Nanotechnology Perceptions
ISSN 1660-6795
www.nano-ntp.com

Harnessing Clustering Algorithm For Article Recommendation System: K-Means Clustering

Prabhawati Rajwade*[1], Pooja Patre[1], Jasmine Minj[1], Advait Khare[2]

[1]*Vishwavidyalaya Engineering College, Ambikapur, Chhattisgarh, India*
[2]*Shri Shankaracharya Institute of Professional Management and Technology, Bhilai, Chhattisgarh, India*
[*]*Address for correspondence: prabharajwade58@gmail.com*

Recommendation systems improve user experiences by providing individualized suggestions in a variety of sectors, including e-commerce, entertainment, and education. Among the strategies used, clustering algorithms have emerged as a useful tool for segmenting persons or products based on shared traits, allowing for more focused and effective suggestions. The article recommendation system is one them. In this paper, we use TFIDF, Woldcloud for better visualization of the word, and K-means clustering for the effective clustering of article recommendation system. In this research, we use dataset "This rich and rare dataset contains a real sample of 12 months' logs (Mar. 2016 - Feb. 2017) from CI&T's Internal Communication platform (DeskDrop)."The K-Means technique to generate data clusters that optimize the differences between the generated collections and the similarity of attributes within each group.In order to give researchers and practitioners a thorough grasp of the state of clustering-based recommendation systems today and to suggest future research avenues to improve their effectiveness and adaptability in changing environments, this work will synthesize findings from recent studies.

Keywords: Recommendation system, k-means, tfidf, woldcloud, clustering.

1. Introduction

Recommendation systems are now essential for delivering individualized experiences in a variety of sectors, such as e-commerce, streaming services, online education, and healthcare, in the age of digital transformation[1]. These systems assess user preferences and behaviors in order to recommend items, services, or content that are relevant to their interests. While classic systems such as collaborative filtering and content-based filtering have gained substantial success, they frequently encounter issues including data sparsity, scalability, and cold-start problems[2]. To solve these restrictions, clustering algorithms have emerged as a promising alternative, allowing the segmentation of persons or products into groups with comparable characteristics[3]. An unsupervised machine learning method called clustering groups related things to find patterns or structures in datasets[4]. Because clustering reduces the problem area and increases computational efficiency, it can be used in recommendation systems to improve suggestion accuracy and efficiency[5]. Because they can reveal hidden patterns in user-item interactions, techniques like k-means, hierarchical clustering, and density-based clustering have been extensively studied. This lays the groundwork for suggestions that are more focused and efficient[6].

Clustering algorithms of the machine learning plays important role for recommendation systems and for other unsupervised learning based applications.

The rest of the paper organized is as follows- section 2 addresses the background. Section 3 discusses the literature review of the proposed work. Section 4 addresses the proposed framework. Section 5 describes the discussions and finally, in section 6 deals the conclusion remark.

2. BACKGROUND

The clustering algorithm of the machine learning technique divides the data into different subsets or clusters as per the distance or similarities measured between subsets. For the recommendation system the clusters might be item clusters or user clusters based on system.Based on similarity measures and objective function of optimization that captures the desired clusters based on the mathematical fundamentals of the clustering technique. The mathematical background and preliminaries are as follows.

2.1 Distance Metrics and Similarity Measures

2.1.1 Euclidean Distance

The Euclidean distance between two points in Euclidean space is the length of the line segment between two points. It is considered as the traditional metric for problem. It is expressed as the equation given (1)-

$$d(s,t) = \sqrt{\sum_{i=1}^{n}(s_i - t_i)^2} \tag{1}$$

Where d is distance and s, t are data points. The Euclidean distance is used in K-Means clustering algorithms.

2.1.2 Cosine Similarity

In machine learning, cosine similarity is defined as the measure of similarity between two datasets as a vectors. It measures the angle between two datasets (vectors) in machine learning application. It is popularly used in text analysis like comparison between two documents, recommendation system and in searching queries. The cosine similarity is defined as equation (2)-

$$\mathrm{sim}(s,t) = \frac{s.t}{||s||X||t||} \tag{2}$$

Where s and t are two vectors. $s.t$ dot product of two vectors. $||s||$and $||t||$ are magnitude of s and t. $||s||X||t||$ is the product of s and t.

2.1.3 Manhattan Distance

For the computation of the absolute difference between data points of objects, a Manhattan distance is used. It is also referred to as "city block distance[7]". Let there be two data points s and t, then the Manhattan distance is defined between these data points as follows:

$$d(s,t) = \sum_{i=1}^{n} |s_i - t_i| \qquad (3)$$

The Manhattan distance is applicable whenever data set has sparse features.

