Jurnal Teknik Informatika Vol. 19 No. 1, April 2026 (62-72)
ISSN: p-ISSN 1979-9160 (Print)| e-ISSN 2549-7901 (Online)
DOI: https://doi.org/10.15408/jti.v19i1.48705

JURNAL TEKNIK INFORMATIKA
Homepage : http://journal.uinjkt.ac.id/index.php/ti

Ensemble Hybrid Recommender System (CBF, CF, KNN, NBC) With Multi-View TF-IDF for Robust Preliminary Medical Diagnosis

Jamilah Karaman[1, 2*], Triyanna Widiyaningtyas[1]

[1]Department of Electrical and Informatics Engineering, Faculty of Engineering, Universitas Negeri Malang, Malang, Indonesia
[2] Department of Informatics Engineering, Faculty of Engineering, Universitas Muhammadiyah Ponorogo, Indonesia

Article:
Accepted: February 20, 2026
Revised: December 12, 2025
Issued: April 30, 2026

***Correspondence Address:**
milafergie.mf@gmail.com

ABSTRACT

Advances in health information technology require intelligent systems capable of supporting rapid and accurate diagnosis. This study proposes a Hybrid Recommender System (HRS) for preliminary medical diagnosis based on electronic medical records. The developed system combines K-Nearest Neighbor and Naïve Bayes Classifier with Multi-View TF-IDF feature representation. A total of 948 doctor-annotated medical records were used in the evaluation using a 10-Fold Cross-Validation scheme to improve the reliability of performance assessment. The results show that the hybrid model provides the best performance with an accuracy of 87.37% and an F1-score of 84.20%, consistently surpassing all comparison methods. These findings confirm that the integration of similarity-based and probabilistic learning can improve the quality of initial diagnosis recommendations in medical decision support systems. Further research will focus on expanding the dataset and clinical validation to ensure the reliability of the system in real-world practice.

Keywords : *Hybrid Recommender System; K-Nearest Neighbor (KNN); Medical Decision-Making; Multi-View TF-IDF; Naïve Bayes Classifier.*

1. INTRODUCTION

The development of information technology and artificial intelligence has had a significant impact on the healthcare sector in Indonesia. [1]. One technology that has shown rapid progress is the recommendation system, which is capable of generating clinical advice based on historical data patterns and patient characteristics. In primary health care services such as community health centers and clinics, the need for this technology is increasing due to limited medical personnel, high workloads, and dependence on subjective assessments in the diagnosis process [2]. Various studies in Indonesia have noted that the rate of misdiagnosis in primary care can reach 10-20%, particularly for diseases with similar symptoms such as acute respiratory infections, typhoid fever, and several non-communicable diseases [3]. The Indonesian Ministry of Health report also shows that delayed diagnosis plays a major role in increasing complications and referrals to hospitals, thus requiring a decision support system that is objective, efficient, and capable of providing rapid preliminary analysis [3].

The development of machine learning-based medical decision support systems has encompassed various computational approaches, ranging from deep learning, transformers, gradient boosting, to BERT-based models for clinical text processing [4]. Although these methods have been widely applied, research that specifically combines similarity-based recommendation approaches and probabilistic models into a single hybrid architecture to support initial diagnosis is still very rare. Various machine learning methods such as deep learning, transformers, gradient boosting, and BERT have been used in the development of medical decision support systems. However, research that explicitly combines similarity-based recommendation approaches and probabilistic models in a single hybrid architecture for initial diagnosis is still very limited [5]. To date, there have been no studies designing a Hybrid Recommender System with the integration of Content-Based Filtering, Collaborative Filtering, K-Nearest Neighbor, and Naïve Bayes Classifier in medical record data. This indicates an important research gap, especially in the application of multi-view representation that combines TF-IDF n-gram, cosine similarity, and probabilistic models for primary health care services [6]. These findings are consistent with the latest systematic review by Sabiri et al., which confirms that the comprehensive integration of hybrid methods in the health domain remains an uncharted area of research. [5]

This study developed a multi-view TF-IDF-based Hybrid Recommender System architecture that utilizes patient identity, symptoms, examination results, diagnoses, and medical history features to improve the accuracy of initial diagnoses. The proposed system integrates the advantages of CBF in handling new patients, the capacity of CF in mapping similarities between patients, the accuracy of KNN in calculating case proximity, and the stability of NBC inference on complex and high-dimensional clinical data. All of these components are implemented in a machine learning-based web platform that allows medical personnel to obtain real-time analysis, thereby supporting a more systematic, consistent, and evidence-based clinical process.

The contributions of this research include the utilization of structured and unstructured medical record datasets to form TF-IDF n-gram-based multi-view representations that enrich the clinical context; the design of a hybrid architecture that integrates CBF, CF, KNN, and NBC into a single recommendation framework for initial diagnosis, an approach that has not been specifically applied to medical decision support systems in Indonesia; and the presentation of a comprehensive evaluation using accuracy, precision, recall, and F1-score metrics to provide a comprehensive overview of the system's performance in a clinical environment. Thus, this study offers methodological novelty through multi-model integration, empirical novelty through evaluation results that are superior to single models, and practical novelty through system implementation on a patient data platform that is relevant to primary health facilities in Indonesia.

