Tadulako Science and Technology Journal
Vol. 3, No. 2, 2023, pp. 27-36
e-ISSN 2776-4893

Implementation Of The Best Training Product Sales Recommendation System Using Coseni Similarity Algorithm

Anugrah Aidin Yotolembah[a,1,*] , Hajra Rasmita Ngemba[b,2] Syaiful Hendra[a,2,*] , Muhammad Nauval Daffa Ulhaq [a,2]

[a] Informatics Engineering, Faculty of Engineering, Tadulako University
[b] Information Systems, Faculty of Engineering, Tadulako University
[1] didiyololembah19@gmail.com*;
* corresponding author

ARTICLE INFO

Keywords
Content-Based Filtering
Cosine Similarity
TFIDF-Vectorizer
Streamlight
REST API Django

ABSTRACT

Introduction: Sales of training products often encounter difficulties in providing the right recommendations according to customer needs and preferences. This research aims to develop a Content-Based Filtering-based recommendation system to address this challenge, especially for PT Menara Indonesia. **Method**: This research focuses on the Cosine Similarity algorithm, this research designs and implements the model through Django's REST API and Streamlit's Dashboard Website. The data used consists of tabular data that includes company profiles and product models. The Content-Based Filtering method is used to improve sales efficiency and support marketing strategies. **Results and Discussion**: The results show that the implementation of the Cosine Similarity algorithm provides training product recommendations that are suitable and relevant to customer needs. The suitability and relevance of the recommendation results are marked by the similarity between the two items that have been combined to produce a value of 10% to 100% which has been sorted according to the recommendation results. 10% is a recommendation result that is not very similar to the input needs and 100% is the most appropriate and accurate recommendation with the input needs. **Conclusion**: With the use of Natural Language Processing technology, this system can produce more accurate and relevant recommendations. This research is expected to help PT Menara Indonesia in increasing sales of training products and the company's marketing

1. Introduction

PT Menara Indonesia is an Indonesian company established in 2003 that focuses on consulting services in the field of training and education, through its Jakarta and Surabaya branches. PT Menara Indonesia has successfully served various companies with satisfactory quality of work, according to its expertise and consulting approach based on *Applicative, contemporary, High Impact, Fun, and Motivational*. faces challenges in increasing the effectiveness of its product sales in the midst of intense competition and complex market dynamics. In this context, the application of information technology becomes crucial, and one of the solutions is through the implementation of a recommendation system.

A recommendation system is software or a method that generates suggestions in the form of specific items that may be of interest to a particular user. In the recommendation system, there are two approaches to generating user suggestions, namely *personalized* and *non-personalized* recommendation systems, there is one main method that can be done, namely *Content-Based Filtering* [1].

The Content-Based Filtering method approach has the potential to improve the personalization of product recommendations. In this research, the *Content-Based Filtering* approach, with a focus on the *Cosine Similarity* algorithm, is proposed as a method to improve the accuracy of the suitability and relevance of training product recommendations. The *Cosine Similarity* algorithm was chosen for its ability to assess the degree of similarity between product characteristics and customer preferences. This research will make a significant contribution to increasing the efficiency of training product sales and decisions [2] The purpose of making this recommendation system is to identify the potential for implementing a recommendation system based on the *Content-Based Filtering* method using the *Cosine Similarity* algorithm in increasing the effectiveness of training product sales at PT Menara Indonesia. To achieve these objectives, this research will identify the problems faced by PT Menara Indonesia in selling training products, analyze product characteristics and consumer preferences, develop and implement a recommendation system based on *Content-Based Filtering* using the *Cosine Similarity* algorithm, and evaluate the performance of the proposed recommendation system.

Thus, this research has significant relevance in the context of increasing the effectiveness of sales of training products at PT Menara Indonesia and can contribute to the development and application of information technology-based recommendation systems in the consulting and education services industry.

2. Method

2.1. Type of Research

This research uses the type of research that is applied research. This research tends to be practical and produces a system implementation, namely the best training product sales recommendation system. The recommendation system method used is the *Content-Based Filtering* method and uses the *Cosine Similarity* algorithm, showing aspects of using information technology and artificial intelligence to provide sales recommendations. This research is also contextual and can provide solutions and direct implementation in the world of industry and business.