2.1.4 Jaccard Similarity

The Jaccard Similarity, often referred to as the Jaccard index, is a Machine Learning (ML) statistic to compute the similarity between two data points. It is computed by dividing the intersection of two data sets (e.g., X and Y) by the union of two datasets (X, Y). The formula is given as:

$$Jsim = \frac{|X \cap Y|}{|X \cup Y|} \qquad (4)$$

2.2 Mathematical Background for Clustering Algorithm

The goal of clustering algorithms is to divide data into groups, or clusters, based on the fact that data points in one cluster are more similar to one another than to those in other clusters. Here are some mathematical expressions for different clustering algorithms.

2.2.1K-means clustering algorithm

K-means is among the most often applied clustering methods in recommendation systems. Using this centroid-based method—where K is a predefined number—data points are arranged into K number of clusters.The method iteratively changes the centroid of every cluster until the cluster membership stops varying. The K-means clustering algorithm is based on the shown Figure 1.

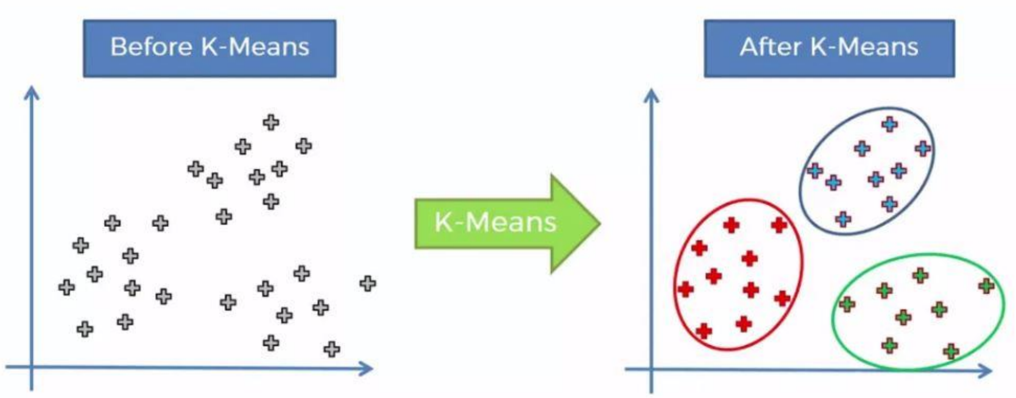

Figure 1: K-means clustering

One kind of partitional clustering algorithm is the K-means clustering algorithm. Finding the minimum squared error between each data point in the dataset and the cluster mean is the first step in partitioning provided datasets into clusters. Then, each data point is assigned to the cluster center closest to it. Mathematically, it is expressed as given a dataset $X = \{x_1, x_2, \dots\dots, x_n\}$ where x_i is in d-dimensional dataset of size n. X is clustered into different k clusters such that $S = \{s_1, s_2, \dots\dots, s_n\}$. Then

$$\text{armin}_s = \sum_i^k \sum_{x \epsilon s_i} ||x - \mu_i||^2 \qquad (5)$$ where

$\mu_i = \frac{1}{||S||_i} \sum_{x \epsilon s_i} x$ is centroid of cluster S_i.

$||x - \mu_i||^2$ is the squared Euclidian distance.

2.2.2 Silhouette Score

The silhouette score interpretation assesses the quality of k-means silhouette score by judging how effectively data points group inside their designated clusters in comparison to data points in other clusters[8]. For each data point i the silhouette score is computed as

$$s(i) = \frac{b(i) - a(i)}{\max(b(i), a(i))} (6)$$

Where $a(i)$ is mean-intra cluster distance. A data point's average distance from every other point in the same cluster. Determines how closely the cluster is clustered.

The $b(i)$ is mean nearest cluster distance. The nearest cluster is the smallest average distance between a data point and every point in any other cluster. Calculates the distance between the point and the next best cluster.

The interpretation of the equation (10) is as follows-

a. The value of s is near to 1: The point is well clustered i.e. $(b(i) > a(i)$.
b. The value of s is around 0: The point falls precisely between two clusters. That is $(b(i) = a(i)$.
c. The value of s is closer to -1: the point may belong to another cluster. That is $(b(i) < a(i))$.

2.2.3 Tfidf Vectorizer

A feature extraction method called TfidfVectorizer (Term Frequency-Inverse Document Frequency Vectorizer) is used in Natural Language Processing (NLP) to transform textual input into numerical form. It gives words weights according to how important they are in a document compared to a corpus, or group of documents. For the given word w in a document d then TF-IDF score is expressed as

$$ID - IDF(w, d) = TF(w.d) \times IDF(w) (7)$$

The equation has following terms-

$$TF(w, d) = \frac{\text{No. of times w appreared in document d}}{\text{Total words in d}} (7.1)$$

The term frequency (TF) is the measure of how frequently the word w appears in d.