1.1. Recommendation Systems in Healthcare

A recommender system is a branch of artificial intelligence used to provide users with the best suggestions or choices based on data analysis and historical preferences. [7]. In the field of healthcare, recommendation systems have been widely used to assist doctors and

ISSN: p-ISSN 1979-9160 (Print)| e-ISSN 2549-7901 (Online)
DOI: https://doi.org/10.15408/jti.v19i1.48705

patients in medical decision-making, such as initial diagnosis, drug selection, and appropriate treatment. [8]. The application of recommendation systems in the medical field plays an important role in Clinical Decision Support Systems (CDSS), as it can improve diagnostic accuracy and reduce the workload of medical personnel [9]. Several previous studies have shown that machine learning and natural language processing (NLP) models are capable of identifying patterns in patient data such as symptoms, medical history, and laboratory results to generate relevant medical advice [10].

1.2. Hybrid Recommender System

A Hybrid Recommender System (HRS) is a combination of two or more recommendation methods, such as content-based filtering and collaborative filtering, which aims to improve accuracy and overcome the limitations of each method [11]. Content-based filtering works by analyzing attributes or features of items (e.g., symptoms or diagnoses) to recommend similar items [12].

Collaborative filtering uses previous user interaction data to find patterns of similarity in behavior between users or patient cases [9]. In a medical context, the combination of these two approaches has proven effective in improving the performance of early diagnosis systems. A study by Derevitskii I et al. (2022) shows that a hybrid recommendation system based on KNN and Bayesian Network can improve the accuracy of chronic disease diagnosis by 12% compared to single methods [13].

1.3. Machine Learning for Medical Diagnosis

Machine learning is the main foundation in the development of modern recommendation systems in the health sector [14]. Some commonly used algorithms include K-Nearest Neighbors (KNN), Naïve Bayes, Decision Tree, and Support Vector Machine (SVM) [15]. Research by Yogianto et al. (2023) shows that the use of KNN algorithms for classifying patient symptoms can achieve a diagnosis accuracy rate of up to 92%, especially for diseases with recurring symptom patterns [16].

Meanwhile, Naïve Bayes provides advantages in probabilistic interpretation that is suitable for diagnosis based on high-dimensional medical data [17]. The combination of these two algorithms is widely used in hybrid recommendation systems because it is able to combine a data similarity-based approach with diagnosis probability [18]. Thus, machine learning plays an important role in generating diagnostic recommendations that are not only accurate but also logically explainable by the system.

1.4. Evaluation of Medical Recommendation Systems

The medical recommendation system was evaluated to measure the reliability and effectiveness of the model in providing recommendations that are appropriate for the patient's condition. The evaluation process uses performance metrics of accuracy, precision, recall, and F1-score, which are commonly used to assess the performance of classification models [19]. The accuracy metric is used to measure the proportion of correct predictions against all test data, while precision and recall assess the accuracy and ability of the model to detect relevant classes [20].

The F1-score provides a balance between precision and recall, enabling it to represent model performance more comprehensively. Three main approaches were evaluated, namely content-based filtering, collaborative filtering, and hybrid recommender systems, to determine the method with the most optimal performance in the context of medical diagnosis recommendations [21].

2. METHODS

2.1. Data Collection

The research data was obtained from the Patient Data Information System that had been developed in the previous stage. The dataset included 1,000 anonymous patient records with the structure shown in Table 1. All patient identities were anonymized to maintain the confidentiality of personal data.

Table 1. *Description of Patient Data Attributes*

Attribute	Attribute
Patient ID	Unique patient identification number
Age	Patient's age in years
Gender	Male / Female
Symptoms	List of reported symptoms
Examination Results	Lab results or physical measurements
Medical History	Previous illness data
Doctor's Diagnosis	Manual examination diagnosis

Jurnal Teknik Informatika Vol. 19 No. 1, April 2026 (62-72)
ISSN: p-ISSN 1979-9160 (Print)| e-ISSN 2549-7901 (Online)
DOI: https://doi.org/10.15408/jti.v19i1.48705

Medical Action	Recommendations or therapy given by the doctor

2.2. Proposed Hybrid Recommender System Architecture

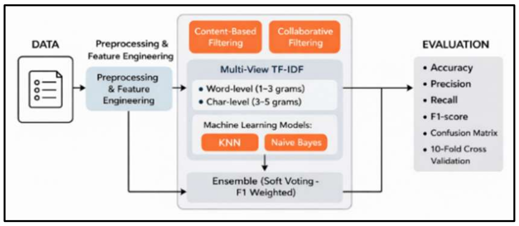


Figure 1. *Hybrid Recommender System Architecture*

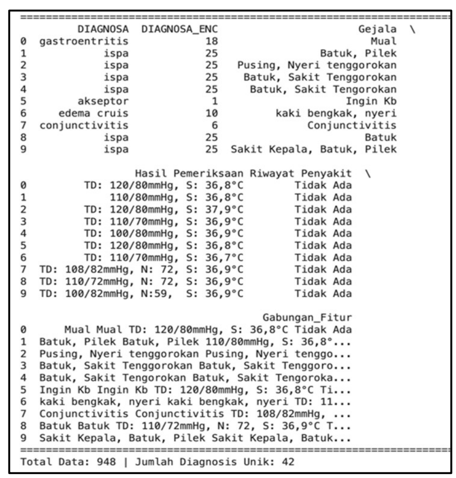
This study proposes a Hybrid Recommender System architecture for preliminary medical diagnosis that integrates multiple approaches to improve prediction accuracy and robustness. The process begins with data preprocessing and feature engineering, where patient data is cleaned, normalized, and transformed into a structured representation. The processed data is then converted into numerical features using a multi-view TF-IDF approach, combining word-level and character-level representations to capture both semantic and morphological patterns.

