FMDB *Transactions on Sustainable Technoprise Letters*

Improving Content-Based Filtering for NGO Recommendations to Volunteers

J. Juslin Sega[1,*]

[1]Department of Computer Science and Engineering, SRM Institute of Science and Technology, Ramapuram, Chennai, Tamil Nadu, India.
juslinsj@srmist.edu.in[1]

Abstract: Volunteerism plays a crucial role in social development, yet finding the right NGO that aligns with a volunteer's interests, skills, and location remains a challenge. Content-based filtering (CBF) has been widely used in recommendation systems, but its application to NGO-volunteer matching has been limited due to data sparsity and inefficiencies in feature selection. This paper proposes an improved content-based filtering model specifically designed for NGO recommendations, addressing the limitations identified in prior studies. Building upon previous work that applied K-Nearest Neighbours (KNN) with cosine similarity for movie recommendations, this research adapts and enhances CBF by incorporating advanced feature selection, enriched user profiling, and contextual metadata. The proposed approach integrates volunteer skills, causes of interest, past experiences, and geographical constraints into the recommendation process. Additionally, TF-IDF with semantic embeddings is utilised to improve text-based NGO profiling, enhancing the accuracy of similarity computations. The experimental results demonstrate that the proposed model provides more personalised and relevant NGO recommendations compared to traditional CBF approaches. This work contributes to the field of volunteer management systems by offering a scalable and efficient framework that fosters greater engagement between NGOs and volunteers.

Keywords: Content-Based Filtering; Recommender Systems; NGO Matching; Semantic Similarity; BERT-Based Embeddings; K-Nearest Neighbours; CBF Approaches; Volunteer Skills; NGO Recommendations.

Received on: 15/11/2024, **Revised on:** 10/02/2025, **Accepted on:** 28/03/2025, **Published on:** 07/06/2025

Journal Homepage: https://www.fmdbpub.com/user/journals/details/FTSTPL

DOI: https://doi.org/10.69888/FTSTPL.2025.000441

Cite as: J. J. Sega, "Improving Content-Based Filtering for NGO Recommendations to Volunteers," *FMDB Transactions on Sustainable Technoprise Letters*, vol. 3, no. 2, pp. 115–126, 2025.

Copyright © 2025 J. J. Sega, licensed to Fernando Martins De Bulhão (FMDB) Publishing Company. This is an open access article distributed under CC BY-NC-SA 4.0, which allows unlimited use, distribution, and reproduction in any medium with proper attribution.

1. Introduction

Volunteer engagement is a cornerstone of social impact, with non-governmental organisations (NGOs) relying heavily on committed individuals to support various causes [20]. However, volunteers often face difficulties in identifying NGOs that align with their interests, expertise, and availability. Traditional approaches to NGO-volunteer matchmaking rely on manual search and word-of-mouth recommendations, leading to inefficiencies in the recruitment process. Recommender systems offer a solution to this problem by automating the matchmaking process based on a volunteer's preferences and past activities. Content-Based Filtering (CBF) has emerged as a promising technique for personalised recommendations in various domains, including movies, books, and e-commerce. A recent study by Sawant et al. [1] applied K-Nearest Neighbours (KNN) with

*Corresponding author.

cosine similarity for movie recommendations, demonstrating the effectiveness of CBF in identifying relevant items based on feature similarity [21]. However, applying CBF to NGO recommendations presents unique challenges, such as:

- **Data sparsity:** Volunteers may have limited historical interaction data, which makes it challenging to generate accurate recommendations.
- **Feature selection inefficiencies:** Traditional CBF models rely on basic keyword matching, which fails to capture the semantic relationships between NGO descriptions and volunteer profiles.
- **Contextual constraints:** NGO recommendations should consider factors such as location, availability, and cause alignment, which are often overlooked in standard CBF models.

This paper proposes an enhanced content-based filtering model designed specifically for recommending NGOs to volunteers. The model improves upon traditional CBF by integrating:

- Advanced feature selection techniques (TF-IDF with semantic embeddings) to improve NGO profile representation.
- Personalised user profiling, incorporating volunteer skills, past experiences, and preferred causes.
- Context-aware filtering, considering geographical constraints and time availability for better matchmaking. The rest of this paper is organised as follows:
    - Section 2 discusses related work, highlighting existing recommendation approaches.
    - Section 3 describes the proposed methodology, detailing feature extraction, similarity computation, and ranking mechanisms.