The Inverse Document Frequency (IDF) that determines how unique a word is throughout all publications. It is expressed as

$\mathrm{IDF}(\mathrm{w}) = \log\left(\frac{\mathrm{N}}{1+\mathrm{DF}(\mathrm{w})}\right)$(7.2) where, N is total number of documents and $\mathrm{DF}(\mathrm{w})$ is number of documents containing w.

2.2.4 Word Cloud

Data visualization is a critical stage in data analysis. It allows us to understand the facts more readily and rapidly. Word Clouds are one of the easiest and most effective ways to visualize text data. A Word Cloud is a graphic made up of words, with the size of each word representing how frequently it appears in the dataset. They help us quickly discover the most common and significant terms in a document. In this lesson, we will learn about word clouds and how to construct them with Python.

3. Related Work

Rodrigues et al[9]proposed a system in which k-means algorithm clusters perfumes, allowing for individualized smell recommendations. Consumer feedback is then used to guide chemical design through graph neural networks (GNNs). By turning SMILES notations into molecular graphs, the GNN creates new fragrance molecules that are in line with tastes. This method helps consumers find similar perfumes while also assisting manufacturers in developing market-ready scents. The combination of clustering, feedback integration, and AI-driven molecule creation improves both recommendation accuracy and product innovation.

Pandey et al.[10]presented a study emphasizing how artificial intelligence might help us change the way we manufacture, distribute, and control food, hence producing a more safe and sustainable future for everyone.

Roumeliotis et al[11]proposed a new approach to enhance product recommender systems by combining unsupervised models—K-means clustering, content-based filtering, and hierarchical clustering—with GPT-4. GPT-4 is uniquely used for model evaluation, enhancing recommendation accuracy through its advanced language understanding. A Flask-based API enables easy implementation using CSV product data. The method empowers e-commerce platforms with sophisticated algorithms while GPT-4 refines the semantic interpretation of product features. This integration results in more personalized, effective recommendations. Experimental results demonstrate the superiority of the framework, offering businesses a scalable and efficient solution for optimizing their recommendation systems.

Yuri et al[12]Proposed an approach in authors have used K-Means technique to generates data clusters maximizing the similarity of traits inside each group and maximizing the disparities between the generated collections. The created recommendation system website contains information on the student data clustering findings from the K-Means procedure at SMAN 1 Durenan.

Zhong et al [13]Published a review paper on the role of Natural Language Processing(NLP) in the context of industry 4.0 maintenance. Authors have highlighted the use of NLP and how can NLP analyse the maintenance report for improving the machine uptime and reduce costs. Authors have addressed the challenges and issues to apply NLP in maintenance in Industry.

Diop et al [14]presents SIMREC, a system that recommends suitable similarity measure pairs for mixed data clustering using meta-learning. It analyzes dataset characteristics and algorithm performance to guide recommendations. SIMREC uses 130 similarity measure pairs, four clustering algorithms (e.g., K-Prototypes, K-Medoids), and three validity indices (Silhouette, Accuracy, ARI) to support recommendations for new and unknown datasets.

Koutsandreas et al [15]proposed a framework to enhance the machine learning in energy efficiency investments by finding project archetypes. In this framework data is used from Europe and the USA and then applied on PAM & K-means clustering to classify investment. Key factors include profitability, risk, intervention type, and building sector, aiding investment decisions.

Irina et al[16]published a review paper that provides systematic review on modern recommendation system. The review mainly focused on use of clustering techniques to enhance the performance. Key problems such data sparsity, diversity, consistency, and changing user preferences can be addressed by clusterings. For both novices and professionals, the paper summarizes current developments in recommender systems and clustering models.

Xu et al[17] proposed a framework that investigates tailored recommendation systems as a means of addressing information overload, improving user experience and business interaction. It looks at their uses in e-commerce, media, and content platforms as it contrasts conventional e-commerce classification with recommendation systems. The BERT model and closest neighbor algorithm are used in this research to present a recommendation system specifically designed for eBay.

Isinkaye et al [18] Proposed a framework that investigates Bayesian Personalized Ranking Smart Linear Model (BPRSLIM) based improvement of recommendation accuracy in collaborative filtering. While conventional BPRSLIM depends just on user-rating matrices, this work improves its performance by adding item feature information. Extensive tests on real-world datasets demonstrate notable gains in nDCG (22.1%) and precision (30.6%) over alternative top-N collaborative filtering techniques. By showing how this method improves suggestion accuracy and overall system efficacy, the study emphasizes the benefits of recreating the user-item matrix using extra data.