The system employs a hybrid modeling strategy that integrates Content-Based Filtering (CBF), Collaborative Filtering (CF), and machine learning algorithms, namely K-Nearest Neighbor (KNN) and Naïve Bayes Classifier (NBC). The outputs from these models are combined using a soft voting ensemble method with F1-score-based weighting to generate the final diagnosis prediction. Finally, the system performance is evaluated using multiple metrics, including accuracy, precision, recall, F1-score, and confusion matrix, along with 10-fold cross-validation to ensure model reliability.

2.3. Data Preprocessing and Feature Engineering

The preprocessing stage is carried out to ensure data quality and consistency before entering the modeling process. Diagnosis normalization is performed by converting all labels to lowercase, removing excess spaces, and combining different spelling variations into a single uniform category. Diagnoses with a frequency of occurrence of less than two cases are eliminated because they are considered noise and do not contribute significantly to model learning. Next, symptoms, examination results, and medical history features are combined into a single medical text representation. At this stage, symptoms are given double weight by duplicating the text before merging because, clinically, patient complaints are the most dominant indicator in determining a diagnosis.

For example, the data: Symptom "Nausea," examination results "BP 120/80 mmHg; Temperature 36.8°C," and history 'None' are engineered into "Nausea Nausea BP 120/80 mmHg Temperature 36.8°C None." This weighting reinforces the role of symptom words, which are generally short but highly informative. Statistically, this technique increases the model's sensitivity to key features, and empirically, it has been proven to increase the F1-score because the model more easily recognizes patterns directly related to diagnosis. After this process, the diagnosis label is converted to a numerical form through label encoding to facilitate the classification process. All of these steps produce a final dataset that is clean, consistent, informative, and representative, thus providing a strong foundation for building an accurate and reliable diagnosis classification model.



```
===========================================================
          DIAGNOSA  DIAGNOSA_ENC                     Gejala  \
0  gastroentritis            18                       Mual
1             ispa            25               Batuk, Pilek
2             ispa            25   Pusing, Nyeri tenggorokan
3             ispa            25    Batuk, Sakit Tenggorokan
4             ispa            25     Batuk, Sakit Tengorokan
5         akseptor             1                   Ingin Kb
6      edema cruis            10         kaki bengkak, nyeri
7  conjunctivitis             6              Conjunctivitis
8             ispa            25                      Batuk
9             ispa            25  Sakit Kepala, Batuk, Pilek

            Hasil Pemeriksaan Riwayat Penyakit  \
0        TD: 120/80mmHg, S: 36,8°C        Tidak Ada
1            110/80mmHg, S: 36,8°C        Tidak Ada
2        TD: 120/80mmHg, S: 37,9°C        Tidak Ada
3        TD: 110/70mmHg, S: 36,9°C        Tidak Ada
4        TD: 100/80mmHg, S: 36,9°C        Tidak Ada
5        TD: 120/80mmHg, S: 36,8°C        Tidak Ada
6        TD: 110/70mmHg, S: 36,7°C        Tidak Ada
7  TD: 108/82mmHg, N: 72, S: 36,9°C        Tidak Ada
8  TD: 110/72mmHg, N: 72, S: 36,9°C        Tidak Ada
9  TD: 100/82mmHg, N:59,  S: 36,9°C        Tidak Ada

                                   Gabungan_Fitur
0     Mual Mual TD: 120/80mmHg, S: 36,8°C Tidak Ada
1  Batuk, Pilek Batuk, Pilek 110/80mmHg, S: 36,8°...
2  Pusing, Nyeri tenggorokan Pusing, Nyeri tenggo...
3  Batuk, Sakit Tenggorokan Batuk, Sakit Tenggoro...
4  Batuk, Sakit Tengorokan Batuk, Sakit Tengoroka...
5  Ingin Kb Ingin Kb TD: 120/80mmHg, S: 36,8°C Ti...
6  kaki bengkak, nyeri kaki bengkak, nyeri TD: 11...
7  Conjunctivitis Conjunctivitis TD: 108/82mmHg, ...
8  Batuk Batuk TD: 110/72mmHg, N: 72, S: 36,9°C T...
9  Sakit Kepala, Batuk, Pilek Sakit Kepala, Batuk...
===========================================================
Total Data: 948 | Jumlah Diagnosis Unik: 42
```

Figure 2. *Patient Data Preprocessing Results*

The results of this entire preprocessing series produced a final dataset containing 948 samples with 42 unique features representing a variety of symptoms, physical conditions, and patient medical histories. Pattern analysis shows different clinical characteristics for each diagnosis. For example, gastroenteritis is characterized by symptoms of nausea, cough,

Jurnal Teknik Informatika Vol. 19 No. 1, April 2026 (62-72)
ISSN: p-ISSN 1979-9160 (Print)| e-ISSN 2549-7901 (Online)
DOI: https://doi.org/10.15408/jti.v19i1.48705

runny nose, dizziness, and sore throat, while conjunctivitis is dominated by symptoms of conjunctivitis. Thus, this preprocessing and feature engineering stage ensures that the dataset is in the most optimal condition—clean, consistent, informative, and representative—to serve as a strong foundation for building an accurate and reliable diagnosis classification model.

