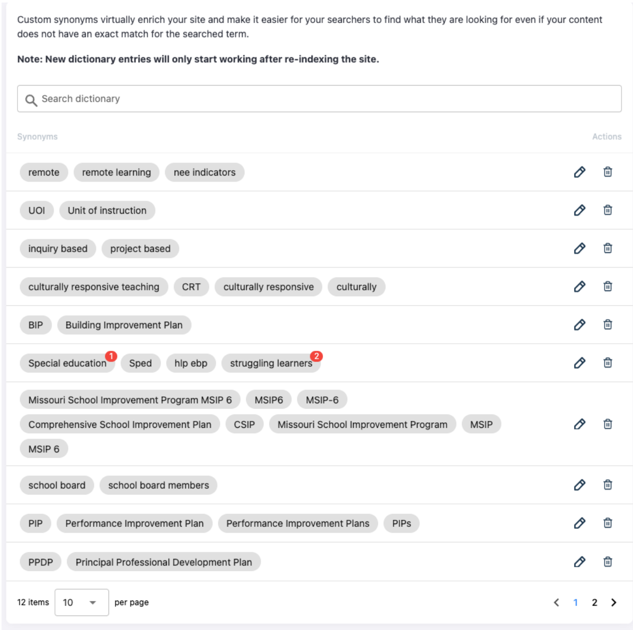


FIGURE 8. **Custom dictionary from the administrative console.**

inferences about users' information goals based on the distinct clusters found in the predictive model of CTR. Second, the results allow PD developers to develop custom dictionaries that generate search recommendations for generic terms and classroom observation evaluation terminologies. Third, 36 search terms with significant search results are a foundation for a data-driven approach in evaluating and improving the search engine function for the most utilized search terms. Thus far, a dictionary of synonym terms was created to address acronyms to develop more targeted search results from the administrative search engine console, including the *Principal Professional Development Plan* (PPDP), *Teacher Professional Development Plan*(TPDP), and *Culturally Responsive Teaching* (CRT). The present dictionary will be expanded based on the 36 search terms to reduce cognitive load in search results. Figure 8 shows an example of the custom dictionary for classroom observation evaluation terminologies created to date.

B. LIMITATIONS

The study presented a limitation regarding the grammatical correction feature when interacting with the search engine. The grammatical correction feature in the search engine is disabled to reduce the large number of search results. Even with the grammatical correction feature turned on, such corrections are not tracked in the search engine dashboard, which may lead to misleading CTR values. Although the proposed methodology included two steps in reducing the dimensionality in the exploratory and feature selection phases, this approach only examined each predictive attribute independently in relation to the target variable based on the Chi-square statistic. It did not analyze mutual information among one-hot encoded search terms.

VI. CONCLUSION

In the analysis of 1,317 search terms, 296 (22.5%) search terms were selected for feature engineering using one-hot encoding and feature selection. Excluded terms comprised 836 (63.5%) search terms with a CTR of 0 and 185 (14%) with a CTR of 1. In predicting CTR based on predictor variables, feature scores showed *Unique Searches*, *Total Searches*, and *Total Search Results* as the top three search engine features, followed by 47 search terms ranked in four groups based on their feature scores. The search terms described in the four groups represented the search priorities of educators when looking for PD materials. Program managers can use this proposed methodology to assess educators' PD priorities and the effectiveness of search terms, leading to appropriate search term customization efforts each school year.

The future directions of the study involve the following:

- Explore educators' navigation trajectories with process mining techniques when interacting with the search engine.
- Examine related or associated search terms through *SelectKBest* mutual information statistics and association rule mining to understand educators' search intentions.
- Develop data ingestion pipelines and end user interfaces to support the automatic ranking of search terms with Streamlit and Tableau.

REFERENCES

[1] J. Leung, "Discovering utilization patterns in an online K-12 teacher professional development platform: Clustering and data visualization methods," *Quart. Rev. Distance Educ.*, vol. 19, no. 3, pp. 17–37, 2018.

[2] J. Leung, "Design features of online teacher professional development: A design case for re-developing the EdHub library to improve usability and alignment of content with teacher standards," *Int. J. Designs Learn.*, vol. 12, no. 2, pp. 79–92, Jun. 2021, doi: 10.14434/ijdl.v12i2.29578.

[3] I. Lopez-Arevalo, E. Aldana-Bobadilla, A. Molina-Villegas, H. Galeana-Zapién, V. Muñiz-Sanchez, and S. Gausin-Valle, "A memory-efficient encoding method for processing mixed-type data on machine learning," *Entropy*, vol. 22, no. 12, p. 1391, Dec. 2020, doi: 10.3390/e22121391.