Florestiyanto et al [19]proposed a framework in which a systematic review of 33 studies explored methods for collecting emotional data, machine learning algorithms, adaptability, camera resolution, and ethical concerns. Emotion detection systems show promise in enhancing online learning by analyzing students' emotional responses in real time, allowing educators to adjust teaching strategies and materials. This improves student engagement and academic outcomes, though challenges and ethical considerations remain.

Algarni n et al[20]presented a comprehensive review of recommender systems using clustering algorithm of machine learning and its impact on learning. Authors have taken from

2017-2022 by selecting 35 key studies from 1,938 academic papers found via CADIMA. Through a systematic literature review (SLR), it evaluates various recommender system methodologies used for course selection, aiming to identify the most effective evidence-based approach. The research highlights trends and insights to enhance decision-making in educational settings.

4.PROPOSED METHODOLOGY

The proposed methodology of the article as shown in the Figure 2. In the proposed methodology, it has following steps- 1) Selection of dataset. The dataset is taken from https://www.kaggle.com/datasets/gspmoreira/articles-sharing-reading-from-cit-deskdrop?select=shared_articles.csv.

The dataset is CI&T created Deskdrop, an internal communications tool targeted at businesses utilizing Google G Suite. Among other things, this platform enables workers at firms to interact around pertinent content and share it with their peers. A genuine sample of 12 months' worth of logs (March 2016–February 2017) from CI&T's Internal Communication platform (DeskDrop) are included in this extensive and uncommon collection.

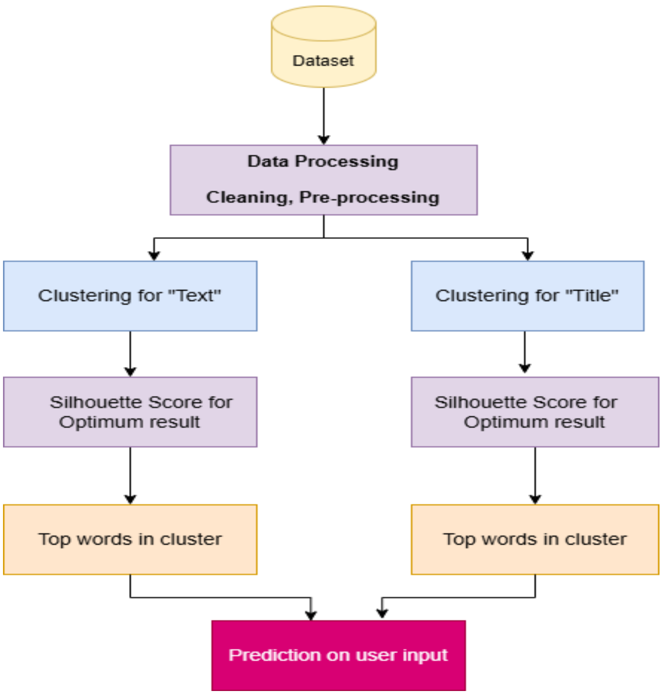

Figure 2. Proposed Methodology

2) Data pre-processing: In this step pre-processing has been done of the raw data, it involves convert some short letters into the forms that one could easily understand respectively (decontracted of the phrase), stemming, and lemmatize, pre-processing of text and pre-processing of title etc.

3) Now in step 3 we apply clustering algorithm based on "text" and "Title"

4) In step 4, apply the silhouette score to optimize the result for "text" and "title" respectively.

5) Now find the top words in clustering using WordCloud library used in Python.

6) Based on user input we predict the correct cluster. Based on the prediction 10 random articles are recommended from the whole data-frame.

5. Evaluation of the result

After the loading the dataset as in the form of .csv, a cleaning process is performed in which convert the contractions in the text string to their expanded forms using python library. Then text pre-processing has been done such as lemmatization in dataset. Next, TF-IDF vectorization applied on the text and title of the data set to transform textual input into numerical form. Now using Euclidian distance, in which Inertia determines the sum of the distances between all locations in a cluster and its centroid.The Dunn index considers not only the distance between the centroid and points, but also the distance between two clusters. Here we have applied K-mean clustering for the "text" and "title". The Plotting the curve with 'k'-value vs distance as shown in the Figure 3.