2.2. Research Type

This research uses the development research type. This research focuses on the development of a system, namely the best training product sales recommendation system. By using *the Content-Based Filtering* method and *Cosine Similarity* algorithm, this research tries to implement a practical solution to improve product sales efficiency. The results of the research can have a direct impact on improving the company's sales performance.

2.3. Data Analysis Method

This research uses data analysis methods using a process of collecting and processing company data that is already available and redeveloped, modeling recommendation systems using *Natural Language Pre-processing* technology and *Cosine Similarity* algorithms and *Content-Based Filtering* recommendation system methods, *and* implementing recommendation systems. In this research, the data analysis method used is quantitative data analysis using *Google Colab.*

2.4. Data Collection

In this research, the data used is tabular data originating from PT Menara Indonesia company provided by the company and developed by the author by interviewing and based on case studies at PT Menara Indonesia company.

2.5. Data Analysis

This research conducts data analysis through several stages analyzing data consists of *Data Preprocessing*, namely cleaning data from invalid values, managing and transforming data into appropriate formats such as from categorical data to numeric, data representation consists of using the *Cosine Similarity* algorithm to measure the similarity of recommended products based on the needs of company clients, modeling of the recommendation system consists of implementing a recommendation system based on the *Cosine similarity* algorithm and applying the *Content-Based Filtering* method, analysis of the results consists of analyzing the recommendation results to understand the extent to which the recommendation system can provide the best product recommendations, system implementation consists of implementing the recommendation system using *Djanggo's REST API* and *Streamlit's Dashboard Website.*

2.6. Implementation

This research in implementing the best product sales recommendation system is implemented using *Djanggo's EndPoint REST API* and *Streamlit Dashboard Website.*

2.7. Testing

Performance testing in the creation and implementation of the recommendation system conducts tests, namely:

A. *Cosine Similarity*

Testing the *cosine similarity* algorithm aims to verify the accuracy and consistency of similarity calculations on two *vectors* to perform equalization in the system.

Recommendation. The main focus of this *rarity* is to ensure that the *cosine similarity* algorithm has been implemented as specified. The expected result is an algorithm implementation that is technically compliant and reliable in providing appropriate product recommendations [3].

B. *Content-Based Filtering*

Testing the *Content-Based Filtering* method is carried out to evaluate the extent to which the recommendation system is able to provide relevant recommendations based on the characteristics of certain content or items. Using predefined inputs, this test ensures that the algorithm can effectively filter and recommend items or information that match the user's preferences or needs [4].

C. *TFIDF - Vectorizer*

TFIDF-Vectorizer testing provides word weighting for calculation in *Cosine Similarity*. evaluates the importance of words in one document compared to the entire document collection. Giving the highest value to words that frequently appear in one document but rarely appear in the whole collection emphasizes sentences that have similarities in one document. After the text is converted into *a TFIDF vector* or *numeric* data type, *Cosine Similarity* measures the extent to which the two vectors are similar, giving a score between -1 (opposite), 0 (no similarity), and 1 (exact match). The combination of *TFIDF - Vectorizer* and *Cosine Similarity* algorithm gives a good representation [5].

D. *BlackBox*

BlackBox testing is the process of evaluating software or systems without paying attention to or having in-depth knowledge of the internal structure or implementation of the code. In this test, the main focus is on the input and output generated by the system, thus testing the functionality and response capabilities of the system to various scenarios. *BlackBox* testing allows the identification of mismatches between specifications and implementation without requiring a detailed understanding of the internal logic of the system [6].

3. Results and Discussion

3.1. Results

This research implements the best product sales recommendation system using the *Cosine Similarity* algorithm at the PT Menara Indonesia company with the recommendation system method, namely the *Content-Based Filtering* method. This system processes training product data at PT Menara Indonesia, analyzes the cosine similarity between products, and uses calculations in the Cosine *Similarity* algorithm, namely (0, 1, -1) *Cosine similarity* value 1 indicates perfect similarity between two vectors while value 0 indicates total discrepancy. A value of -1 may indicate extreme dissimilarity between two vectors. *Cosine Similarity* value to provide recommendations for training products according to client needs. [7]. The results of this study show the success of the system in recommending the best sales product names that match client needs and are almost accurate. It is expected to improve sales efficiency and personalized shopping experience.

3.2. Discussion

A. *Flowchart* or flow of making a recommendation system

At this stage is an explanation of the implementation of the flow for making a recommendation system. The following is an explanation of the *Flowchart for* making a recommendation system.

