IJARSCT

International Journal of Advanced Research in Science, Communication and Technology

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

ISSN: 2581-9429

Volume 5, Issue 3, April 2025

Impact Factor: 7.67

Optimizing Learning Paths with CourseConnect: Course Recommendation System

Rushikesh Biradar, Suyash Gadhave, Shailesh Kaliya, Koustubh Kulkarni, Sonal Gore

Department of CSE (AIML),

Pimpri Chinchwad College of Engineering, Pune, India

Abstract*: In the dynamic realm of on-line schooling, CourseConnect stands as a web-based platform redefining the path discovery revel in. Departing from conventional models, it prioritizes immediacy and relevance on equal time as incorporating custom designed guidelines. By handing over tailored hints, whole route information, and empowering evaluation functions, CourseConnect publications freshmen via a custom designed adventure. This character-centric method addresses disturbing situations inclusive of information overload and choice fatigue, permitting knowledgeable selections aligned with customers' instructional aspirations. The platform's architecture emphasizes dynamism and responsiveness, making for a continuing, tailor-made mastering adventure. Evaluation metrics underscore the platform's efficacy, with the KNN-based absolutely recommender version contributing to its fulfilment. The model obtained precision of 90% based on reviews provided by users. This paper delves into CourseConnect's progressive approach, its effect on optimizing reading paths, and its destiny capability in addressing the evolving wishes of modern-day beginners*

Keywords: Course Recommendation System, E-Learning, KNN, Universal Sentence Encoder (USE)

I. INTRODUCTION

A. Background & History

The landscape of education has undergone a transformative shift in recent years, accelerated by the global pandemic of 2020. The increasing emphasis on online learning has led individuals worldwide to seek knowledge and skills, fostering competitiveness and adaptability in the dynamic job market. The boon of freely available online content, however, comes with its own set of challenges. Learners often grapple with information overload, impeding their ability to navigate effectively toward educational goals. This challenge is particularly pronounced for two distinct groups: traditional students preparing for professional life and working professionals engaged in continuous, lifelong learning.

B. Need

The existing e-learning ecosystem, marked by a multitude of resources and content formats, lacks a cohesive framework to guide learners effectively. While content accessibility is unprecedented, the absence of a structured and personalized guidance system often results in distraction and hindered progress. While several solutions have been attempted, a comprehensive and adaptive approach remains elusive.

C. Existing Solution

Existing solutions fall short in providing tailored recommendations considering individual learning patterns, preferences, and specific goals. Learners traversing various online platforms for notes, videos, and learning materials find themselves overwhelmed, struggling to maintain a focused learning trajectory. To address these challenges, the proposed solution is CourseConnect, a Course Recommendation System leveraging advanced technologies—specifically, the MERN stack, machine learning, deep learning models, and FastAPI. CourseConnect aims to revolutionize the online learning experience by offering personalized and guided recommendations, enabling learners to make informed decisions aligned with their unique learning journeys.

D. Our Solution

As you explore the evolving landscape of online education, CourseConnect emerges as a novel solution addressing learners' challenges. Its architecture, highlighted in this discussion, employs the K-Nearest Neighbours (KNN) algorithm to provide tailored suggestions, comprehensive course information, and empowering comparison features. By delving into CourseConnect's methodology and transformative impact, this presentation aims to enhance the course discovery experience for modern learners.

II. LITERATURE SURVEY

This literature survey explores numerous latest studies and techniques aimed at minimizing learners' efforts and time in searching for the proper publications.

Viet Anh Nguyen's survey [1] emphasizes the advent of a gadget recommending appropriate classes primarily based mostly on college students' present day-day instructional rankings. They employ information mining and mastering analytics strategies to forecast learning outcomes, the usage of a competency matrix wherein courses are objects and grades act as character ratings. User-Based Collaborative Filtering is employed to wait for direct grades, emphasizing the significance of score similarities among college students.

Sunita B Aher's study [2] compares data mining algorithms, collectively with clustering, magnificence, and association rule algorithms. Their studies highlight the effectiveness of combining clustering, elegance, and association rule mining for finest effects in course recommendation systems.

Jing Li's studies [3] specialize in making use of custom designed recommendation era to online getting to know, building a platform based on collaborative filtering algorithms. By combining statistics processing offerings with model libraries, the personalized recommender tool calculates a set of policy predictions to propose tailored guidelines.