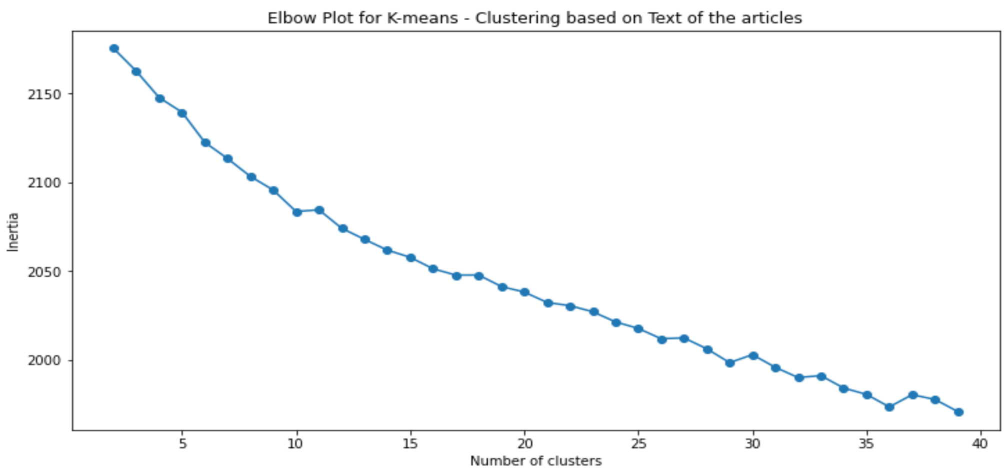

Figure 3: Elbow plot for K-means-clustering based on the text of the article

By using Silhouette scoreto find optimal number of clusters ranges from 2 to 40 e.g. 39 for the "text", as shown in the Table 1.

Table 1: Optimal numbers of clusters

Range	Silhouette score	Range	Silhouette score	Range	Silhouette score
2	0.00403545468308	15	0.01318385930648	28	0.0196695453993
3	0.00575308450570	16	0.01374618845911	29	0.0195391395800
4	0.00647692387426	17	0.01484321248321	30	0.0187919613770

5	0.00744569247380	18	0.01471237419420	31	0.0217659928848
6	0.00708910100843	19	0.01619691386765	32	0.0218376450951
7	0.00887421247487	20	0.01549200831622	33	0.0228775025347
8	0.00792634966595	21	0.01769709429815	34	0.0215368188056
9	0.01073922370625	22	0.01745218860622	35	0.0222367707199
10	0.01142035980851	23	0.01839445266089	36	0.0231290776893
11	0.00984308199886	24	0.02072673924405	37	0.0236251763896
12	0.01143847082728	25	0.01844626465645	38	0.0261591751475
13	0.01427423717894	26	0.01864063863106	39	0.0242396456665
14	0.01427423717894	27	0.02041297643204		

Previewing top 39 (k=39) words in each cluster for the “text” using wordcloud python library as shown in the Figure4.

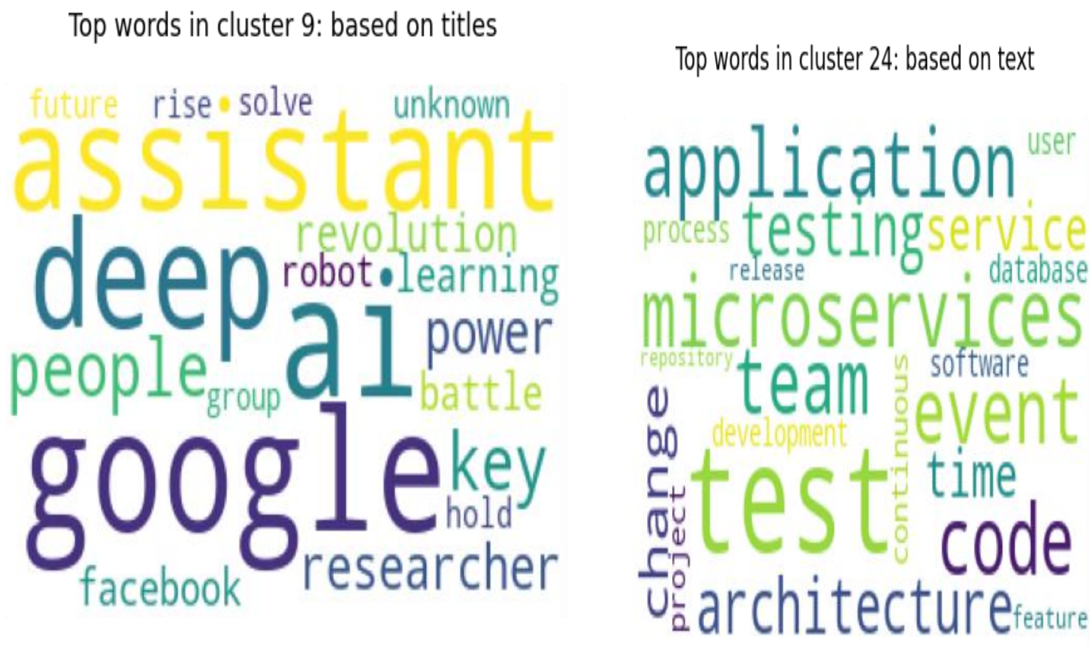

Figure 4: Two top words among the 39 top words

Now, clustering process using k-means clustering is applied on the “title” on the cleaned dataset. The Plotting the curve with 'k'-value vs distance as shown in the Figure 5.