[4] L. Darling-Hammond and J. Bransford *Preparing Teachers for a Changing World: What Teachers Should Learn and be Able to Do*. Hoboken, NJ, USA: Wiley, 2007.

[5] B. A. Ermeling, "Tracing the effects of teacher inquiry on classroom practice," *Teaching Teacher Educ.*, vol. 26, no. 3, pp. 377–388, Apr. 2010, doi: 10.1016/j.tate.2009.02.019.

[6] J. Lindvall, O. Helenius, and M. Wiberg, "Critical features of professional development programs: Comparing content focus and impact of two large-scale programs," *Teaching Teacher Educ.*, vol. 70, pp. 121–131, Feb. 2018, doi: 10.1016/j.tate.2017.11.013.

[7] L. M. Desimone, "A primer on effective professional development," *Phi Delta Kappan*, vol. 92, no. 6, pp. 68–71, Mar. 2011, doi: 10.1177/003172171109200616.

[8] B. S. Billingsley, "Special education teacher retention and attrition: A critical analysis of the research literature," *J. Special Educ.*, vol. 38, no. 1, pp. 39–55, May 2004, doi: 10.1177/00224669040380010401.

[9] M. Coldwell, "Exploring the influence of professional development on teacher careers: A path model approach," *Teaching Teacher Educ.*, vol. 61, pp. 189–198, Jan. 2017, doi: 10.1016/j.tate.2016.10.015.

[10] E. M. Weiss, "Perceived workplace conditions and first-year teachers' morale, career choice commitment, and planned retention: A secondary analysis," *Teaching Teacher Educ.*, vol. 15, no. 8, pp. 861–879, Nov. 1999, doi: 10.1016/s0742-051x(99)00040-2.

[11] J. Leung, ''Mapping self-regulated events and actions in online teacher professional development with process mining techniques,'' *Quart. Rev. Distance Educ.*, vol. 23, no. 4, pp. 31–62, 2022, doi: 10.6084/m9.figshare.21130299.v3.

[12] T. Seymour, D. Frantsvog, and S. Kumar, ''History of search engines,'' *Int. J. Manag. Inf. Syst. (IJMIS)*, vol. 15, no. 4, pp. 47–58, Sep. 2011, doi: 10.19030/ijmis.v15i4.5799.

[13] L. P. Page, ''Method for node ranking in a linked database,'' U.S. Patent 6 285 999 B1, Jan. 9, 1998. [Online]. Available: https://patents.google.com/patent/US6285999B1/en

[14] M. Chau, X. Fang, and O. R. Liu Sheng, ''Analysis of the query logs of a web site search engine,'' *J. Amer. Soc. Inf. Sci. Technol.*, vol. 56, no. 13, pp. 1363–1376, Nov. 2005, doi: 10.1002/asi.20210.

[15] S. Park, J. Ho Lee, and H. Jin Bae, ''End user searching: A web log analysis of NAVER, a Korean web search engine,'' *Library Inf. Sci. Res.*, vol. 27, no. 2, pp. 203–221, Mar. 2005, doi: 10.1016/j.lisr.2005.01.013.

[16] J. Silverstein, ''Digital reference services: Recommendations for supporting children's informal learning,'' in *Youth Information-Seeking Behavior II. Context Theories Models and Issues*. Lanham, MD, USA: Scarecrow, 2007, pp. 257-277.

[17] A. Usta, I. S. Altingovde, R. Ozcan, and Ö. Ulusoy, ''Learning to rank for educational search engines,'' *IEEE Trans. Learn. Technol.*, vol. 14, no. 2, pp. 211–225, Apr. 2021, doi: 10.1109/TLT.2021.3075196.

[18] B. Carterette, E. Kanoulas, and E. Yilmaz, ''Simulating simple user behavior for system effectiveness evaluation,'' in *Proc. 20th ACM Int. Conf. Inf. Knowl. Manag.*, Oct. 2011, pp. 611–620, doi: 10.1145/2063576.2063668.

[19] H. A. Feild, J. Allan, and R. Jones, ''Predicting searcher frustration,'' in *Proc. 33rd Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, Jul. 2010, pp. 34–41, doi: 10.1145/1835449.1835458.

[20] F. Radlinski, M. Kurup, and T. Joachims, ''How does clickthrough data reflect retrieval quality?'' in *Proc. 17th ACM Conf. Inf. Knowl. Manag.*, Oct. 2008, pp. 43–52, doi: 10.1145/1458082.1458092.

