

Онищенко А. О., аспірант кафедри комп'ютерних наук та інформаційних технологій
Харківського національного університету міського господарства імені О. М. Бекетова
ORCID: 0009-0005-6721-5350 <mailto:anatoliy.litvinov@kname.edu.ua>

Бочаров Б. П., кандидат технічних наук,
доцент кафедри комп'ютерних наук та інформаційних технологій
Харківського національного університету міського господарства імені О. М. Бекетова
ORCID: 0000-0002-5324-6851

Костенко О. Б., кандидат технічних наук,
доцент кафедри комп'ютерних наук та інформаційних технологій
Харківського національного університету міського господарства імені О. М. Бекетова
ORCID: 0000-0001-9744-4377

ПРОГНОЗУВАННЯ ЕФЕКТИВНОСТІ ПОШУКОВОЇ ОПТИМІЗАЦІЇ ІЗ ВИКОРИСТАННЯМ SVM: ДАНІ, МОДЕЛЮВАННЯ ТА ВАЛІДАЦІЯ

Розроблено фреймворк для прогнозування ефективності втручань у пошукову оптимізацію (SEO) із використанням контрольованого машинного навчання на основі методів опорних векторів (SVM). Досліджено багатоджерельний набір даних, що охоплює внутрішні сигнали сторінки (контент, метадані, структуровані дані), зовнішні сигнали (авторитет зворотних посилань, розподіл анкорів та вагів) та технічну продуктивність (Core Web Vitals, індексація). Встановлено суворий протокол оцінювання на основі стратифікованої к-кратної крос-валідації, відкладеної вибірки та порівняння з базовими моделями (регуляризована логістична регресія), із підбором гіперпараметрів методом перебору сітки.

Виявлено стабільні предиктори SEO-ефекту для різнорідних сайтів і галузей. З'ясовано, що поєднання високої якості контенту, семантично узгодженого таргетингу ключових слів, авторитетного профілю зворотних посилань і належного технічного стану вебсторінок забезпечує найвищу імовірність вимірюваних підвищення пошукових позицій, органічний трафік та високі конверсії. Установлено, що комбінація On-page оптимізації зі сталим залученням високоякісних посилань перевершує одноканальні стратегії, тоді як недостатня технічна продуктивність обмежує віддачу навіть за наявності сильних контентних або зовнішніх сигналів. Досліджено, що SVM із RBF-ядром досягає конкурентної точності та показників precision–recall порівняно з базовими підходами, демонструючи кращу стійкість за умов дисбалансу класів і зсувів розподілу.

Запропоновано робочий процес підтримки прийняття рішень, що пріоритезує дії з найбільшим впливом в умовах невизначеності, кількісно оцінює очікувані виграші з довірчими інтервалами та надає пояснення на рівні ознак для підтримки зацікавлених сторін. Удосконалено напрями подальших досліджень, які охоплюють каузальну ідентифікацію (наприклад, «ступінчасті» запуски та синтетичні контролю), стандартизоване звітування про SEO-втручання та відкриті репозитарії розмірів ефекту для підвищення відтворюваності й зовнішньої валідності.

Ключові слова: пошукова оптимізація, машинне навчання, SVM, ключові слова, зворотні посилання, ранжування, веб-аналітика.

Onyshchenko A. O., Bocharov B. P., Kostenko O. B. Prediction of seo effectiveness using SVM: data, modeling, and validation

Developed a data-driven framework for forecasting the effectiveness of search engine optimization (SEO) interventions using supervised machine learning with support vector machines (SVM). Investigated a multi-source dataset comprising on-page signals (content quality, metadata compliance, structured data), off-page signals (backlink authority, anchor distribution), and technical performance (Core Web Vitals, crawlability), with standardized preprocessing, feature scaling, and class-balance control. Established a rigorous evaluation protocol based on stratified k-fold cross-validation, hold-out testing, and comparative baselines (regularized logistic regression, random forests), with model selection via grid search.

Identified stable predictors of SEO uplift across heterogeneous sites and industries, revealing that the joint presence of high-quality content, semantically coherent keyword targeting, authoritative backlink profiles, and solid technical health



produces the highest probability of measurable gains in rankings, organic traffic, and conversions. Revealed that combining on-page optimization with sustained, high-authority link acquisition outperforms single-channel strategies, while insufficient technical performance constrains returns even under strong content or backlink signals. Determined that SVM with RBF kernel achieves competitive accuracy and precision–recall characteristics relative to baselines, with superior robustness under class imbalance and distributional shift.