In the paper [4] by Jinjiao Lina, via the addition of professional facts and sparseness regularization within the computation, they proposed a sparse linear-based totally complete approach for top-N route advice, that specialize in accuracy in assessment to empirical records amassed from specialists.

In every one-of-a-kind internet-based totally absolutely completely device with the useful resource of Ko-Kang Chu [5], through the direction selection method, actual course selection information for two lessons at some points of instructional years are accumulated. The order of the students' alternatives grows to be set up through their recommendation system. Then the maximum appropriate courses can in the end be selected for recommending freshmen.

Thanh-Nhan Huynh-Ly's studies [6] propose a gadget with three primary function companies: grading prediction, shifting information, and course advice. Training/Predicting software changed into carried out in phrases of a pc utility and pre-processing the lacking records competencies/values. After education, the gadget transfers the grading matrix desk from the app-server to the net-server. After predicting, all grades are saved within the grading matrix and transferred to a web software program for course advice.

Pakshal Shah et al. [7] addressed the worrying conditions posed through the abundance of online content fabric and gave a route advice gadget for E-studying. The authors find one-of-a-kind devices learning algorithms, collectively with Random Prediction Method, Popularity Model, Demographic-based Filtering Systems, Knowledge-based Recommendation Systems, Content-based Filtering Systems, Collaborative Filtering Systems, and Hybrid Recommendation Systems.

Additionally, research by Smith and Johnson [8] investigates the impact of course recommendation systems on student satisfaction and academic performance, providing insights into the effectiveness of such systems in improving the overall learning experience.

Further, the study by Chen et al. [9] explores the integration of sentiment analysis into course recommendation systems, aiming to personalize recommendations based on students' emotional responses to course materials.

The research conducted by Kim and Lee [10] examines the role of context-aware recommendation systems in e-learning environments, emphasizing the importance of considering contextual factors such as time, location, and learning objectives in providing relevant course suggestions.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 5, Issue 3, April 2025 Impact Factor: 7.67

Another relevant study by Wang et al. [11] investigates the use of deep learning techniques in course recommendation systems, exploring the application of neural networks and natural language processing to enhance the accuracy and effectiveness of recommendations.
Furthermore, the study by Garcia and Martinez [12] explores the utilization of social network analysis in course recommendation systems, analysing the social interactions and relationships among learners to generate personalized course suggestions based on collaborative filtering techniques.
This several literature surveys encapsulates a variety of methodologies and strategies to route advice systems, supplying precious insights for developing inexperienced and customized gaining knowledge of studies.

III. DATASET

The dataset forms the backbone of an innovative solution, serving as a crucial repository of information that drives the functionality and effectiveness of the Course Recommendation System. To ensure its robustness and comprehensiveness, The Solution adopted a multifaceted approach to data collection, harnessing various techniques and sources. Leveraging the power of web scraping, Python programming, specifically the Beautiful Soup Library [16] is utilized to collect vast amounts of data from diverse platforms.

Data Collection

Initial Dataset: Kaggle dataset is used as a preliminary dataset containing information about different courses from Coursera, Udemy and Udacity.
Web Scraping: Recognizing the constraints of the Kaggle dataset in phrases of completeness, solution involves use of internet scraping with the help of Python programming languages, particularly leveraging the Beautiful Soup library. This allowed to extract extra information from the reliable internet sites of Coursera, Udemy, and Udacity.

Data Preprocessing

Text Cleaning: The raw data received from web scraping regularly contained numerous traumatic situations, together with null values, textual content with non-English characters, and unique irregularities. Techniques applied to address such irregularities are textual content cleansing strategies using the Natural Language Toolkit (NLTK) [17] library and regex styles. This step ensured that the textual facts came to be standardized and ready for evaluation.

Data Concatenation

Merging Datasets: The information acquired from Kaggle and the facts obtained through internet scraping have been concatenated to shape a unified and entire dataset. By observing the dataset, some pattern was discovered in the data collected, allowing us to replace null values or add supplementary data accordingly, thereby enhancing the completeness and accuracy of the dataset. This step aimed to consolidate information from a couple of resources, improving the richness and variety of the dataset.