2.4. Distribution Diagnosis and Imbalance Analysis

The following is the distribution of diagnoses obtained after normalization:

```
                          Diagnosis  Jumlah  Persentase (%)
0                              ispa     248            24.8
1                         gastritis     128            12.8
2                          akseptor     125            12.5
3                       cephalalgia     113            11.3
4                           myalgia      68             6.8
..                              ...     ...             ...
88                      pedal edema       1             0.1
89            dizziness dengan fever      1             0.1
90  pharyngitis dengan dizziness          1             0.1
91                          dyspnea       1             0.1
92                            diare       1             0.1

[93 rows x 3 columns]
```

Figure 3. *Distribution of Patient Diagnoses*

The imbalance analysis shows that the dataset has an unbalanced class distribution, with several dominant diagnoses such as ISPA (24.8%), gastritis, and acceptors, while many other diagnoses only have a very small number of samples (<5 cases). This condition is a long-tail class imbalance that can reduce model performance, especially in minority classes that are prone to being overlooked in the learning process. To minimize this bias, this study applies two main strategies: first, removing diagnoses with a frequency of occurrence of less than two cases to reduce noise and prevent the formation of unrepresentative classes; second, applying a soft-voting mechanism with weighting based on the F1-score value so that minority classes still have a proportional influence in the final prediction process. This approach is expected to maintain model performance stability while improving accuracy across all classes.

2.5. Multi-View TF-IDF Vectorization and Model Performance Analysis

a. Multi-View TF-IDF Vectorization

At this stage, text representations are constructed using the Multi-View TF-IDF method to strengthen the semantic context and linguistic patterns in patient data. Two perspectives are used, namely word-level n-grams (1–3) to capture the meaning of medical phrases, and character-level n-grams (3–5) to recognize morphological patterns such as word form variations, prefixes, and suffixes. The vectorization process produces two feature matrices, each measuring (948 × 1500) for word-level and (948 × 800) for character-level. These two representations are then combined using the concatenation technique, which is the horizontal merging of vectors, to form a richer and more informative multi-view representation measuring (948 × 2300). This approach was chosen because it can retain semantic and morphological information without performing reductions that could potentially eliminate important features.

```
============================================================
TAHAP 3: MULTI-VIEW TF-IDF VECTORIZATION (VARIASI SPLIT)
============================================================

[3.1] Membuat Representasi Multi-View TF-IDF...
✓ TF-IDF Word shape: (948, 1500)
✓ TF-IDF Char shape: (948, 800)

[3.2] Evaluasi Setiap Variasi:
------------------------------------------------------------
(90:10) = 0.8316
✓ Train: 853 | Test: 95

(80:20) = 0.8158
✓ Train: 758 | Test: 190

(75:25) = 0.7975
✓ Train: 711 | Test: 237

(70:30) = 0.7930
✓ Train: 663 | Test: 285
```

Figure 4. *Results of Multi-View TF-IDF Vectorization*

The data was then stratified into training and test data to maintain the balance of diagnosis distribution. The evaluation in Figure 3 shows that Multi-View TF-IDF representation has high and stable performance across various training and test data ratios. The best accuracy was obtained at a ratio of 90:10 (0.8316), followed by 80:20 (0.8158), 75:25 (0.7975), and 70:30 (0.7930). The decrease in accuracy of 0.03–0.04 as the proportion of test data increases is a natural phenomenon, as the reduced amount of training data limits the model's ability to recognize text patterns. However, the stability of performance across all ratio variations shows that the combination of word-level and character-level successfully produces features that are robust to changes in data size. Thus, Multi-View TF-IDF is capable of providing a strong representation basis to

Jurnal Teknik Informatika Vol. 19 No. 1, April 2026 (62-72)
ISSN: p-ISSN 1979-9160 (Print)| e-ISSN 2549-7901 (Online)
DOI: https://doi.org/10.15408/jti.v19i1.48705

support the diagnosis recommendation model in the next stage.

b. Accuracy Analysis Based on Train-Test Split

The test results shown in Figure 4 aim to analyze the performance comparison of three recommendation system models, namely Collaborative Filtering (CF), Content-Based Filtering (CBF), and Hybrid Recommender, against various variations of training and testing data splits (train:test split).

This test was conducted to determine the extent to which the proportion of training data affects the accuracy of the recommendation model, as well as to determine the most optimal and stable model under various training data conditions. The data distribution proportions used in this study included four variations, namely 90:10, 80:20, 75:25, and 70:30.

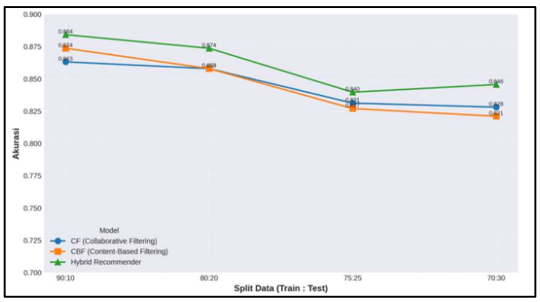


Figure 5. *Comparison of Model Accuracy Based on Data Split*

Based on the results obtained, all three models showed a downward trend in accuracy as the proportion of training data decreased. This illustrates that the smaller the amount of training data used, the more limited the model's ability to recognize patterns of relationships between users and recommended items. However, the rate of decline in accuracy varied for each model, indicating differences in generalization ability among the three.

