



MACHINE LEARNING BASED SEARCH ENGINE OPTIMIZATION

Rudramurthy V C¹, Anindita Chakraborty², Prathiksha P³, Sahana S T⁴

Assistant professor, Department of Computer Science Engineering, Global Academy of Technology, Bangalore, India¹

Student, Department of Computer Science Engineering, Global Academy of Technology, Bangalore, India^{2,3,4}

Abstract: In recent years, due to the COVID-19 pandemic, Search Engine Optimization (SEO) has hit an all-time high. As consumers shifted to online shopping in droves, even the most traditional businesses realized they needed to catch up with the digital shift. SEO is important since it enhances the visibility of websites, which provides more traffic and possibilities to convert leads into consumers. SEO may help you figure out what people are looking for on the internet, what answers they're looking for, what terms they're using, and what kind of content they want to consume. The explanations will enable site owners to reconnect with people who are looking for solutions to their problems online. Because most internet users don't look above the first page of search results, the higher a website ranks for a specific term or phrase, the more likely it is to attract new clients. In recent years, practically all website owners have been scrambling to boost their search engine rankings. Business digitization has become a requirement. The situation hasn't changed once the website is built you'll need to implement a variety of digital marketing methods to boost your business's visibility. The purpose of our project Search engine optimizer (SEO) suggester is to help achieve the same. The solution is to build a system that will take as input a websites URL and the users can choose to audit the website or compare the website with the higher ranked website in the domain. The system then produces the summary of what changes can be done in terms of keywords used on the website and changes that can be done in the structure of website and report it to the user.

Keywords: Search Engine Optimization (SEO), Website rank, Search Engine result page (SERP), Website

I. INTRODUCTION

According to Tom M. Mitchell, a computer scientist and machine learning pioneer, machine learning (ML) is a subfield of artificial intelligence, as shown in fig1 from his book Tom M. Mitchell McGraw-Hill.[1]. "Machine learning is the study of computer algorithms that allow computer systems to automatically improve through experience [2]." Working with machine learning is a paradigm change. The core notion is that we give a computer highly curated data, and the machine learns from it, programs itself, and assists us in analysis by grouping data points, making predictions, or extracting significant patterns from our dataset

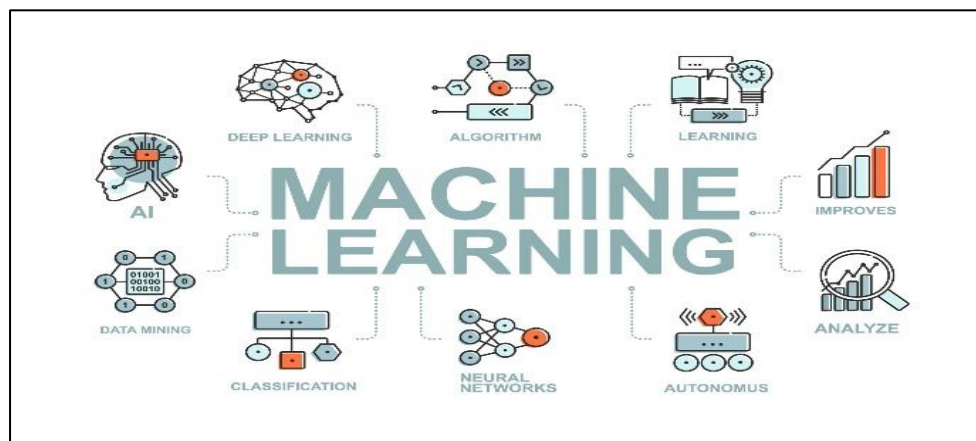


Fig. 1 Machine learning representation

Machine learning can aid SEO by assisting with query comprehension (intent), personalization, and voice search. By using right machine learning techniques time can be saved on tasks like keyword research, technical audits, Content



optimization, Content distribution, Internal linking. It assists us in sifting through large amounts of data to find meaningful insights, taking action, and automating some time-consuming and tiresome jobs [3]