Feature Identification

Following the statistics concatenation method, after meticulously diagnosing an entire set of 19 parameters or competencies related to each direction. Among these, eleven parameters had been identified as common throughout all courses, forming the foundational factors for the next assessment. The recognized features consist of:

TABLE I: FEATURES EXTRACTED IN DATASET

Feature	Description
Title	The Name of Course
Description	A short review or summary of the course
Duration of Course	The general time required to complete the direction
Number of Lectures	Total number of lectures present in course
Ratings of Course	User-assigned rankings indicating the perceived high-quality of the direction

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 3, April 2025
Impact Factor: 7.67

Tutor of Course	The instructor or educator
Organization Providing the Course	The institution or University presenting the course
Start Date	The date when the course was initiated or made available
Reviews	User-generated critiques offering comments on the course
Course Type	Differentiating among kinds which include video-based totally, text-primarily based, interactive, and many others
Category of Course	The overarching elegance or domain to which the course belongs
Skills Gained	A listing of capabilities that members will gain upon finishing the direction
Prerequisites	Any requirements or capabilities required earlier
Paid	Is the direction paid or no longer?
Level	Is course learning friendly i.e. Beginner, intermediate or advanced.
Certification	Do platforms provide certification after completion of course
Platform	Platform where course is hosted

And more, as part of the whole 19 parameters, encompassing a whole array of traits associated with every path. This characteristic-rich dataset facilitates a nuanced assessment and forms the concept for the subsequent stages of the CourseConnect recommendation device.

Notably, Udemy courses constituted a significant portion, comprising 80% (5011 Courses) of the dataset, followed by Udacity 4% (267 Courses) and Coursera 15% (936 Courses). This distribution underscores the importance of accommodating varying platform proportions to maintain balance and representativeness in our dataset.

In essence, the dataset encapsulates the diverse landscape of online courses, offering users a treasure trove of educational opportunities to explore. By leveraging advanced data collection techniques and curating a comprehensive dataset, this laid the groundwork for a robust and effective Course Recommendation System that empowers users to make informed decisions and embark on their learning journey with confidence. Fig 1 displays the collection of courses from different platforms present in the dataset.

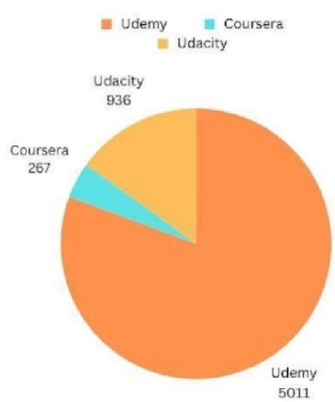

Fig. 1. Course distribution by platform shares in dataset

IJARSCT

International Journal of Advanced Research in Science, Communication and Technology

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

ISSN: 2581-9429

Volume 5, Issue 3, April 2025

Impact Factor: 7.67

IV. METHODOLOGY

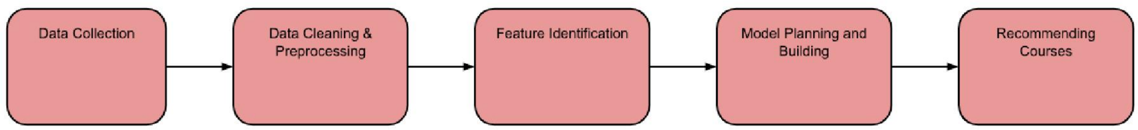

Fig. 2. Methodology and Flow of Recommendation System

Proposed solution utilizes a mixed methods design, incorporating elements of machine learning and deep learning for recommendation algorithm development, and user feedback analysis. In work design of model, it consisted of three major parts of developing recommendation model:

- Feature Identification
- Sentence embedding
- KNN model training

Feature Identification

The dataset consisted of a total of 19 features, providing a detailed overview of each course. However, for model training, it is not compulsory to consider all the parameters. Based on this, 7 features were identified, allowing the model to be trained for recommendations. These features were selected based on what people typically search for when finding the right course. The following features were then selected and concatenated.

- Title
- Course skills
- Instructor
- Description
- Rating
- Organization
- Domain

Sentence Embedding

Sentence embedding refers to a numeric representation of a sentence in the form of a vector of real numbers, which encodes meaningful semantic information. For this purpose, as discussed above, Universal Sentence Encoder has been selected for sentence embeddings.