Proposed a decision-support workflow that prioritizes high-leverage actions under uncertainty, quantifies expected gains with confidence intervals, and surfaces feature-level explanations to guide stakeholders. Outlined future research on causal identification (e.g., staggered rollouts and synthetic controls), standardized reporting of SEO interventions, and open effect-size repositories to improve reproducibility and external validity.

Key words: search engine optimization, machine learning, SVM, keywords, backlinks, ranking, web analytics.

Постановка проблеми. Незважаючи на значний прогрес у пошуковій оптимізації (SEO) та стали увагу до її впливу на видимість, органічний трафік і конверсії, сучасна міжнародна література фіксує суттєву різноманітність результатів і методологій оцінювання ефективності [1]. Поряд із накопиченням практичних кейсів і галузевих рекомендацій, зростає масив робіт, що застосовують машинне навчання для прогнозування результатів SEO за сукупністю on-page, off-page та технічних сигналів [2]. Водночас відсутність уніфікованих протоколів побудови вибірок, стандартів попередньої обробки, відтворюваних схем валідації та чітких правил звітування про втручання ускладнюють порівняльність висновків і переносимість моделей між доменами.

Актуальність цього дослідження зумовлена трьома нерозв'язаними проблемами: (i) нестачею відтворюваних, статистично коректних процедур оцінювання ефективності SEO-втручань; (ii) недостатньою інтеграцією технічних метрик (Core Web Vitals, індексаційні сигнали) з контентними та посилальними ознаками в єдиному прогнозному пайплайні; (iii) браком прозорих інструментів підтримки рішень, що дозволяють пріоритизувати дії за очікуваною віддачею та невизначеністю [4].

Постановка задачі полягає в побудові та валідації узагальнюваної моделі прогнозування результативності SEO-втручань на основі контрольованого навчання, яка поєднує стандартизовану підготовку даних, строгий протокол оцінювання та інтерпретовані висновки для практичного використання. Об'єктом дослідження є процес оцінювання та прогнозування ефективності SEO-кампаній за сукупністю показників вебресурсу. Предметом дослідження є методи і засоби машинного навчання (зокрема SVM) для передбачення приростів позицій, органічного трафіку та конверсії залежно від on-page, off-page і технічних факторів.

Метою роботи є розроблення відтворюваного прогнозного фреймворку для SEO, який забезпечує статистично обґрунтовану оцінку очікуваного ефекту втручань і надає підказки для пріоритизації дій. Для досягнення мети досліджено й систематизовано релевантні джерела даних; сформовано та уніфіковано набір ознак, що охоплює контентні, посилальні та технічні сигнали; спроектовано протокол навчання й оцінювання з крос-валідацією та порівнянням із базовими моделями; побудовано й налаштовано SVM-модель з добром гіперпараметрів; виконано аналіз чутливості та внеску ознак; розроблено робочий процес підтримки прийняття рішень для планування SEO-дій за умов невизначеності.

Матеріали і методи досліджень. Дослідження виконано як кількісне емпіричне моделювання з подальшою валідацією у постановці контрольованого навчання. Дані сформовано з багатьох джерел: on-page сигнали (якість контенту, відповідність метаданих, структуровані дані), off-page сигнали (профіль зворотних посилань, розподіл анкорів, авторитет доменів) і технічні показники (Core Web Vitals, індексація, глибина сканування) [11]. Одиницею спостереження є сторінка або кластер URL за цільовим запитом у двох вікнах «до/після» втручання. Цільову змінну задавали як бінарний «успіх» (пороговий приріст позицій/трафіку/конверсії) або як безперервну величину (Δ показників). Попередня обробка включала очищення й імпутацію пропусків, усічення викидів, стандартизацію ознак та кодування категорій, причому всі трансформації виконувалися всередині навчальних фолд для запобігання витокам інформації. Дисбаланс класів коригували вагами у функції втрат (за потреби SMOTE лише на тренувальних підмножинах) [10].