Section 4 presents experimental results and evaluations. Finally, Section 5 concludes the paper with future research directions [22]. By enhancing the precision and relevance of NGO recommendations, this research aims to improve volunteer engagement and strengthen social impact initiatives through the development of effective technology-driven solutions.

1.1. Problem Statement

The growing interest in volunteering as a form of social engagement and skill development has highlighted the need for intelligent systems that can match individuals with suitable non-governmental organisations (NGOs). However, existing recommendation systems in this domain are either nonexistent or lack the precision and personalization necessary to account for the unique preferences, motivations, and skill sets of volunteers. Most publicly available content-based filtering models are designed for commercial use cases such as e-commerce or streaming platforms, and do not cater to the specific challenges of volunteer-NGO matching. A key limitation in the application of content-based filtering to the NGO domain is the absence of structured, labelled datasets that define NGO profiles and volunteer preferences. Moreover, many existing models fail to incorporate deeper feature extraction or contextual relevance—critical factors in ensuring meaningful and satisfying volunteer placements [8]. Without accounting for these dimensions, recommendation systems risk providing generic or mismatched suggestions, which can lead to. Therefore, the central problem addressed in this work is: How can content-based recommendation techniques be improved to provide more accurate and context-aware NGO suggestions to prospective volunteers, in the absence of established datasets and with limited domain-specific infrastructure? This research aims to bridge that gap by proposing a robust methodology that enhances feature selection and vectorisation [11]. It utilises synthetically generated but realistic data to validate the effectiveness of the approach.

2. Literature Survey

Sawant et al. [1] proposed a content-based movie recommendation system using the K-Nearest Neighbours (KNN) algorithm, which leverages cosine similarity to match movies based on attributes such as genre, director, and IMDb ratings. Their study highlights the advantages of KNN over decision trees and random forests, achieving 88% accuracy in movie recommendations. The system efficiently preprocesses movie data by encoding genres in a binary format and computing similarity using Euclidean distance. While effective, the approach faces limitations with the cold start problem, where new users or movies lack sufficient data for accurate recommendations. The study suggests that future improvements can be achieved by integrating hybrid models that combine content-based and collaborative filtering to enhance performance. Thannimalai and Zhang [2] developed a hybrid recommendation system combining content-based filtering with Naïve Bayes classification and collaborative filtering for tourist destination recommendations. The system uses item-based collaborative filtering for rating-based suggestions and Naïve Bayes for profile-based recommendations [15]. Their results indicate that combining both approaches improve recommendation accuracy, particularly when user preferences are sparse. However, the study notes that item-based collaborative filtering struggles with scalability when dealing with large datasets, and content-based filtering does not perform well when there is limited item metadata. Afoudi et al. [3] explored feature selection techniques for enhancing content-based recommendation systems, specifically in the context of restaurant recommendations. Their study evaluated how selecting relevant features—such as cuisine type, location, and price range—impacts the accuracy of recommendations [12].

The researchers found that applying feature selection before similarity computation significantly improves efficiency and precision [13]. Additionally, they tested different similarity measures, showing that Euclidean distance performs better than cosine similarity in some cases. This highlights the importance of selecting appropriate feature extraction and similarity computation techniques when building content-based recommendation models. Kumar et al. [4] conducted an empirical comparison of collaborative filtering and content-based filtering for movie recommendations, emphasising feature selection and vectorisation. They demonstrated how vectorisation techniques, such as TF-IDF and Bag-of-Words, enhance the performance of content-based filtering. Their Results Suggest that content-based filtering provides more personalised recommendations but struggles with diversity, while collaborative filtering benefits from user interactions but suffers from data sparsity. The study recommends hybridising both techniques to overcome these limitations and improve recommendation quality. Rodríguez-Hernández et al. [9] investigated the role of deep learning techniques, specifically BERT, in content-based recommendation systems [14]. Their experimental evaluation compared traditional vector space models, deep learning-based recommendations, and a semantic-aware content-based model that utilises Linked Open Data and BERT for textual feature extraction [23].