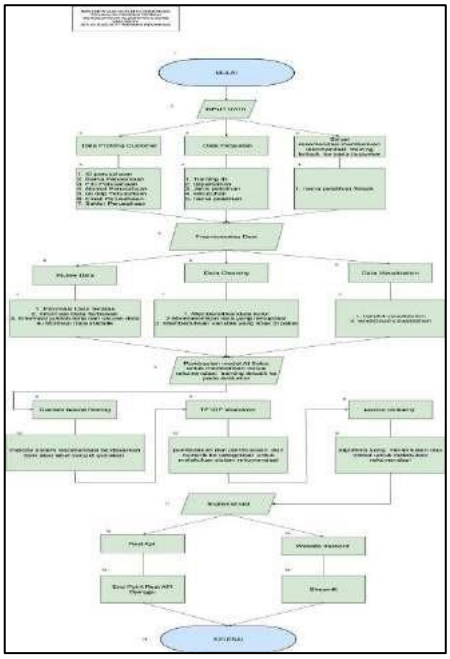


Fig. 1. *Flowchart of* recommendation system creation flow

The *Flowchart* above contains the flow of making a recommendation system that starts from inputting data through the d a t a review stage, cleaning data from duplicated data, and cleaning data or variables that are not needed, the data visualization stage. After the data input stage, then the data modeling stage which consists of the recommendation system method stage to combine the two items in order to get recommendations, namely the *Content-Based Filtering* method, *the TF-IDF Vectorizer stage* which is the stage of weighting words from text to numeric, then to the stage of equating two vectors in the calculation of the *Cosine Similarity* algorithm in order to create appropriate and relevant recommendation results. Next to the implementation stage, namely implementation into the *Djanggo EndPoint REST API* and *Streamlit Dashboard Website.*

B. Recommendation Results

This stage is a stage to display 10 recommendation results that have been carried out Processing

and data modeling which consists of calculating TF-IDF Vectorizer, calculating the Cosine Similarity algorithm, to equalize two vectors, sorting recommendations based on the calculation of Score Cosine Similarity, and adding Compatibility for ranking.

Table. 1. Recommendation Results

Needs	Product Name		Compatibility y
Innovation Center Product Development	1	Skills Development Product or Innovation	100%
	2	Creativity in campaign development marketing	90%
	3	Skills analysis that powerful	80%
	4	Skills legal analysis	70%
	5	Skills management project	60%
	6	Communication skills that both	50%
	7	Skills management IT projects	40%
	8	Skills interpersonal and communication both	30%
	9	Skills Problem-solving production	20%
	10	Skills enlargement and promotion	10%

C. Implementation

This stage is the stage to display the final results of the *data processing, data modeling,* and testing stages. This implementation stage is also the stage of displaying the output of the *project* that has been made by the author. This implementation stage consists of implementation into the *Djanggo REST API Endpoint,* and into the *Streamlit Dashboard Website.*

1. Implementation of j into *EndPoint point REST API Django*
   This stage is the implementation stage of *Djanggo's EndPoint REST API. Django REST API EndPoint.* This is the stage of creating a *REST API* in the form of points and *variables* that are synchronized and needed.

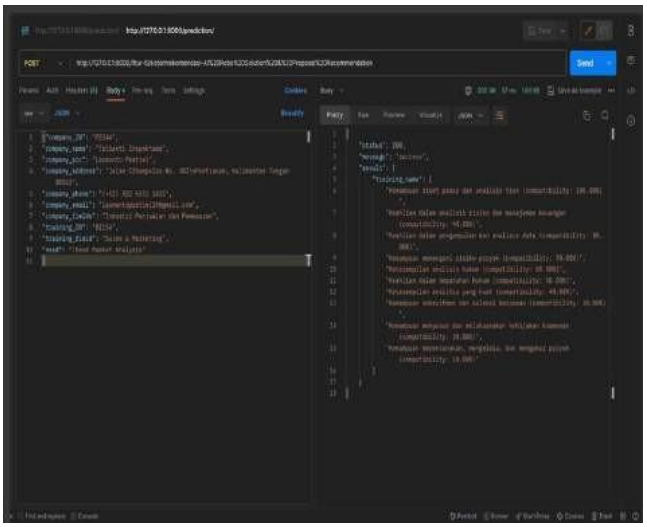


Fig. 2. *Djanggo REST API Implementation*

2. In this implementation stage, it uses the *Streamlit Dashboard Website* as a simple interface to be able to implement the recommendation system.