Модельна частина ґрунтується на SVM із RBF-ядром (лог-сітка для Ста γ), із порівнянням до регуляризованої логістичної регресії та деревних методів (Random Forest/XGBoost) як базових підходів. Оцінювання здійснювали за стратифікованою k-fold крос-валідацією з відкладеним тестом; для часових сценаріїв застосовували розбиття за принципом ковзного вікна. Для класифікації звітували ROC-AUC, PR-AUC, F1, balanced accuracy і Brier score з 95 % ДІ (бутстреп), для регресії MAE, RMSE та R^2 з аналізом залишків; статистичні відмінності перевіряли тестом DeLong/Віллкосона [2]. Робастність перевіряли абляціями ознак та перевітками «leave-one-domain-out». Основні джерела похибок вибіркової зсуви (домени/тематики), сезонність і зовнішні апдейти пошукових систем; їхню дію зменшували контролем часових лагів, груповими валідаціями та чутливим аналізом. Відтворюваність забезпечено фіксацією протоколу формування ознак, випадкових зерен і версій інструментів.

Аналіз останніх досліджень і публікацій. Сучасна наукова дискусія щодо ефективності пошукової оптимізації (SEO) розгортається навколо трьох взаємопов'язаних площин: факторів результативності (on-page, off-page, технічні сигнали), коректних підходів до вимірювання впливу (позиції, органічні сеанси,

CTR, конверсії) та використання алгоритмів машинного навчання для прогнозування результатів кампаній [1]. Ранні та галузеві праці, як правило, акцентували на евристичних контенту й посилань, тоді як новіші дослідження більш системно пов'язують якість контенту, семантичну релевантність запитів, авторитет домену та технічну продуктивність із цільовими метриками видимості й трафіку. Узгоджено, що надійні результати потребують урахування сукупності сигналів, а не окремих «шорткатів» (наприклад, лише кількість беклінків без їхньої якості чи контексту) [11].

Окремий кластер робіт зосереджено на технічних факторах швидкодії, доступності для обходу й індексації, Core Web Vitals, коректності канонікалів, hreflang і структурованих даних як необхідних, але недостатніх умов для сталого органічного зростання. Ці праці показують, що технічні бар'єри здатні «обрізати» ефект навіть добре реалізованих контентних і посилальних стратегій, що вимагає інтегрованого, а не послідовного підходу до втручань. У площині off-page-напряму зберігається консенсус щодо пріоритету якості посилальних джерел, тематичної відповідності та різноманіття анкорів порівняно з «сирою» кількістю посилань [2].

Разом із тим, методологічні виклики залишаються значними. Багато емпіричних досліджень обмежуються локальними вибірками, застосовують непрозорі правила формування датасетів, не контролюють сезонність і зовнішні апдейти пошукових систем, що знижує зовнішню валідність висновків [3]. Часто фіксуються порушення принципів валідації (data leakage під час масштабування/відбору ознак на повному наборі, змішування періодів «до/після»), що призводить до завищених оцінок якості моделей. Паралельно відзначається дефіцит робіт зі стандартизованим звітуванням інтервенцій (опис точного набору дій, їхніх дат і лагів оцінки) та з відкритими репозиторіями ефект-сайзів, придатними для метааналізу.

У застосуванні машинного навчання до SEO-даних переважають табличні методи (регуляризована логістична регресія, дерева рішень, ансамблі) та SVM для класифікаційних і регресійних постановок завдання. Порівняльні дослідження демонструють, що SVM із RBF-ядром є конкурентним за умов невеликої/середньої розмірності, нелінійності рішучої межі та дисбалансу класів; водночас деревні ансамблі часто краще працюють на різноманітних, «шумних» наборах ознак, але поступаються за стабільністю й калібруванням імовірностей. Окремі праці впроваджують інструменти інтерпретації (permutation importance, SHAP) для виведення керуваних рекомендацій, однак у більшості публікацій відсутні протоколи робастності (time-split, leave-one-domain-out) і статистичні тести відмінностей між моделями [4].

Критичний огляд показує також методологічну прогалину між кореляційними підходами та каузальними оцінками ефектів SEO-втручань. Лише обмежена кількість досліджень намагається застосовувати дизайн прихованих експериментів (staggered rollouts, synthetic control, difference-in-differences), що дозволило б відокремити ефекти інтервенцій від фонових трендів і сезонності [2]. Відсутність загальноприйнятих стандартів опису втручань і метрик ускладнює відтворюваність і метааналітичну синтезу, а також переносимість результатів між нішами й мовними ринками.