A search engine is a program that employs keywords or phrases to assist users in finding the information they require on the web. When a user puts in a search query, the search engine looks at the page titles, contents, and keywords it has collected, then uses algorithms (step-by-step methods) to build a list of sites, with the most web pages at the top [4]. The art and science of getting pages to rank higher in search engines like Google is known as Search Engine Optimization (SEO). It's a type of digital marketing that refers to tactics aimed at increasing the visibility and number of visits to a website or Web page. In general, the higher a site ranks in the Search Engine Results Page (SERP), the more users it attracts. Search engine optimization (SEO) is concerned with the operation of search engines and the computer-programmed algorithms that govern their behaviour. Users tend to choose one of the top five options presented by a search engine while looking for a service or product online. SEO aims at providing information on how to implement best SEO practices so that a particular website is more discoverable online.

Search engines like Google use an algorithm or set of rules to choose which pages to show for any given query. These algorithms have evolved to be exceedingly intricate in order to determine the rankings of their SERPs, taking into account hundreds or even thousands of different ranking indications, as seen in fig2. However, search engines use four key factors. Links from other websites are extremely important in determining a site's rating in Google and other search



Fig 2: Components of Search Engine Optimization.

engines. The rationale for this is that a link from another website can be interpreted as a vote of quality. Another important SEO aspect is header tags, which comprise title tags, meta descriptions, and more. Make SEO work for you by learning how to design SEO and user-friendly header tags that highlight the relevance and utility of your content. One of the most important aspects of SEO is creating content that is tailored toward the keywords that search engine customers are looking for. Since the web sites are written in HTML, the HTML code's structure may influence how a search engine assesses a page. Site owners can improve their SEO by including significant keywords in their page titles, URLs, headers, and making sure their site is crawlable. [5].

SEO is more than just employing the right keywords and links. There are several types of SEO that may be used to increase website traffic. Technical SEO is exactly what it sounds like: technical factors that affect a website's search engine exposure. The main goal of technical SEO is to make it easy for search engine crawlers to crawl and index a website. On-Page SEO is the optimization of visible page elements like page title, Meta description and on page content for increasing website ranking Off-page SEO is implementation of optimization elements not related to page itself.



II. LITERATURE REVIEW

Employing Search Engine Optimization Techniques in Online News Articles: With the fast growth of internet use in recent years, a vast number of people used search engines to find information. As a result, the goal of a story is to reach readers through search engines as well. Journalists continue to write to be read. According to several surveys conducted over the years, digital technologies are continually changing people's reading patterns, and online journalism is forced to pursue Web traffic. As a result, it can be thought of as a part of journalistic work. Web specialists can change a title or the substance of an article to improve its ranking in Google's search results and raise its chances of getting noticed. The pursuit of the greatest possible ratings and search engine exploitation has become a new reality, a shift in behaviour. (Stabiner, 2011). Many big online news organizations like Daily Mail, British Broadcasting Company (BBC), LATimes have hired SEO consultants to accomplish the task in recent years. SEO can be used in a variety of ways in online journalism. Some have an immediate effect, while others have a delayed effect. Off-page SEO entails attracting many and high-quality inbound links as well as leveraging the power of social media. On-page SEO entails writing titles and headings with keywords that people typically use when searching the internet, using the Meta keywords tag, and using proper HTML, and off-page SEO entails attracting many and high-quality inbound links. New Internet technologies and Web 2.0 have posed a challenge to traditional journalism in recent years by altering critical journalistic labour routines (Spyridou & Veglis, 2008; Chung, 2008). In this essay, we looked at SEO and the possibilities of Web search in the context of online journalism and media websites.[6].

Search Engine Optimization Techniques Practiced in Organizations: A Study of Four Organizations: The study includes an examination and comparison of the target firms' search engine optimization techniques. The goal of the study was to discover the answers to the following questions.

Q1. What are the most frequent SEO techniques utilized by businesses in the United Kingdom and Pakistan?

Q2. What are the benefits and drawbacks of using Search Engine Optimization techniques in UK and Pakistani businesses?