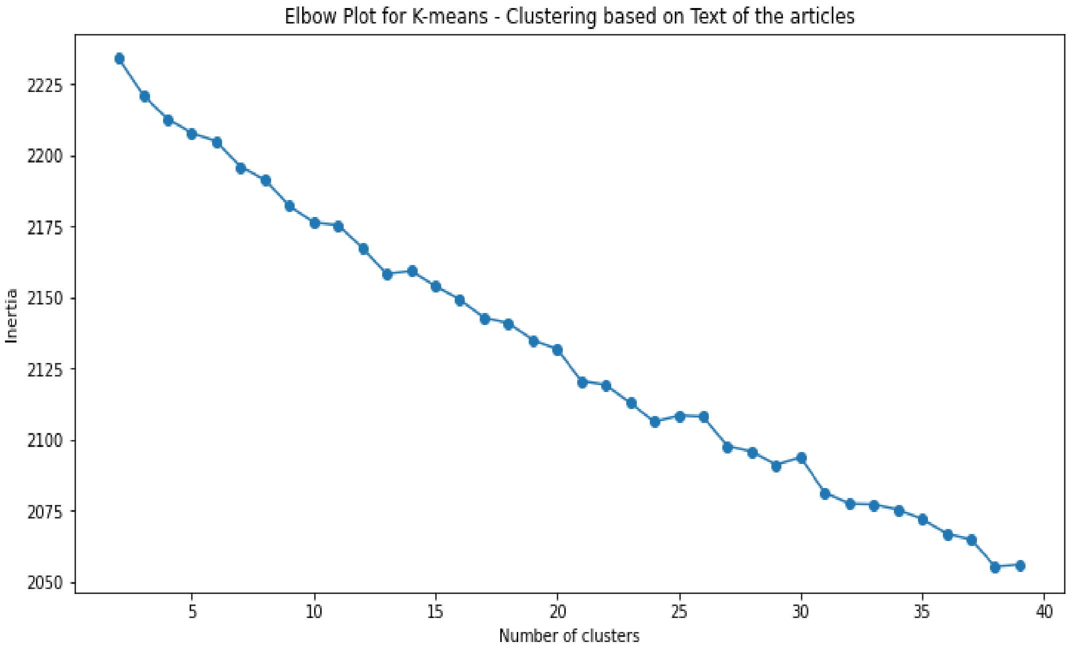

Figure 5: Elbow plot for K-means-clustering based on the text of the article

By using Silhouette scoreto find optimal number of clusters ranges from 2 to 40 e.g. 39 for the "title" , as shown in the Table 2.

Table 1: Optimal numbers of clusters for "title"

Range	Silhouette score	Range	Silhouette score	Range	Silhouette score
2	0.00403545468308	15	0.01318385930648	28	0.0196695453993
3	0.00575308450570	16	0.01374618845911	29	0.0195391395800
4	0.00647692387426	17	0.01484321248321	30	0.0187919613770
5	0.00744569247380	18	0.01471237419420	31	0.0217659928848
6	0.00708910100843	19	0.01619691386765	32	0.0218376450951
7	0.00887421247487	20	0.01549200831622	33	0.0228775025347
8	0.00792634966595	21	0.01769709429815	34	0.0215368188056
9	0.01073922370625	22	0.01745218860622	35	0.0222367707199
10	0.01142035980851	23	0.0183944526608	36	0.0231290776893
11	0.00984308199886	24	0.02072673924405	37	0.0236251763896
12	0.01143847082728	25	0.01844626465645	38	0.0261591751475
13	0.01427423717894	26	0.01864063863106	39	0.0242396456665
14	0.01441481365343	27	0.02041297643204		

Previewing top 39 (k=39) words in each cluster for the "title" using wordcloud python library as shown in the Figure 6.