c. Formula and Pseudocode Implementation

TF-IDF:

$$tfidf(t,d) = tf(t,d) \cdot log\left(\frac{N}{df(t)}\right) \quad (1)$$

Concatenation Multi-View:

$$X = [X_{\text{word}} \;\; || \;\; X_{\text{char}}] \in \mathbb{R}^{n\times(a+b)} \quad (2)$$

Pseudocode:

```
text = combine(gejala*2, hasil_pemeriksaan, riwayat)
tfidf_word = TFIDF(ngram=(1,3), max_features=1500).fit_transform(text)
tfidf_char = TFIDF(analyzer='char', ngram_range=(3,5), max_features=800).fit_transform(text)
X = concatenate(tfidf_word, tfidf_char)
```

2.6. Hyperparameter Model

Table 2. *Model Hyperparameter Configuration*

Model	Hyperparameter	Value
TF-IDF Word	ngram_range	(1,3)
TF-IDF Word	max_features	1500
TF-IDF Char	ngram_range	(3,5)
TF-IDF Char	analyzer	char
TF-IDF Char	max_features	800
KNN	k	5
NBC	smoothing	Laplace
CBF	similarity metric	Cosine
CF	similarity metric	Cosine

Table 2 shows the hyperparameter configuration for each model used. At the text representation stage, word-based TF-IDF uses an ngram range (1–3) with a maximum limit of 1500 features to capture a broader context of words, while character-based TF-IDF uses an ngram range (3–5), a character-based analyzer, and 800 features to recognize letter patterns and spelling variations. The K-Nearest Neighbor model uses the five closest neighbors as a classification reference, while the Naive Bayes Classifier applies Laplace smoothing to overcome the problem of zero probability. In the recommendation method, both Content-Based Filtering and Collaborative Filtering use Cosine Similarity as a measure of similarity to generate recommendations based on content similarities and user interaction patterns.

2.7. Hybrid Filtering

Hybrid Filtering is a recommendation system approach that integrates several methods to improve the accuracy and relevance of recommendation results [22]. In this study, a mixed hybrid technique was applied, which is a combination of the results of Content-Based Filtering (CBF), Collaborative Filtering (CF), K-Nearest Neighbor (KNN), and Naïve Bayes Classifier (NBC) to obtain more comprehensive and accurate diagnosis recommendations.

a. Content-Based Filtering (CBF)

CBF calculates the similarity between new patient data and previous patients based on symptom attributes, examination results, and

Jurnal Teknik Informatika Vol. 19 No. 1, April 2026 (62-72)
ISSN: p-ISSN 1979-9160 (Print)| e-ISSN 2549-7901 (Online)
DOI: https://doi.org/10.15408/jti.v19i1.48705

medical history. Similarity is calculated using Cosine Similarity, with the formula:

$$similarity(A, B) = \frac{A \cdot B}{\parallel A \parallel \times \parallel B \parallel} \quad (3)$$

where A and B are vectors representing patient features. The highest similarity value indicates the closeness of patient characteristics and forms the basis for diagnosis recommendations and appropriate medical actions [23].

b. Collaborative Filtering (CF)

CF utilizes a patient-diagnosis matrix to find patterns of similarity between patients. A user-based collaborative filtering approach is used to identify patients with similar conditions. The similarity value between patients is calculated using Cosine Similarity:

$$sim(u, v) = \frac{\sum_i \ r_{u,i} r_{v,i}}{\sqrt{\sum_i \ r_{u,i}^2} \times \sqrt{\sum_i \ r_{v,i}^2}} \quad (4)$$

where $r_{u,i}$ and $r_{v,i}$ are the diagnostic values of patients u and v for item (disease) i. These similarity values form the basis for providing new diagnostic recommendations based on similar patient patterns [24].

c. K-Nearest Neighbor (KNN)

The KNN algorithm classifies new patients based on the k nearesst neighbors with the smallest distance. The Euclidean distance formula is used to calculate the proximity between patients:

$$d(x, y) = \sqrt{\sum_{i=1}^{n} \ (x_i - y_i)^2} \quad (5)$$

New patients are classified based on the majority diagnosis of the k nearest patients, with optimal performance obtained at k = 5 [25].

d. Naïve Bayes Classifier (NBC)

NBC is used to calculate the probability of diagnosis based on observed symptoms by applying Bayes' theorem:

- P $(C|x)$ = probability of diagnosis C for symptom x
- P $(x|C)$ = probability of symptom x occurring in diagnosis C
- P (C) = overall probability of diagnosis C
- P (x) = probability of symptom x

The NBC model generates probability-based diagnostic predictions which are then combined (ensemble averaging) with KNN prediction results to improve the stability and accuracy of the medical recommendation system [26].

e. Soft Voting

Soft voting combines the probabilistic predictions of each base model (CBF, CF, KNN, NBC). For each diagnosis Cj, the final score is calculated as:

$$Score(C_j) = \sum_{m=1}^{M} \quad w_m \cdot P_m(C_j) \quad (6)$$

where:

- Pm (Cj) = diagnosis probability of the mth model
- wm = weight of the mth model
- M = number of models (4 models in this study)

The final diagnosis is selected based on the highest score.

The weight of each model is not given statically, but is calculated based on its actual performance in the training evaluation. The weight is determined proportionally to the F1-score of each model:

$$w_m = \frac{F1_m}{\sum_{k=1}^{M} \quad F1_k} \quad (7)$$

With this mechanism:

- Higher-performing models contribute more to the final prediction.
- Weaker models are not removed, but their contribution is reduced.
- This approach improves prediction stability on clinical datasets with high symptom variation.