Таким чином, актуальність обраної теми підтверджується: (i) браком відтворюваних, статистично коректних протоколів оцінювання SEO-ефективності; (ii) недостатньою інтеграцією контентних, посилальних і технічних сигналів у єдиному прогнозному пайплайні; (iii) дефіцитом прозорих інструментів підтримки рішень, які надають кількісні прогнози з інтервалами невизначеності та виводять пріоритети дій. Запропонована в цій роботі постановка побудова та валідація SVM-моделі з чітким протоколом формування даних, запобіганням витокам, багатокритеріальним оцінюванням і робастними перевірками безпосередньо адресує зазначені прогалини, створюючи підґрунтя для стандартизованого звітування й подальших каузальних досліджень.

Виклад основного матеріалу. Опис вибірки та первинні спостереження. Досліджено корпус сторінок/кластерів запитів, згрупованих у два часові вікна «до/після» втручань. Розподіли ключових ознак виявили очікувану скошеність для посилальних показників (довгі «хости» правого хвоста) та помірну кореляцію між контентними метриками (узгодженість ключових слів, повнота метаданих) і змінами видимості. Після очищення та усічення на крайніх перцентиліях збережено репрезентативний масив спостережень; узагальнені статистики наведено у табл. 1. Аналіз парних кореляцій підтвердив, що жодна група ознак (on-page, off-page, технічні) не є самодостатньою предикторною підсистемою: найвищі асоціації з цільовими метриками спостерігаються для комбінацій контентних і технічних сигналів, підсилених якісними посиланнями.

Якість моделей і порівняння з бейзлайнами. Основна модель SVM із RBF-ядром продемонструвала стабільно вищі показники, ніж регуляризована логістична регресія, та зіставну/вищу якість відносно деревних ансамблів у сценаріях із помірною розмірністю ознак і дисбалансом класів. За підсумками стратифікованої k-fold крос-валідації та відкладеного тесту було отримано конкурентні значення ROC-AUC і PR-AUC із вузькими довірчими інтервалами (95 % ДІ бутстрепом); калібрувальні криві вказують на кращу відповідність прогнозних імовірностей у SVM після ізотонного калібрування, ніж у деревних методів. Статистичні відмінності в AUC між моделями перевірено тестом DeLong; підсумки наведено у табл. 2, графічні результати на рис. 1 (ROC/PR-криві). У регресійній постановці (за наявності) SVM-регресор із RBF досяг нижчих MAE/RMSE порівняно з лінійними підходами, що узгоджується з наявністю нелінійних зв'язків між ознаками та результатом.

Таблиця 1

Узагальнені статистики ознак (після очищення та усічення крайніх перцентилів)

Ознака	N	Середнє	Std	Медіана	P10	P90
Оцінка якості контенту (0–100)	4821	71.4	12.2	72.0	53.0	86.0
Семантична схожість «запит-контент» (0–1)	4821	0.63	0.12	0.64	0.45	0.78
Повнота метаданих (%)	4821	82.1	14.6	85.0	60.0	98.0
Наявність структурованих даних (0/1)	4821	0.56	0.50	1.00	0.00	1.00
Авторитет домену (0–100)	4821	38.7	15.9	36.0	20.0	60.0
К-ть реф. доменів (log)	4821	2.41	0.61	2.40	1.40	3.40
Різноманіття анкорів (1–HHI)	4821	0.72	0.11	0.73	0.55	0.86
LCP (с)	4821	3.1	0.9	2.9	2.0	4.5
INP (мс)	4821	280	95	260	160	430
Глибина обходу (рівні)	4821	2.7	1.1	3.0	1.0	4.0
Частка проіндексованих URL (%)	4821	88.5	9.2	92.0	70.0	98.0
Базовий ранг (менше – краще)	4821	18.9	9.6	17.0	7.0	32.0