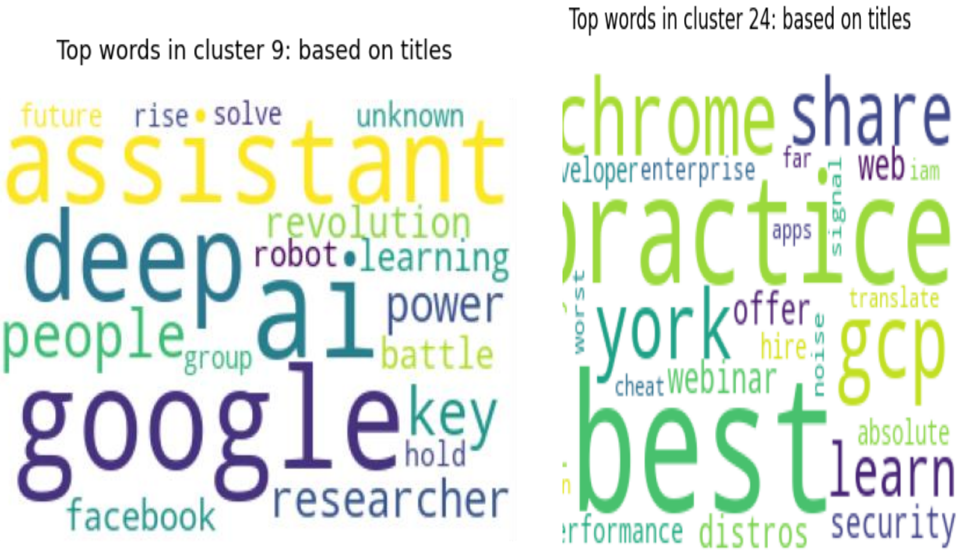

Figure 6: Two top words among the 39 top words of "title"

5.1 Recommendation based on clustering

Based on the user input, cluster prediction is performed in this section, On the basis of predicted cluster, performed the 10 random articles are recommended from the whole data-frame as shown on the below-

```
# provide user-input text based
user_input = ['The robot cannot see the human arm during the entire dressing process. In particular, it cannot always see the elt
res = recommend(user_input, True, False)

print('Recommending top 5 articles according to the user input based on text:')
res
```

```
Cluster number is:  22
Recommending top 5 articles according to the user input based on text:

{'8 Old School SEO Practices That Are No Longer Effective - Whiteboard Friday',
 'Blab shuts down, but founders promise new product on the way',
 'Google unleashes DeepMind on energy-hungry datacenter, cutting cooling bill by 40 percent',
 "Here's what Viv looks like, the next generation AI assistant built by Siri creator",
 'Virtual Weapons Are Turning Teen Gamers Into Serious Gamblers'}
```

We have provided the input as "The robot cannot see the human arm during the entire dressing process. In particular, it cannot always see the elbow or determine its precise position or bearing. That, in turn, affects the amount of force the robot has to apply to pull the article of clothing — such as a long-sleeve shirt — from the hand to the shoulder.". Then system recommends on the cluster 22. Likewise we have tested two different inputs as shown below with recommendation.

```
# provide user-input title based
user_input = ['Learning to think critically about machine learning'] #change the input
res = recommend(user_input, False, True)

print('Recommending top 5 articles according to the user input based on title:')
res
```

```
Cluster number is:  1
Recommending top 5 articles according to the user input based on title:

{'Apple Acquires Machine Learning Company Turi For $200 Million',
 "Google Calendar's newest feature uses machine learning to help you actually accomplish your goals",
 'Google saves Allo conversations, a win for machine learning but a loss for privacy - Tech2',
 'Key trends in machine learning and AI',
 'Learning at the speed of business'}
```

```
# provide user-input title based
user_input = ['Block Chain and Crypto Currency'] #change the input
res = recommend(user_input, False, True)

print('Recommending top 5 articles according to the user input based on title:')
res
```

```
Cluster number is:  5
Recommending top 5 articles according to the user input based on title:

{'Game of Loans',
 'Hotpatching a C Function on x86',
 "Lady Gaga's startup Backplane burns out and sells assets",
 'Paragraphs',
 "The Biggest Crowdfunding Project Ever Was Supposed to Create Manager-free Companies. But It's a Mess"}
```

6. Conclusion

Using K-Means clustering for article recommendation systems is an efficient and scalable way to group comparable content based on user preferences or article characteristics. By grouping articles into relevant clusters, the system may make personalised recommendations that improve user engagement and content discovery. K-Means has drawbacks, such as sensitivity to the initial number of clusters and the assumption of spherical cluster shapes, but its simplicity and speed make it an excellent starting point for recommendation jobs. Future developments could incorporate hybrid models or more sophisticated clustering algorithms to improve suggestion accuracy and respond to changing user behavior.

References

[1] A. I. Stoumpos, F. Kitsios, and M. A. Talias, “Digital Transformation in Healthcare: Technology Acceptance and Its Applications,” Int. J. Environ. Res. Public Health, vol. 20, no. 4, 2023, doi: 10.3390/ijerph20043407.

[2] S. Natarajan, S. Vairavasundaram, S. Natarajan, and A. H. Gandomi, “nnn,” Expert Syst. Appl., vol. 149, 2020, doi: 10.1016/j.eswa.2020.113248.