2.8. Ethics & Consent for Data Use

The research follows the following medical ethics principles:

a. Informed Consent
All data is obtained from patients with consent for use in the development of the system.

Jurnal Teknik Informatika Vol. 19 No. 1, April 2026 (62-72)
ISSN: p-ISSN 1979-9160 (Print)| e-ISSN 2549-7901 (Online)
DOI: https://doi.org/10.15408/jti.v19i1.48705

b. Data Anonymization
All identities (names, addresses, contacts) are deleted/changed so that they cannot be traced back to the patient.
c. Data Security
Data is stored on an internal server protected by authentication.
d. Non-Maleficence and Beneficence
The system is designed to support medical decisions, not to replace doctors.
e. Compliance
All procedures follow ethical standards and relevant healthcare facility policies. [27].

3. RESULTS AND DISCUSSION

This section presents the modeling results, performance evaluation, ablation study, statistical analysis, and clinical implications of the Hybrid Filtering-based diagnostic recommendation system. All testing was conducted on 948 anonymous patient medical records that had undergone preprocessing, including medical term normalization, duplication removal, and the merging of diagnoses with clinical equivalence. The validity of the experiment was reinforced by applying 10-Fold Cross-Validation so that each data set played a balanced role as training data and test data.

The distribution of disease diagnoses in the dataset is shown in Table 3 to demonstrate the dominance of major classes such as ARI and gastritis, as well as the presence of minor classes that are relatively small in number but clinically important.

Table 3. Distribution of Diagnoses

Primary Diagnosis	Number of Cases
Acute Respiratory Infection	248
Gastritis	128
Cephalalgia (headache)	113
Myalgia	68
Contraceptive Acceptors	125
Vertigo	8
(Other minor categories)	258
Total	948

Feature representation using the Multi-View TF-IDF approach combines word-level n-grams (1–3) to capture semantic relationships and character-level n-grams (3–5) to minimize errors due to variations in the spelling of medical terms. The ablation study results in Table 4 show a consistent improvement in performance compared to the use of a single view. The Multi-View approach provides an improvement of up to +1.24% in Accuracy and F1-Score.

Table 4. *Multi-View TF-IDF Ablation Study*

Method	Accuracy	F1-Score
Word Only	86.71%	84.58%
Character Only	86.60%	84.39%
Multi-View	88.08%	85.82%

Performance evaluation of the system was conducted on five methods, namely Content-Based Filtering (CBF), Collaborative Filtering (CF), Naïve Bayes (NB), K-Nearest Neighbors (KNN), and Hybrid Filtering. The results in Table 3 show that Hybrid Filtering provides the best performance with an Accuracy of 87.37% and an F1-Score of 84.20%, outperforming all comparison models. The consistency of performance between models can be seen in Figure 5, where Hybrid maintains a stable performance margin across all validation folds.

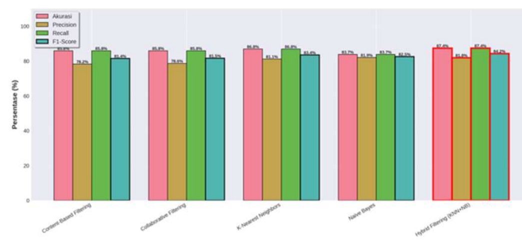


Figure 6. *Model Performance Comparison Chart*

Table 5. *Comparison of the performance of all methods*

Model	Accuracy	Precision	Recall	F1-Score
CBF	85.79	78.21	85.79	81.39
CF	85.79	78.55	85.79	81.50
NB	83.68	81.94	83.68	82.47
KNN	86.84	81.10	86.84	83.43
Hybrid	87.37	81.83	87.37	84.20

Further analysis at the class level is shown in Table 6, which indicates that the model has high sensitivity to major diagnoses such as Acute Respiratory Infection, Gastritis, and Contraceptive Acceptors, with an F1-Score above 96%. However, minor classes still show low performance due to limited sample size and overlapping symptoms. These misclassification patterns are shown in the Confusion Matrix in Figure 6, especially in diagnoses with similar clinical characteristics such as vertigo → ARI or odontalgia ↔ dental caries.

Jurnal Teknik Informatika Vol. 19 No. 1, April 2026 (62-72)
ISSN: p-ISSN 1979-9160 (Print)| e-ISSN 2549-7901 (Online)
DOI: https://doi.org/10.15408/jti.v19i1.48705

Table 6. *Evaluation results per diagnosis class*

Diagnosis	Precision	Recall	F1-Score
Acute Respiratory Infection	98.35%	96.37%	97.35%
Gastritis	99.18%	94.53%	96.80%
Contraceptive Acceptors	97.62%	98.40%	98.01%