Fig. 3. *Djanggo REST API Implementation*

D. System testing and test results

The system test carried out is a test of the best training product recommendation system application using the *Cosine Similarity* algorithm with the recommendation method, namely *Content-Based Filtering*, system testing is carried out using the *black box* method to find out the functions contained in the system running or not. For *Cosine Similarity* algorithm testing is done by calculating in program code and manually by calculating the equation of two vectors to produce recommendations that are appropriate and similar. Testing the *Content-Based Filtering* method is done by combining two variables that will be equalized so that calculations can be made in the *TF-IDF Vectorizer* test.

1. *BlackBox* Testing

Blackbox testing is testing that is done by observing the results of execution through test data and checking the functionality of the software. The functions of each system will be tested whether successful or not. From the buttons that exist to the pages that exist whether they appear or not on the existing system. In system testing, namely, *Blackbox testing* consists of several tests, namely general functionality testing, namely testing every button or action on the website, then input testing which consists of testing inputs that are appropriate or not, then output testing, namely testing the output generated in the *streamlet dashboard website* that has been created.

General Functional Testing Results:

- *The Streamlit Dashboard website* has been deployed and can be accessed by everyone.
- All the main features of the recommendation system within the *Stemlit Dashboard Website* can be accessed and function as expected through the *Streamlit* user interface.

Input Testing Results:

- The system accepts various types of input without error
- The system's response to invalid inputs is as expected, e.g. providing informative error messages to the user

Output Testing Results:

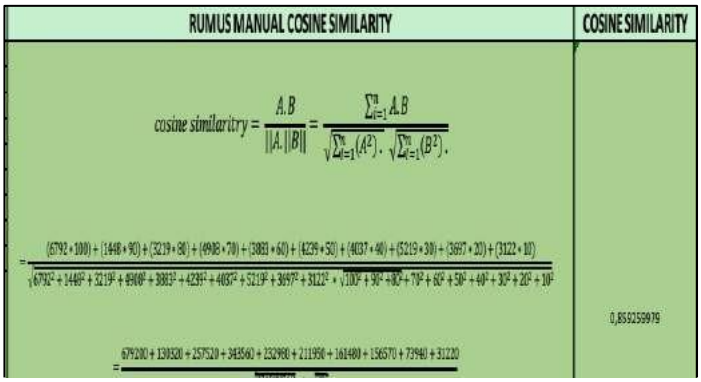
- The system provides recommendation results that are relevant and in accordance with the input entered.
- The quality of the recommendation results has been tested using the Cosine Similarity algorithm *and content-based Filtering* and is as expected.

2. *Cosine Similarity* algorithm testing

Testing the implementation of the *Cosine Similarity* algorithm aims to verify the accuracy and consistency of similarity calculations in the recommendation system. The main focus of this step is to ensure that the algorithm has been implemented in accordance with the specified technical specifications.

Table. 5. *Cosine Similarity Algorithm* Testing

NO	KEBUTUHAN	REKOMENDASI PRODUK NAME	TERM TFIDF	COMPATIBILITY
		Keterampilan komunikasi tertulis dan lisan yang baik	6792	100
		Keterampilan komunikasi yang baik	1448	90
		Keterampilan interpersonal dan komunikasi yang baik	3219	80
		Keterampilan analitis yang kuat	4908	70
		Kreativitas dalam penyusunan pesan dan kampanye komunikasi	3883	60
		Pemahaman tentang proses produksi dan optimisasi operasional	4239	50
		Pemahaman yang kuat tentang sistem dan infrastruktur TI	4037	40
		Kemampuan mengelola risiko proyek	5219	30
		Keterampilan pemasaran dan promosi	3697	20
1	susaku jelas modul komunikasi			
		Keterampilan pengawasan dan pemantauan	3122	10