[3] S. Tripathi, A. Bhardwaj, and P. E, “Approaches to Clustering in Customer Segmentation,” Int. J. Eng. Technol., vol. 7, no. 3.12, p. 802, 2018, doi: 10.14419/ijet.v7i3.12.16505.

[4] N. B. A. Warif et al., “Copy-move forgery detection: Survey, challenges and future directions,” J. Netw. Comput. Appl., vol. 75, pp. 259–278, 2016, doi:

10.1016/j.jnca.2016.09.008.

[5] M. Wasid and R. Ali, “An improved recommender system based on multi-criteria clustering approach,” Procedia Comput. Sci., vol. 131, pp. 93–101, 2018, doi: 10.1016/j.procs.2018.04.190.

[6] C. X. Gao et al., “An overview of clustering methods with guidelines for application in mental health research,” Psychiatry Res., vol. 327, no. May, p. 115265, 2023, doi: 10.1016/j.psychres.2023.115265.

[7] E. M. Z. R Suwanda, Z Syahputra, “Analysis of Euclidean Distance and Manhattan Distance in the K-Means Algorithm for Variations Number of Centroid K Analysis of Euclidean Distance and Manhattan Distance in the K-Means Algorithm for Variations Number of Centroid K,” J. Phys. Conf. Ser., 2020, doi: 10.1088/1742-6596/1566/1/012058.

[8] K. R. Shahapure and C. Nicholas, “Cluster quality analysis using silhouette score,” Proc. - 2020 IEEE 7th Int. Conf. Data Sci. Adv. Anal. DSAA 2020, pp. 747–748, 2020, doi: 10.1109/DSAA49011.2020.00096.

[9] B. C. L. Rodrigues, V. V. Santana, L. P. Queiroz, C. M. Rebello, and I. B. R. Nogueira, “Harnessing graph neural networks to craft fragrances based on consumer feedback,” Comput. Chem. Eng., vol. 185, no. March, p. 108674, 2024, doi: 10.1016/j.compchemeng.2024.108674.

[10] D. K. Pandey and R. Mishra, “Towards sustainable agriculture: Harnessing AI for global food security,” Artif. Intell. Agric., vol. 12, pp. 72–84, 2024, doi: 10.1016/j.aiia.2024.04.003.

[11] K. I. Roumeliotis and N. D. Tselikas, “Precision-Driven Product Recommendation Software : Unsupervised Models , Evaluated by GPT-4 LLM for Enhanced Recommender Systems,” pp. 62–80, 2024.

[12] Y. Ariyanto, W. I. Sabilla, and Z. S. As Sidiq, “Recommendation System for Clustering to Allocate Classes for New Students Using The K-Means Method,” Compiler, vol. 13, no. 1, p. 27, 2024, doi: 10.28989/compiler.v13i1.1962.

[13] K. Zhong, T. Jackson, A. West, and G. Cosma, “Natural Language Processing Approaches in Industrial Maintenance: A Systematic Literature Review,” Procedia Comput. Sci., vol. 232, no. 2023, pp. 2082–2097, 2024, doi: 10.1016/j.procs.2024.02.029.

[14] A. Diop, N. El Malki, M. Chevalier, A. Péninou, G. R. Jimenez, and O. Teste, “Simrec : a similarity measure recommendation system for mixed data clustering algorithms,” J. Big Data, 2025, doi: 10.1186/s40537-024-01052-y.

[15] D. Koutsandreas and I. Keppo, “Harnessing machine learning algorithms to unveil energy efficiency investment archetypes,” Energy Reports, vol. 12, no. January, pp. 3180–3195, 2024, doi: 10.1016/j.egyr.2024.09.009.

[16] B. Irina and K. Mikhail, “Review of Clustering-Based Recommender Systems,” no. September, 2021, doi: 10.13140/RG.2.2.34745.90725.

[17] K. Xu, H. Zhou, H. Zheng, M. Zhu, and Q. Xin, “Intelligent Classification and Personalized Recommendation of E-commerce Products Based on Machine Learning.”

[18] F. O. Isinkaye, “Harnessing Item Features to Enhance Recommendation Quality of Collaborative Filtering,” J. Appl. Intell. Syst., vol. 8, no. 2, pp. 162–172, 2023, doi:

10.33633/jais.v8i2.7915.

[19] M. Y. Florestiyanto, H. D. Surjono, and H. Jati, Harnessing Emotions for Effective Online Learning : A Systematic Review of Emotion Recognition Techniques, no. 1. Atlantis Press International BV, 2024.

[20] S. Algarni and F. Sheldon, “Systematic Review of Recommendation Systems for Course Selection,” Mach. Learn. Knowl. Extr., vol. 5, no. 2, pp. 560–596, 2023, doi: 10.3390/make5020033.