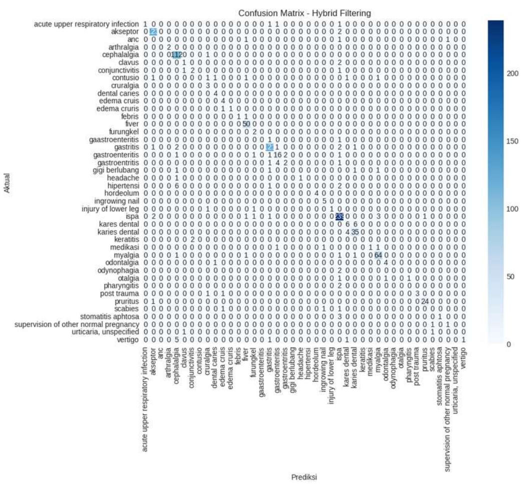


Figure 7. *Confusion Matrix model Hybrid*

Statistically, the improvement in Hybrid performance compared to the baseline is significant with a p-value of 0.018 < 0.05 based on the paired t-test, indicating that the increase in accuracy has a strong statistical basis and is not merely the result of random variation. This confirms that combining the generative power of Naïve Bayes with the instance-based approach in KNN can improve the system's ability to recognize both dominant and variable symptoms. From a clinical perspective, this system has the potential to support the initial triage process and rapid decision-making, especially in cases with common symptoms that have the potential to be misleading. However, data quality improvement strategies such as oversampling of minor classes or the addition of clinically-based data need to be implemented so that the model's performance is more consistent and safe to use across all disease categories.

CONCLUSION

This study developed a Hybrid Recommender System (HRS) based on a combination of Content-Based Filtering, Collaborative Filtering, KNN, and Naïve Bayes through a soft voting mechanism. The evaluation results show that the hybrid model provides the best performance with an accuracy of 87.37% and an F1-score of 84.20%, making it more effective in supporting initial diagnosis than individual methods. The integration of Multi-View TF-IDF also improves symptom representation, resulting in more stable predictions across various disease categories.

However, this study has limitations in terms of the relatively small dataset size, originating from a single healthcare facility, and does not yet cover rare diagnoses and has not been validated through clinical trials. Therefore, the generalization of the results still needs to be further examined. Future research could focus on expanding the multi-facility data sources, adding clinical features such as laboratory results and medical images, applying class balancing techniques, and evaluating through pilot implementation in healthcare facilities to test the effectiveness of the system in real operational conditions.

REFERENCES

[1] G. Costantini, A. S. Alanazi, A. S. Alanazi, And H. Benlaria, “Academic Editors: Daniele Giansanti Enhancing Healthcare For People With Disabilities Through Artificial Intelligence: Evidence From Saudi Arabia,” 2025, Doi: 10.3390/Healthcare.

[2] R. Ramadhani, R. Ramadhanu, A. Fajri, A. Abdillah, And M. Ridwan, “Studi Komparatif Multinomial Naïve Bayes, Decision Tree, Dan K-Nearest Neighbor Dalam Klasifikasi Validasi Ulasan Clash Of Clans Oleh Pengguna Ahli,” *Justin (Jurnal Sistem Dan Teknologi Informasi)*, Vol. 12, No. 4, Pp. 653–660, 2024, Doi: 10.26418/Justin.V12i4.81638.

[3] Zubayri, Abdulloh, Et Al. "Kualitas Layanan, Waktu Tunggu Dan Fasilitas Terhadap Kepuasan Pasien Pelayanan Medical Check Up Klinik Westerindo." *Jutranis* 1.02 (2024): 1-16

[4] Onsu, Romario, Daniel Febrian Sengkey, And Feisy Diane Kambey. "Implementasi Bi-Lstm Dengan Ekstraksi Fitur Word2vec Untuk Pengembangan Analisis Sentimen Aplikasi Identitas Kependudukan Digital." *Jurnal Teknologi Terpadu* 10.1 (2024): 46-55

ISSN: p-ISSN 1979-9160 (Print)| e-ISSN 2549-7901 (Online)
DOI: https://doi.org/10.15408/jti.v19i1.48705

[5] B. Sabiri, A. Khtira, B. El Asri, And M. Rhanoui, "Hybrid Quality-Based Recommender Systems: A Systematic Literature Review †," Jan. 01, 2025, *Multidisciplinary Digital Publishing Institute (Mdpi)*. Doi: 10.3390/Jimaging11010012.

[6] A. Rachmaniar, S. Widayati, K. Rokoyah, P. Studi Manajemen Informatika, P. Studi Sistem Informasi, And S. Jakarta Sti, "Sistem Rekomendasi Produk E-Commerce Menggunakan Collaborative Filtering Dan Content-Based Filtering (E-Commerce Product Recommendation System Using Collaborative Filtering And Content-Based Filtering)," *Journal Of Information System, Informatics And Computing Issue Period*, Vol. 9, No. 1, Pp. 40–54, 2025, Doi: 10.52362/Jisicom.V9i1.1904.

[7] Ko, Hyeyoung, Et Al. "A Survey Of Recommendation Systems: Recommendation Models, Techniques, And Application Fields." *Electronics* 11.1 (2022): 141

[8] Tran, Thi Ngoc Trang, Et Al. "Recommender Systems In The Healthcare Domain: State-Of-The-Art And Research Issues." *Journal Of Intelligent Information Systems* 57.1 (2021): 171-20

[9] W. Rojas-Carabali *Et Al.*, "Natural Language Processing In Medicine And Ophthalmology: A Review For The 21st-Century Clinician," Jul. 01, 2024, *Elsevier B.V.* Doi: 10.1016/J.Apjo.2024.100084.

[10] F. Farah And S. Dewi, "Peran Clinical Decision Support System (Cdss) Dalam Meningkatkan Efisiensi Dan Akurasi Triase Di Igd,"2025. [Online]. Available: Https://Www.Researchgate.Net/Publication/392510051

[11] R. M. Harahap And A. N. Rachman, "Sistem Rekomedasi Film Berdasarkan Kemiripan Deskripsi Cerita Menggunakan Metode Content Based Filtering," *Jurnal Informatika Dan Teknik Elektro Terapan*, Vol. 13, No. 3, Jul. 2025, Doi: 10.23960/Jitet.V13i3.6577.