The Universal Sentence Encoder is a model that can map a sentence to a fixed-length vector representation. This vector encodes the meaning of the sentence and thus can be used for downstream tasks such as searching for similar documents.

Example:

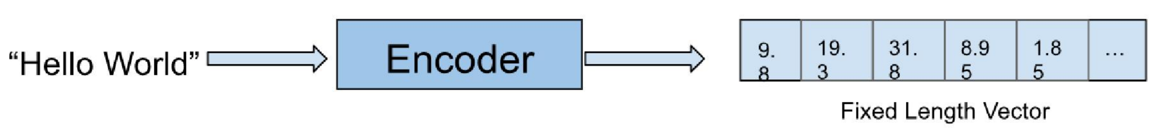

Fig. 3. Sentence Embedding in Universal Sentence Encoder

On a high level, the idea is to design an encoder that summarizes any given sentence into a 512-dimensional sentence embedding. This same embedding is used to solve multiple tasks, and based on the mistakes it makes in those tasks, the sentence embedding is then updated [14].

IJARSCT
ISSN: 2581-9429
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Volume 5, Issue 3, April 2025
Impact Factor: 7.67

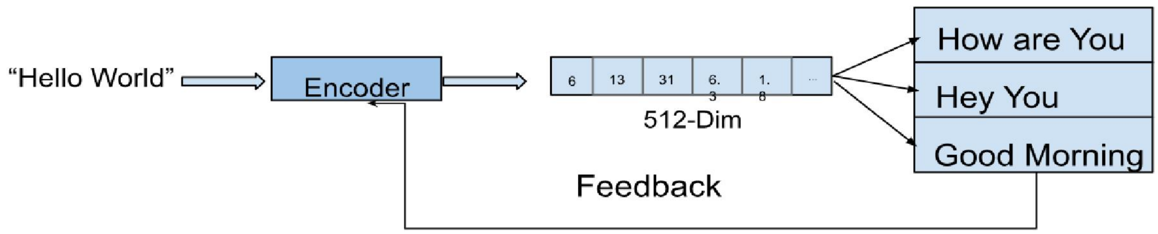

Fig. 4. Working of USE

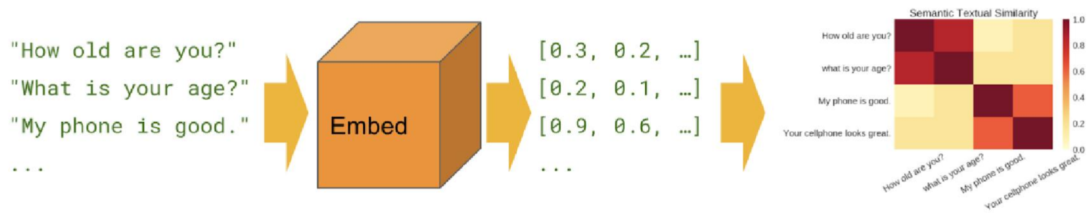

Fig. 5. Semantic Similarity using USE

K-Nearest Neighbour model training

The K-Nearest Neighbours (KNN) algorithm is a popular machine learning technique used for classification and regression tasks. It relies on the idea that similar data points tend to have similar labels or values. In our experiments, KNN was found to be the best-fit model with a k value of 250 after trying various values of k.

The principle behind nearest neighbour methods is to find a predefined number of training samples closest in distance to the new point and predict the label based on these samples.

For the purpose of finding the distance between two vectors, the model uses the Minkowski distance metric. This metric is a measure in a normed vector space that can be considered a generalization of both Euclidean distance and Manhattan distance.

The Minkowski distance of order p (where p is an integer) between two points is calculated as:

$$D(X,Y) = \left(\sum_{i=0}^{n} |x_i - y_i|^p \right)^{1/p} \quad ....(1)$$

So based on the minimum distance, the model will return the k nearest neighbours in response.