[21] K. Wang, T. Walker, and Z. Zheng, ''PSkip: Estimating relevance ranking quality from web search clickthrough data,'' in *Proc. 15th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Jun. 2009, pp. 1355–1364, doi: 10.1145/1557019.1557164.

[22] D. Lewandowski, ''Users' understanding of search engine advertisements,'' *J. Inf. Sci. Theory Practice*, vol. 5, no. 4, pp. 6–25, 2017.

[23] Y. Kammerer and P. Gerjets, ''Chapter 10 how search engine users evaluate and select web search results: The impact of the search engine interface on credibility assessments,'' in *Proc. Web Search Engine Res.*, 2012, pp. 251–279, doi: 10.1108/S1876-0562(2012)002012a012.

[24] A. Popov, ''Feature engineering methods,'' in *Advanced Methods in Biomedical Signal Processing and Analysis*. Cambridge, MA, USA: Academic Press, 2023, pp. 1–29, doi: 10.1016/B978-0-323-85955-4.00004-1.

[25] P. F. Velleman and D. C. Hoaglin, ''Exploratory data analysis,'' Amer. Psychol. Assoc., Washington, DC, USA, Tech. Rep., 2023, doi: 10.1037/0000320-003.

[26] J. T. Hancock and T. M. Khoshgoftaar, ''Survey on categorical data for neural networks,'' *J. Big Data*, vol. 7, no. 1, pp. 1–41, Dec. 2020, doi: 10.1186/s40537-020-00305-w.

[27] T. Al-Shehari and R. A. Alsowail, ''An insider data leakage detection using one-hot encoding, synthetic minority oversampling and machine learning techniques,'' *Entropy*, vol. 23, no. 10, p. 1258, Sep. 2021, doi: 10.3390/e23101258.

[28] V. Kumar, ''Feature selection: A literature review,'' *Smart Comput. Rev.*, vol. 4, no. 3, pp. 211–229, Jun. 2014, doi: 10.6029/smartcr.2014.03.007.

[29] I. Guyon and A. Elisseeff, ''An introduction to variable and feature selection,'' *J. Mach. Learn. Res.*, vol. 3, pp. 1157–1182, Mar. 2003.

[30] H. Peng, F. Long, and C. Ding, ''Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1226–1238, Aug. 2005, doi: 10.1109/TPAMI.2005.159.

[31] M. Saarela and S. Jauhiainen, ''Comparison of feature importance measures as explanations for classification models,'' *Social Netw. Appl. Sci.*, vol. 3, no. 2, pp. 1–12, Feb. 2021, doi: 10.1007/s42452-021-04148-9.

[32] R. Rofik and N. Hidayat, ''Improving the accuracy of the logistic regression algorithm model using SelectKBest in customer prediction based on purchasing behavior patterns,'' *Future Comput. Sci. J.*, vol. 1, no. 1, pp. 9–17, 2023.

[33] (Jan. 17, 2022). *Search Insights | Site Search 360 Docs. (N.D)*. [Online]. Available: https://docs.sitesearch360.com/search-insights

[34] (Jan. 18, 2022). *Pandas_Profiling API Documentation. (N.D)*. Pandas Profiling. [Online]. Available: https://pandas-profiling.github.io/pandas-profiling/docs/master/index.html

[35] S. Brugman. (2021). *Introduction-Pandas-Profiling 3.0. 0 Documentation. Pandas Profiling*. [Online]. Available: https://github.com/ydataai/ydata-profiling

[36] *Plotly: Low-Code Data App Development (N.D)*. Plotly. Accessed: Jan. 17, 2022. [Online]. Available: https://plotly.com/

[37] (Jan. 17, 2022). *Feature Selection (N.D)*. Scikit-Learn. [Online]. Available: https://scikit-learn.org/stable/modules/feature_selection.html

JAVIER LEUNG (Member, IEEE) a seasoned instructional designer, brings over 15 years of expertise in learning design, asset management, educational data mining, and learning analytics. With a proven track record, he successfully managed 500+ digital assets in the EdHub Library, benefiting 38,000 K-12 educators since 2014. Dr. Leung's proficiency extends to user experience design, machine learning, and natural language processing, showcased in his publications. He actively contributes to the academic community by exploring knowledge structures and professional practices in instructional design. Dr. Leung holds a Ph.D. in Information Science and Learning Technologies (2023) and graduate degrees in Data Science and Analytics (2019) from the University of Missouri, Instructional Technology (2009), and Educational Leadership (2009) from the University of West Florida. Dr. Leung's publications can be accessed at (www.JavierLeung.com).