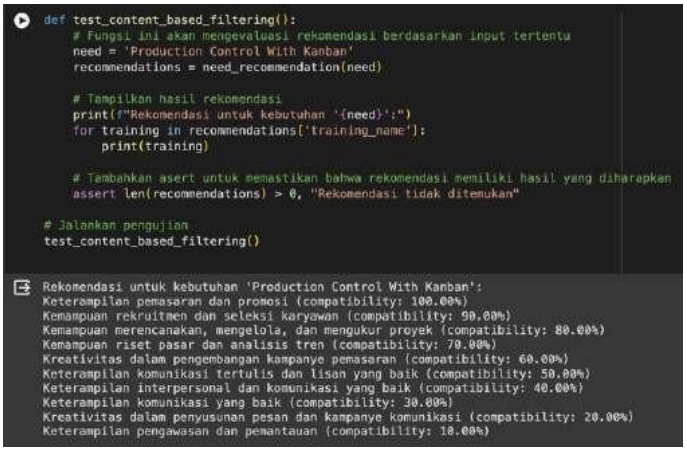
Cosine Similarity Algorithm Testing Results:

- *The Cosine Similarity* algorithm produces similarity values that are in line with expectations indicating that the implementation works well.

- *The Cosine Similarity* algorithm provides a similarity value so that it provides appropriate and relevant recommendation results.

3. *Content-Based Filtering* Testing

   Content-based filtering method testing is carried out to evaluate the extent to which the recommendation system is able to provide relevant recommendations based on certain content characteristics or attributes. By using the input that has been determined.



```
def test_content_based_filtering():
    # Fungsi ini akan mengevaluasi rekomendasi berdasarkan input tertentu
    need = 'Production Control With Kanban'
    recommendations = need_recommendation(need)

    # Tampilkan hasil rekomendasi
    print(f"Rekomendasi untuk kebutuhan '{need}':")
    for training in recommendations['training_name']:
        print(training)

    # Tambahkan assert untuk memastikan bahwa rekomendasi memiliki hasil yang diharapkan
    assert len(recommendations) > 0, "Rekomendasi tidak ditemukan"

# Jalankan pengujian
test_content_based_filtering()
```

```
Rekomendasi untuk kebutuhan 'Production Control With Kanban':
Keterampilan pemasaran dan promosi (compatibility: 100.00%)
Kemampuan rekruitmen dan seleksi karyawan (compatibility: 90.00%)
Kemampuan merencanakan, mengelola, dan mengukur proyek (compatibility: 80.00%)
Kemampuan riset pasar dan analisis tren (compatibility: 70.00%)
Kreativitas dalam pengembangan kampanye pemasaran (compatibility: 60.00%)
Keterampilan komunikasi tertulis dan lisan yang baik (compatibility: 50.00%)
Keterampilan interpersonal dan komunikasi yang baik (compatibility: 40.00%)
Keterampilan komunikasi yang baik (compatibility: 30.00%)
Kreativitas dalam penyusunan pesan dan kampanye komunikasi (compatibility: 20.00%)
Keterampilan pengawasan dan pemantauan (compatibility: 10.00%)
```

Fig. 4. *Content-Based Filtering* Testing

Content-Based Filtering Testing Results*:*

- *The Content-Based Filtering method recommendation system* provides recommendations that match the given items or client needs, showing that the implementation using the *Content-Based Filtering* method works well.

4. *TF - IDF Vectorizer* Testing

   Testing the *TF-IDF Vectorizer* on the best-selling product recommendation system using the *Cosine Similarity* algorithm ensures that the text or content is effectively represented in the vector space.

TF - IDF Vecto

```
[ ] def calculate_tfidf_matrix(data):
        tfidf_vectorizer = TfidfVectorizer(stop_words = 'english')
        tfidf_matrix = tfidf_vectorizer.fit_transform(data['combined_attributes'])

        return tfidf_matrix
```

```
[ ] print(tfidf_matrix)

  (0, 361)      0.3753852708867026
  (0, 77)       0.1448416324778541
  (0, 300)      0.32193245563222023
  (0, 59)       0.49083175534582524
  (0, 234)      0.3883524668004702
  (0, 24)       0.4239041716675823
  (0, 238)      0.40377976253154907
  (1, 222)      0.470837500415639
  (1, 158)      0.55403956199531
  (1, 13)       0.47440098696356621
  (1, 357)      0.39249720370530036
  (1, 297)      0.3037127323447119
  (2, 324)      0.34779809236918346
  (2, 453)      0.20102949788268412
  (2, 67)       0.4569818352407284
  (2, 166)      0.4569818352407284
  (2, 39)       0.38061724382623999
  (2, 158)      0.5254420483245901
  (3, 201)      0.35946812358226937
  (3, 232)      0.244055268443712533
  (3, 383)      0.30452605588098625
  (3, 76)       0.21141076402702014
  (3, 265)      0.4858708841517728
  (3, 144)      0.35822001938885363
  (3, 369)      0.4858708841517728
  :     :
  (4996, 158)   0.1655959113321646
  (4996, 13)    0.1417927766654024
  (4997, 362)   0.46574315757451973
  (4997, 468)   0.4780329678955554
  (4997, 385)   0.38258771059016483
  (4997, 426)   0.33688840641823833
  (4997, 361)   0.3954648950299283
  (4997, 77)    0.15258931403599474
  (4997, 300)   0.33915285081001767
```