Q3. What are the consequences of online companies in Pakistan and the United Kingdom employing Search Engine Optimization techniques?

The preceding questions were answered using a qualitative approach. Data on SEO strategies was gathered from four Organizations in two countries: Pakistan and the United Kingdom by speaking with professionals. The findings of the interview were supplied in the form of tables and textual information about search engine optimization approaches. These SEO strategies are significant, according to people interviewed in both nations. Directory Submission, Keyword Generation, Link Exchange. The advantages and disadvantages of each SEO tactic used in organizations were discussed. All the above-mentioned questions were answered using the information gathered. According to William Trochim, the validity of interview results is assessed using four principles: credibility, transferability, dependability, and conformability. During study, it was discovered that directory submission and link exchange both require time to list in directories and link to other websites. So, finding strategies to reduce the time spent on directory submission and link building was considered for future project [7].

Search Engine Optimization in E-Commerce Sites: The research focuses on numerous SEO approaches that are used to improve the ranking of websites. The following are the most important variables for good ranking positions in all the major search engines On-page optimization, External links (back links, backward links) and inbound links, Off Page optimization. Problem formulation was done by using an algorithm that considers the number of clicks on a page of an e-commerce site that was visited to look for a product, regardless of whether that product was purchased from that same site. It is determined by the number of clicks received rather than the profit generated by the site or online page. The objective was instead of being based on the quantity of clicks, the site with more profit clicks and better ratings should appear higher in the search engine result pages. The following objective was reached by only counting clicks that lead to checkout or clicks that lead to the basket and are then processed for making an order for a specific product [8]. The role and importance of search engine and search engine optimization: This article focuses on the importance of SEO and the study of search engines. The research focuses on the history of search engines and SEO, SEO flavours such as white hat SEO, grey hat SEO, and black hat SEO, SEO varieties, and SEO advantages [9].

III. METHODOLOGY

The proposed system makes use of various machine learning techniques to create datasets, crawling website to collect metadata, feature selection and training the machine to recognize the domain of the website that is provided as input. The website's metadata is pulled out and the aspects that increase the page rank are assessed. It includes keywords, semantics, and number of inbound links, number of outbound links and number of times a site is visited. These details are collected and produced as an analysis report by the SEO. It also facilitates comparing a website to top websites



which allows owners to draw out the mistakes that they have been making but went unnoticed. The figure 3 shows the system architecture of the SEO that performs website analysis and website comparison as requested by the user. It takes in URL and their request as the input. The output given contains details on the aspects that are currently present in the website considering which the changes can be made by the website owner. The comparison request also takes in URL of the website as the input and compares with the namely top websites of the domain of the input website and gives the detailed report on by how many ranks is the website below the top ones and also the insights on the number of keywords, links and other aspects that is present in the top website which was already stored in the datasets previously. The datasets are fed to the machines using various algorithms using machine learning. The algorithms such as Random