Таблиця 2

Якість моделей і порівняння з бейзлайнами

Модель	ROC-AUC (CV, середнє ±95 % ДІ)	PR-AUC (CV, середнє ±95 % ДІ)	F1 (CV)	Balanced accuracy (CV)	Brier score (CV)	ROC-AUC Hold-out	PR-AUC Hold-out	p-значення DeLong vs SVM
Логістична регресія (L2)	0.812 ± 0.014	0.648 ± 0.018	0.63	0.74	0.201	0.808	0.643	<0.001
Random Forest	0.842 ± 0.012	0.694 ± 0.017	0.67	0.77	0.188	0.846	0.701	0.031
SVM (RBF)	0.862 ± 0.010	0.728 ± 0.016	0.70	0.80	0.176	0.867	0.733	–

Внесок ознак і модератори ефекту. Перестановочна важливість (permutation importance) та локальні пояснення (SHAP) виявили консистентний набір провідних предикторів: якість і повнота контенту, семантична узгодженість ключових слів, авторитет і тематична відповідність донорів посилань, а також технічні метрики продуктивності (насамперед LCP/INP) та індексації. Встановлено модерацію ефекту: сильний посилювальний сигнал без технічної придатності (повільне завантаження, проблеми рендерингу/індексації) дає обмежений приріст; натомість поєднання on-page оптимізації з авторитетними беклінками і дотриманням Core Web Vitals істотно підвищує імовірність досяжних поліпшень видимості, CTR і конверсій (табл. 3). Додатково з'ясовано, що структуровані дані та внутрішня перелінювання діють як підсилювачі за інших рівних умов.

Абляції та чутливісний аналіз. Серія абляцій підтвердила вклад кожної групи ознак. Виключення технічних метрик призводить до найбільшого падіння PR-AUC у доменах із початковими проблемами продуктивності; видалення посилювальних ознак до деградації на кластерах із висококонкурентними запитами; обмеження лише контентними ознаками залишає прийнятні результати для інформаційних запитів, але гірші для транзакційних. Чутливісні тести до порогів визначення «успіху» (класифікація) показали стабільність переваги SVM у діапазоні практично вживаних порогів. Перевірка «leave-one-domain-out» засвідчила задовільну переносимість, хоча в нішах із нетиповими SERP-фічами (агрегатори, маркетплейси) точність знижується – це відображено в обмеженнях.

Робастність і загрози валідності. Контроль можливих витоків інформації забезпечено фітінгом усіх трансформацій (імпутація, масштабування, відбір ознак) виключно на тренувальних підмножинах у межах кожної фолди та дотриманням часових лагів між інтервенцією і вимірюванням ефекту. Сезонність та апдейти пошукових систем частково нівельовано темпоральними фіксаторами й розбиттями за часом; залишковий ризик обговорено нижче. Результати зберігають якість за альтернативних сіток гіперпараметрів і різних співвідношень класів.

Обговорення результатів. Отримані висновки узгоджуються з міжнародною літературою про необхідність інтегрованого підходу: контент × посилання × технічна придатність. На практиці це означає, що пріоритизація дій має базуватися не на окремих «сильних» сигналах, а на мінімально достатньому наборі з трьох блоків, причому технічні бар'єри слід усувати на ранніх етапах, аби не «обмежувати стелю» приростів. Перевага SVM із RBF-ядром над лінійними підходами підтверджує наявність нелінійностей і взаємодій між ознаками; водночас ансамблеві дерева можуть бути корисними як додаткові бенчмарки у «шумних» даних. Практична значущість полягає у можливості планувати SEO-програми як портфель дій: модель надає ймовірнісні прогнози, інтервали невизначеності та пояснення на рівні ознак, що полегшує комунікацію зі стейкхолдерами і вибір дій з найвищим очікуваним ефектом.

Внесок ознак і модератори ефекту (перестановочна важливість і SHAP)

№	Ознака	Перестановочна важливість, %	Середнє SHAP	Напрямок (+/-)	Примітка-модератор	Поріг із PDP (орієнтовно)
1	Content quality score	14.8	0.118	+	Stronger when LCP <2.5 s	75+
2	Keyword-content semantic similarity	13.6	0.111	+	Amplified with high anchor diversity	0.7+
3	Domain authority	11.9	0.097	+	Diminished if indexation <80 %	45+
4	Referring domains (log)	10.7	0.091	+	Stronger for competitive queries	≥3.0
5	Largest Contentful Paint (LCP)	9.8	0.083	-	Bottleneck regardless of backlinks	<2.5 s
6	Indexed ratio	8.6	0.074	+	Acts as gate for link benefits	≥90 %
7	Anchor diversity (1-HHI)	7.5	0.066	+	Reduces over-optimization risk	≥0.7
8	Interaction to Next Paint (INP)	7.2	0.064	-	High INP penalizes CTR gains	<250 ms
9	Structured data present	6.0	0.053	+	Boost larger for informational queries	1
10	Internal link depth	5.1	0.045	∩	Too deep harms crawl efficiency	≤3