[12] Javed, Umair, Et Al. "A Review Of Content-Based And Context-Based Recommendation Systems." *International Journal Of Emerging Technologies In Learning (Ijet)* 16.3 (2021): 274-306

[13] I. V. Derevitskii, N. D. Mramorov, S. D. Usoltsev, And S. V. Kovalchuk, "Hybrid Bayesian Network-Based Modeling: Covid-19-Pneumonia Case," *J Pers Med*, Vol. 12, No. 8, Aug. 2022, Doi: 10.3390/Jpm12081325.

[14] I. Akbar, F. Supriadi, And D. I. Junaedi, "Pemanfaatan Machine Learning Di Bidang Kesehatan," 2025.

[15] Damayunita, Aina, Rifqi Syamsul Fuadi, And Christina Juliane. "Comparative Analysis Of Naive Bayes, K-Nearest Neighbors (Knn), And Support Vector Machine (Svm) Algorithms For Classification Of Heart Disease Patients." *Jurnal Online Informatika* 7.2 (2022): 219-225.

[16] A. Yogianto, A. Homaidi, And Z. Fatah, "Implementasi Metode K-Nearest Neighbors (Knn) Untuk Klasifikasi Penyakit Jantung," *G-Tech: Jurnal Teknologi Terapan*, Vol. 8, No. 3, Pp. 1720–1728, Jul. 2024, Doi: 10.33379/Gtech.V8i3.4495.

[17] M. J. Faisti, R. Hadapiningradja Kusumodestoni, G. Wahyu, And N. Wibowo, "Mental Health Classification Using Naïve Bayes And Random Forest Algorithms," 2025. [Online]. Available: Http://Jurnal.Polibatam.Ac.Id/Index.Php/Jaic

[18] R. Widayanti, M. Heru, R. Chakim, C. Lukita, U. Rahardja, And N. Lutfiani, "Improving Recommender Systems Using Hybrid Techniques Of Collaborative Filtering And Content-Based Filtering," *Journal Of Applied Data Sciences*, Vol. 4, No. 3, Pp. 289–302, 2023.

[19] F. Amandasari And D. Damayanti, "Perbandingan Kinerja Support Vector Machine Dan Naive Bayes Dalam Klasifikasi Sentimen Twitter Terhadap Pelayanan Bpjs," *Jurnal Pendidikan Dan Teknologi Indonesia*, Vol. 5, No. 3, Pp. 645–653, Mar. 2025, Doi: 10.52436/1.Jpti.680.

ISSN: p-ISSN 1979-9160 (Print)| e-ISSN 2549-7901 (Online)
DOI: https://doi.org/10.15408/jti.v19i1.48705

[20] M. D. Purbolaksono, M. I. Tantowi, A. I. Hidayat, And Adiwijaya, "Perbandingan Support Vector Machine Dan Modified Balanced Random Forest Dalam Deteksi Pasien Penyakit Diabetes," *Jurnal Resti*, Vol. 5, No. 2, Pp. 393–399, Apr. 2021, Doi: 10.29207/Resti.V5i2.3008.

[21] N. Noor Kamala Sari, L. Licantik, And M. Zahra, "Pemanfaatan Sistem Rekomendasi Menggunakan Content-Based Filtering Pada Hotel Di Palangka Raya," *Technologia : Jurnal Ilmiah*, Vol. 15, No. 4, P. 754, Oct. 2024, Doi: 10.31602/Tji.V15i4.16282.

[22] A. Misbullah, M. Akbar, And L. Farsiah, "Sistem Rekomendasi Pemilihan Program Studi Berbasis Hybrid Menggunakan Pendekatan Deep Learning," *Jurnal Pendidikan Teknologi Informasi*, Vol. 9, No. 1, Pp. 75–87, 2025.

[23] A. Tantri Yanuar, Z. N. Aini, A. M. Samhan, D. J. Supriyanto, M. Muhammad, And I. Al'azzam, "Implementasi Sistem Rekomendasi Dokter Berbasis Content-Based Filtering Pada Layanan Konsultasi Kesehatan," *Seminar Nasional Teknologi Informasi Dan Bisnis (Senatib)*, P. 2025.

[24] M. Uta *Et Al.*, "Knowledge-Based Recommender Systems: Overview And Research Directions," 2024, *Frontiers Media Sa*. Doi: 10.3389/Fdata.2024.1304439.

[25] H. A. Dwi Fasnuari, H. Yuana, And M. T. Chulkamdi, "Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Penyakit Diabetes Melitus," *Antivirus : Jurnal Ilmiah Teknik Informatika*, Vol. 16, No. 2, Pp. 133–142, Oct. 2022, Doi: 10.35457/Antivirus.V16i2.2445.

[26] I. Sedek Pakaphaan, I. Sri Dumayanty, S. Manurung, And I. Artikel Abstrak Histori Artikel, "Sistem Pakar Deteksi Penyakit Tanaman Cabai Metode Naïve Bayes," 2024. [Online]. Available: Http://Ojs.Fikom-Methodist.Net/Index.Php/Methotika

[27] V. Chiruvella And A. K. Guddati, "Ethical Issues In Patient Data Ownership," *Interact J Med Res*, Vol. 10, No. 2, P. E22269, May 2021, Doi: 10.2196/22269.