Fig. 5. *TF - IDF Vectorizer* Testing

- *TF-IDF Vectorizer* gives words appropriate weighting, words that appear significantly in a single item get a higher weight than words that appear in common items.

4. Conclusion

Based on the results of research conducted by the author, the following conclusions can be drawn: Based on a case study at PT Menara Indonesia, it can be concluded that the implementation of the *Cosine Similarity* algorithm in creating a recommendation system has succeeded in providing the best training product sales recommendations based on client or customer needs. The integration of the *Content-Based Filtering* method proves its success in providing recommendations that are more personalized and in accordance with product merging items. The use of word weighting becomes numeric using the *TF- IDF Vectorizer library* and is combined with customer needs. The implementation of the recommendation system consists of two *platforms*, namely using the *Django REST API Endpoint* and the *Streamlit Dashboard Website.* Based on the results of the research conducted by the author, there are still some shortcomings, so it is still necessary for future development, namely developing and integrating the *Content-Based Filtering* method with other methods, such as *Collaborative Filtering*, to improve the quality of recommendations. This approach can provide more diverse results by utilizing the advantages of various recommendation methods. Ensure that the *Dashboard* and *Interface* views are optimized for responsiveness to enable device sharing including mobile phones and tablets. A responsive user experience can increase user engagement and satisfaction.

References

[1] M. Alkaff, H. Khatimi, and A. Eriadi, "Book Recommendation System at the Regional Library of South Kalimantan Province Using Content-Based Filtering Method," *MATRIK: Journal of Management, Informatics Engineering and Computer Engineering*,

vol. 20, no. 1, pp. 193-202, Sep. 2020, doi: 10.30812/matrik.v20i1.617.

[2] A. A. Huda, R. Fajarudin, and A. Hadinegoro, "Content-based Filtering Recommendation System Using TF-IDF Vector Similarity for News Article Recommendation," *Building of*

[3] *Informatics, Technology and Science (BITS)*, vol. 4, no. 3, Dec. 2022, doi: 10.47065/bits.v4i3.2511.

[4] Mb. Sujasman and A. Syazili, "IMPLEMENTATION OF COSINE SIMILARITY METHOD FOR PRODUCT RECOMMENDATION ON MOBILE-BASED SALES APPLICATION," *Bina Darma*
Conference on Computer Science.

[5] R. Farina, E. Sitanggang, J. W. Supratman, and K. Limun, "Implementation of Content-Based Filtering and Collaborative Filtering Method in the Tourism Recommendation System in Bali," 2023.

[6] F. D. Adhiatma and A. Qoiriah, "Application of TF-IDF Method and Deep Neural Network for Sentiment Analysis on Hotel Review Data," *Journal of Informatics and Computer Science*.

[7] Y. Dwi Wijaya and M. Wardah Astuti, "BLACKBOX TESTING OF PT INKA (PERSERO) EMPLOYEE PERFORMANCE ASSESSMENT INFORMATION SYSTEM BASED ON
EQUIVALENCE PARTITIONS," *Journal of Digital Information Technology*, vol. 4, p. 2021.

[8] S. A. Zulvian, K. Prihandani, and A. A. Ridha, "COMPARISON OF MSD AND COSINE SIMILARITY METHODS IN THE ITEM-BASED COLLABORATIVE FILTERING RECOMMENDATION SYSTEM," *Journal of Information Technology and Computer Science (INTECOMS)*, vol. 4, no. 2, p. 2021, [Online]. Available: https://nijianmo.github.io/.