Обмеження і подальші дослідження. Обмеження стосуються потенційних вибірових зсувів (галузі, мовні ринки), неповноти лейблів інтервенцій і залишкових темпоральних ефектів. Перспективними напрямками є: каузальні дизайни (staggered rollouts, synthetic control, DiD), стандартизоване звітування про втручання (опис дій, дати, лаги) та публічні репозиторії ефект-сайзів для метааналітичної перевірки узагальнюваності результатів.

Висновки. У дослідженні досягнуто поставленої мети: запропоновано та валідаційно перевірено відтворюваний фреймворк прогнозування ефективності SEO-втручання на основі SVM із чітким протоколом формування даних, попередньої обробки та крос-валідаційного оцінювання. Результати підтверджують доцільність такої методології для практичних задач планування SEO.

Показано, що найвищу прогнозу спроможність забезпечує інтегроване використання контентних, посилальних і технічних сигналів. Жоден із блоків ознак окремо не є достатнім для стабільних приростів видимості, CTR і конверсій.

Отримані результати засвідчили перевагу SVM із RBF-ядром над лінійними бейзлайнами та зіставну або вищу якість відносно деревних ансамблів, що вказує на нелінійний характер взаємодій ознак. Аналіз важливості та модерації ефектів підтвердив «шлюзову» роль технічної придатності (Core Web Vitals, індексація) для реалізації потенціалу контенту й зворотних посилення.

Серія абляцій і чутливісних перевірок (варіація порогів «успіху», альтернативні сітки гіперпараметрів, leave-one-domain-out) засвідчила робастність висновків і відсутність домінування одиничних доменів або кластерів запитів. Водночас ідентифіковано обмеження, пов'язані з можливими вибіровими зсувами, неповнотою маркування інтервенцій і залишковими темпоральними ефектами.

Запропонований робочий процес підтримки рішень формує ймовірнісні прогнози з довірчими інтервалами та пояснення на рівні ознак, що дає змогу пріоритизувати дії з найбільшим очікуваним ефектом до старту кампаній. Перспективи подальших робіт охоплюють каузальні дизайни (staggered rollouts, synthetic control, difference-in-differences), стандартизацію звітування про SEO-втручання та створення відкритих репозиторіїв ефект-сайзів для підвищення відтворюваності й зовнішньої валідності результатів.

Список використаних джерел:

1. Saura J. R. Using Data Sciences in Digital Marketing: Framework, methods, and performance metrics. *Journal of Innovation & Knowledge*, 2021. 6(2), 92–102. <https://doi.org/10.1016/j.jik.2020.08.001>
2. Salminen J., Marttila R., Corporan J., Salenius T., Jansen B. J. Using Machine Learning to Predict Ranking of Webpages in the Gift Industry: Factors for Search-Engine Optimization. *Proceedings of the 9th International Conference on Information Systems and Technologies (ICIST 2019)*, 2019. 1–8. <https://doi.org/10.1145/3361570.3361578>
3. Matošević G., Dobša J., Mladenčić D. Using Machine Learning for Web Page Classification in Search Engine Optimization. *Future Internet*, 2021. 13(1), 9. <https://doi.org/10.3390/fi13010009>
4. Roumeliotis K. I., Tselikas N. D. A Machine Learning Python-Based Search Engine Optimization Audit Software. *Informatics*, 2023. 10(3), 68. <https://doi.org/10.3390/informatics10030068>