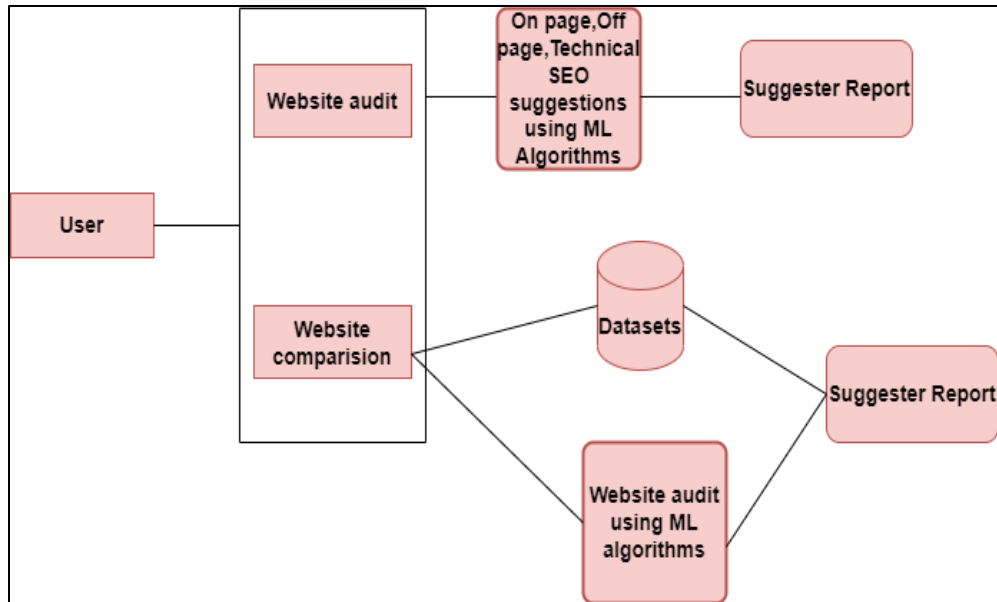


Fig3: System Architecture

Forest, Logistic regression, Multinomial NB, Linear svc are used. The algorithm that gives the best result is considered the final.

IV. CONCLUSION

Virtual technology has become a part of our everyday life. It has become the need of the day. Using technology to raise living standards has become a significant challenge. To any business that functions online the secret key to success is incorporating technologies to market the products to overcome the competition. SEO is such a marketing tool that enables the website owners to present their product in a better and standard ways that converts potential buyers into regular customers. As ssnew technology emerges and evolves, it evolves as well. By boosting traffic, promoting the owner's brand, and engaging with customers, SEO improves the user experience. The use of SEO in a company's digital marketing plan is critical. You'll establish yourself as a dependable supplier if you supply what your customers desire. Building trust boosts SEO, allowing you to grow and strengthen your digital marketing and business over time. When your customers have a positive experience with your business, you will obtain more internet clicks, which will increase traffic and, as a result, enhance your revenues. The most practical and cost-effective way of understanding and contacting clients at vital times is through SEO. As a result, SEO is an effective strategy for increasing brand awareness, building relationships with prospects, and establishing a presence online. The proposed system thus satisfies all the above-mentioned aspects by providing accurate insights on number of keywords, number of inbound and outbound links, semantics and also comparison against the best ranked websites in their respective domains.

REFERENCES

1. For pioneering contributions and leadership in the methods and applications of machine learning. | "Prof. Tom M. Mitchell." National Academy of Engineering. Retrieved October 2, 2011.
2. Machine Learning Definition | Tom M. Mitchell McGraw-Hill (March 1, 1997), Page 1
3. <https://www.searchenginejournal.com/machine-learning-seo/208415/#close>
4. <https://www.bdc.ca/en/articles-tools/entrepreneur-toolkit/templates-business-guides/glossary/search-engine>
5. <https://www.webfx.com/seo/learn/how-seo-works/>



6. Dimitrios Giomelakis , Andreas Veglis – “Employing search engine optimization techniques in online news articles” Redfame Publishing, volume 3, No 1
7. Muhammad Akram, Imran Sohail, Sikandar Hayat, M. Imran Shafi, and Umer Saeed - ”Search Engine Optimization Techniques Practiced in Organizations: A Study of Four Organizations” Journal of computing, Volume 2, Issue 6, (June 2010):134-139
8. Shashank Narayan Srivastava, Suhani Kshatriya, Rajkumar Singh Rathore – “Search engine optimization in E – commerce sites.” International Research Journal of Engineering and technology volume 4 , issue 5 (May 2017)
9. Ayush jain – “The role and importance of search engine and search engine optimization.” International journal of emerging trends and technology in computer science volume 2, Issue 3(June - 2015)
10. Panda/publication/350313250_Search_Engine_Marketing_Does_the_Knowledge_Discovery_Process_Help_Online_Retailers/links/60641b7da6fdccbfea1a6a22/Search-Engine-Marketing-Does-the-Knowledge-Discovery-Process-Help-Online-Retailers.pdf
11. <https://www.outbrain.com/help/advertisers/seo/#:~:text=SEO%20refers%20to%20the%20process,people%20will%20visit%20the%20site.>
12. <https://www.searchenginejournal.com/seo-guide/>