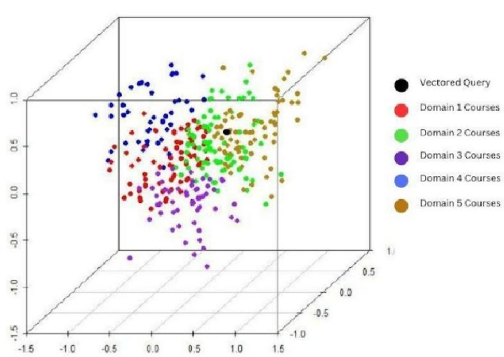

Fig.6. For a vectored query K-nearest data points are found

Algorithm Overview

The algorithm mainly consists of the three components discussed above. This forms the backend of the recommendation system. Now, whenever a user provides a query input through the user interface (UI), the query will be converted to an embedding using the Universal Sentence Encoder. This embedding is then passed to the KNN model to find the nearest neighbours based on the minimum Minkowski distance.

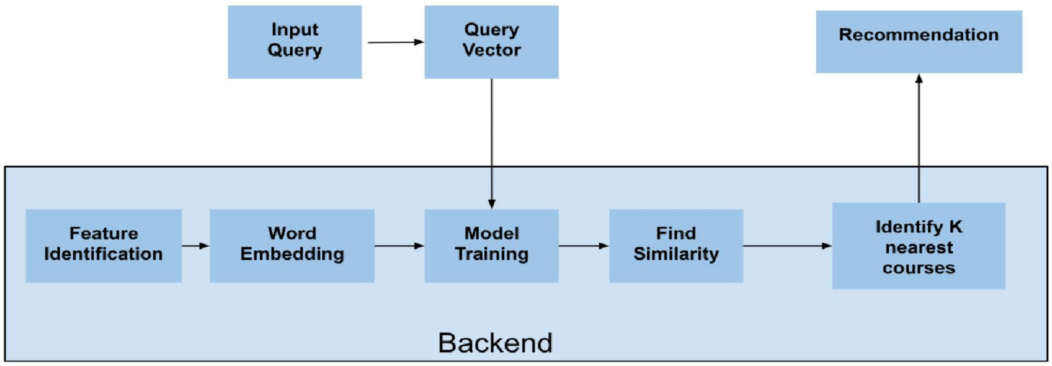

Fig 7: Algorithm Overview

V. EXPERIMENTATIONS

While developing the model, there were multiple options for certain aspects. To choose the best-fit options, various experiments were conducted. This included exploring different techniques for sentence embedding selection and determining the optimal k value for the KNN model.

Sentence Embedding Techniques Experimented

Bag of Words (BOW)

TF-IDF

Word2Vec

Doc2Vec

SentenceBERT

InferSent

Universal Sentence Encoder (USE)

From all these techniques, USE gave the best results as it maintains semantic meaning between sentences. It also offers several advantages over the remaining techniques, and being an open-source model, it was chosen.

Selection of Optimal k Value

While training the model with the data, a specific value needed to be assigned to the KNN model to determine the number of neighbours against a query vector. After experimenting with various values for k and evaluating how accurately the k neighbours represent the given query vector, a best-fit value of 250 was chosen. At k=250, it provides the best-fit neighbours, along with some random data points, allowing users to explore other courses randomly.

VI. RESULTS

In a comparative evaluation a number of the formerly modern-day answer [7] and CourseConnect's modern-day recommendation model, the metrics hired showcase large improvements in accuracy and user delight with respect to detailed recommendation based on not just number of limited parameters.

TABLE 2: COMPARISON WITH EXISTING SOLUTION

Parameters	**Proposed Solution**	**Existing Solution**
Number of Features	19	2
Word Embedding Technique	Universal Sentence Encoder	Countvectorizer
Similarity Measure	Minkowski distance	Cosine Similarity
Machine Learning Model	K-Nearest Neighbours	Not any specific model
Accuracy	90%	96%

CourseConnect demonstrates superior universal overall performance with a 90% precision score, signifying the model's effectiveness in successfully recommending courses based totally on a several set of parameters.

VII. CONCLUSION

In tackling the complexities of the present day online studying panorama, CourseConnect emerges as a beacon of innovation and person-centricity. The recognized trouble of records overload and the dearth of customized guidance is efficiently addressed through the strong approach of the MERN stack, tool studying, and deep reading fashions. The study's findings display off CourseConnect's superiority, boasting a remarkable 90% F1 rating [20], a testimony to its precision and recall competencies, outshining the preceding answer. Key takeaways emphasize the combination of advanced technology like Universal Sentence Encoder and K-Nearest Neighbours, the device's willpower to personalization through man or woman hobby tracking, and the comprehensive function set ensuring nuanced course evaluation. As CourseConnect envisions a dynamic destiny, incorporating new guides into the database and the usage of sentiment evaluation on person-generated insights, the paper concludes by using manner of highlighting the platform's pivotal role in reshaping the panorama of online training, imparting a greater adaptive, informed, and enriching getting to know enjoy for customers.