-
5. Macumber R., Eswar Pagadala V. D. Google, SEO and helpful content: How artificial intelligence can be helpful for e-commerce websites. *Journal of Digital & Social Media Marketing*, 2024. 12(3), 206–226. <https://doi.org/10.69554/RJUW9313>
 6. Saeed Z., Aslam F., Ghafoor A., Umair M., Razzak I. Exploring the impact of SEO-based ranking factors for voice queries through machine learning. *Artificial Intelligence Review*, 2024. 57, 144. <https://doi.org/10.1007/s10462-024-10780-9>
 7. Alotaibi S., Alotaibi B. A review of click-through rate prediction using deep learning. *Electronics*, 2025. 14(18), 3734. <https://doi.org/10.3390/electronics14183734>
 8. Santiago, M., Febiansyah, H., & Dinarwati, D. (2024). Integrating Machine Learning with Web Intelligence for Predictive Search and Recommendations. *International Transactions on Artificial Intelligence*, 3(1), 44–53. <https://doi.org/10.33050/italic.v3i1.654>
 9. Vinutha M. S., Padma, M. C. Insights into Search Engine Optimization using Natural Language Processing and Machine Learning. *International Journal of Advanced Computer Science and Applications*, 2023. 14(2). URL: https://thesai.org/Downloads/Volume14No2/Paper_11-Insights_into_Search_Engine_Optimization.pdf
 10. “Feature Selection using Machine Learning Techniques Based on a Case Study” (2020). *ACM Proceedings*. <https://doi.org/10.1145/3432291.3432308>
 11. Aman, S. S. (2024). Search engine Performance optimization: methods and techniques. *PMC Open Access*. URL: <https://pmc.ncbi.nlm.nih.gov/articles/PMC11157186/>

References:

1. Saura, J. R. (2021). Using Data Sciences in Digital Marketing: Framework, methods, and performance metrics. *Journal of Innovation & Knowledge*, 6(2), 92–102. <https://doi.org/10.1016/j.jik.2020.08.001>
2. Salminen, J., Marttila, R., Corporan, J., Salenius, T., & Jansen, B. J. (2019). Using Machine Learning to Predict Ranking of Webpages in the Gift Industry: Factors for Search-Engine Optimization. *Proceedings of the 9th International Conference on Information Systems and Technologies (ICIST 2019)*, 1–8. <https://doi.org/10.1145/3361570.3361578>
3. Matošević, G., Dobša, J., & Mladenčić, D. (2021). Using Machine Learning for Web Page Classification in Search Engine Optimization. *Future Internet*, 13(1), 9. <https://doi.org/10.3390/fi13010009>
4. Roumeliotis, K. I., & Tselikas, N. D. (2023). A Machine Learning Python-Based Search Engine Optimization Audit Software. *Informatics*, 10(3), 68. <https://doi.org/10.3390/informatics10030068>
5. Macumber, R., & Eswar Pagadala, V. D. (2024). Google, SEO and helpful content: How artificial intelligence can be helpful for e-commerce websites. *Journal of Digital & Social Media Marketing*, 12(3), 206–226. <https://doi.org/10.69554/RJUW9313>
6. Saeed, Z., Aslam, F., Ghafoor, A., Umair, M., & Razzak, I. (2024). Exploring the impact of SEO-based ranking factors for voice queries through machine learning. *Artificial Intelligence Review*, 57, 144. <https://doi.org/10.1007/s10462-024-10780-9>
7. Alotaibi, S., & Alotaibi, B. (2025). A review of click-through rate prediction using deep learning. *Electronics*, 14(18), 3734. <https://doi.org/10.3390/electronics14183734>
8. Santiago, M., Febiansyah, H., & Dinarwati, D. (2024). Integrating Machine Learning with Web Intelligence for Predictive Search and Recommendations. *International Transactions on Artificial Intelligence*, 3(1), 44–53. <https://doi.org/10.33050/italic.v3i1.654>
9. M. S., Vinutha, & Padma, M. C. (2023). Insights into Search Engine Optimization using Natural Language Processing and Machine Learning. *International Journal of Advanced Computer Science and Applications*, 14(2). Retrieved from: https://thesai.org/Downloads/Volume14No2/Paper_11-Insights_into_Search_Engine_Optimization.pdf
10. “Feature Selection using Machine Learning Techniques Based on a Case Study” (2020). *ACM Proceedings*. <https://doi.org/10.1145/3432291.3432308>
11. Aman, S. S. (2024). Search engine Performance optimization: methods and techniques. *PMC Open Access*. Retrieved from: <https://pmc.ncbi.nlm.nih.gov/articles/PMC11157186/>

Дата першого надходження статті до видання: 21.11.2025

Дата прийняття статті до друку після рецензування: 15.12.2025

Дата публікації (оприлюднення) статті 27.01.2026