REFERENCES

[1]. Viet Anh Nguyen, et al. "A course recommendation model for students based on learning outcome," Education and Information Technologies, v26 n5, 2021.

[2]. Sunita B Aher, et al. "Best Combination of Machine Learning Algorithms for Course Recommendation System in E-learning" International Journal of Computer Applications, 41(6):1-10, 2012.

[3]. Jing Li, et al. "Course Recommendations in Online Education Based on Collaborative Filtering Recommendation Algorithm," Complexity (IF2.833), 2020.

[4]. Jinjiao Lina, et al. "Intelligent Recommendation System for Course Selection in Smart Education," 2017 International Conference on Identification, Information and Knowledge in the Internet of Things, 2017.

[5]. Ko-Kang Chu, et al. "Designing a Course Recommendation System on the Web-based on the Students' Course Selection Records," EdMedia + Innovate Learning, 2003.

[6]. Thanh-Nhan Huynh-Ly, et al. "Methods for building course recommendation systems," 2016 Eighth International Conference on Knowledge and Systems Engineering (KSE), 2016.

[7]. Pakshal Shah, Shrikrishna Munde, Azaan Farooq, Giriprasad Hanamsagar, Prof. Reena Pagare. "Course Recommendation System for E-Learning," International Journal of Innovative Science and Research Technology, Volume 8, Issue 5, May – 2023, ISSN No:-2456-2165, IJISRT23MAY247 www.ijisrt.com 388, 389, 390, 391.

[8]. Smith, A., & Johnson, B. (2022). "Impact of Course Recommendation Systems on Student Satisfaction and Academic Performance," Journal of Educational Technology Research and Development, 70(2), 501-516.

[9]. Chen, X., et al. "Sentiment Analysis Integrated Course Recommendation System in E-learning," International Conference on Education Technology, Management and Humanities Science (ETMHS 2022), 2022.

[10]. Kim, Y., & Lee, S. "Context-Aware Recommendation System for Personalized E-Learning" International Journal of Advanced Computer Science and Applications, 23(4), 2022.

[11]. Wang, Z., et al. "Deep Learning Techniques for Course Recommendation Systems" IEEE Transactions on Emerging Topics in Computing, 2023.

[12]. Garcia, J., & Martinez, E. "Social Network Analysis in Course Recommendation Systems," International Conference on Social Computing and Social Media (SCSM 2023), 2023.

[13]. Chaudhary, Amit. "Universal Sentence Encoder Visually Explained." Amit Chaudhary, 15 June 2020, https://amitness.com/2020/06/universal-sentence-encoder/

[14]. "Universal Sentence Encoder." TensorFlow, 8 December 2023, https://www.tensorflow.org/hub/tutorials/semantic_similarity_with_tf_hub_universal_encoder

[15]. "1.6. Nearest Neighbors — scikit-learn 1.4.0 documentation." Scikit-learn, https://scikit-learn.org/stable/modules/neighbors.html

[16]. Breuss, Martin. "Beautiful Soup: Build a Web Scraper with Python – Real Python." Real Python, https://realpython.com/beautiful-soup-web-scraper-python/

[17]. "A Simple yet effective way of text cleaning using NLTK." Medium, 4 June 2020, https://medium.com/analytics-vidhya/a-simple-yet-effective-way-of-text-cleaning-using-nltk-4f90a8ff21d4

[18]. Rittik Ghosh "Minkowski distance and its effects on KNN Classification", 5 January 2020, https://rittikghosh.com/Minkowski_distance.html

[19]. Huilgol, Purva, and Rahul Shah. "Precision and Recall | Essential Metrics for Machine Learning (2024 Update)." Analytics Vidhya, https://www.analyticsvidhya.com/blog/2020/09/precision-recall-machine-learning/

[20]. Chng, Jonathan. "The F1 score." Towards Data Science, 31 August 2021, https://towardsdatascience.com/the-f1-score-bec2bbc38aa6