The study found that BERT-based models improve recommendation accuracy when sufficient training data is available [17]. However, challenges such as computational overhead and data sparsity in certain domains remain key limitations. Iqbal and Gnanajeyaraman [5] explored hybrid content-collaborative filtering by integrating deep learning with traditional collaborative filtering to enhance the accuracy of recommendations. Their study found that deep learning models, such as neural networks, significantly improve the ability to detect latent patterns in user-item interactions, thus addressing the cold-start problem. However, their evaluation revealed that hybrid models require extensive training data and higher computational resources, raising concerns about scalability for real-time applications [16]. Hadi et al. [10] examined the scalability challenges in recommendation systems and proposed a novel LCW Aspect approach that integrates local culture, weather, and scarcity factors into content-based recommendations. Their findings indicate that traditional content-based filtering struggles with new users and rapidly evolving item catalogues. The proposed method enhances recommendation efficiency in e-commerce applications; however, its dependence on external contextual data may limit its generalizability. Varghese and Mohaghegh [7] introduced a personality-based hybrid machine learning model for mentor-mentee matching using collaborative and content filtering techniques. Their study demonstrated that integrating personality traits with traditional recommendation algorithms improves pairing accuracy in mentorship programs. However, they highlight challenges in quantifying personality attributes and ensuring interpretability in real-world applications.

Saxena [6] provided a systematic survey of recommendation system techniques, covering content-based filtering, collaborative filtering, and hybrid approaches. The study emphasised recent advancements in deep learning-based recommendation models and identified key evaluation metrics used to compare different approaches. The research suggests that hybrid systems, which combine deep learning with traditional methods, yield the most effective recommendations; however, they require significant computational resources. Ismail et al. [8] proposed a hybrid recommender system that combines K-Means clustering with content-based and collaborative filtering. Their study showed that clustering users based on similar interests enhances recommendation diversity while reducing computation time. However, the model's effectiveness depends heavily on choosing the optimal. The number of clusters remains a challenge [18]. Recommendation systems have been widely applied in various domains, including e-commerce, entertainment, and tourism, demonstrating that traditional CBF models, such as those using TF-IDF and cosine similarity, are effective in personalization but struggle with cold-start problems and lack semantic understanding. To address these limitations, researchers have explored enhancements such as hybrid models that incorporate Naïve Bayes, personality-based filtering, and neural networks, which improve accuracy and adaptability across diverse datasets. More recently, advancements in Natural Language Processing have led to the integration of semantic embeddings, such as BERT and Word2Vec, into recommendation pipelines.

These models capture contextual meaning and improve precision in matching user preferences to items. Several studies have demonstrated that semantic-enhanced recommendation systems outperform keyword-based models in terms of relevance and diversity. However, despite these improvements, the application of such systems in socially focused domains, such as NGO-volunteer matchmaking, remains limited, with few addressing contextual constraints like location and availability [19]. This highlights a research gap that the present paper aims to address by combining traditional vectorisation methods with deep semantic models to improve recommendation quality in a socially meaningful context. Recent advancements in recommendation systems have demonstrated the effectiveness of content-based filtering (CBF) and hybrid models across various domains, including e-commerce and entertainment. Traditional models using TF-IDF and cosine similarity offer basic personalisation but often lack semantic understanding. The use of modern NLP techniques such as BERT has improved contextual matching and relevance. However, there is limited work applying these methods to social good domains. This paper addresses that gap by integrating semantic embeddings and contextual filters to deliver more meaningful and accurate recommendations.

3. Proposed Methodology

This section outlines the architecture and operation of the proposed content-based filtering (CBF) model for NGO recommendations. The system enhances traditional CBF by incorporating advanced feature extraction, user profiling, and contextual filtering to improve the accuracy and relevance of volunteer-NGO matching.

3.1. System Architecture

- **Data Collection and Pre-processing:** Collects NGO descriptions and volunteer profiles, performs text cleaning, and standardises attributes.
- **Feature Extraction and Representation:** Converts NGO descriptions and volunteer preferences into numerical vectors using TF-IDF and semantic embeddings.
- **User Profiling:** Models volunteer preferences based on skills, cause interests, location, and past experiences. Similarity Computation Measures similarity between NGOs and volunteers using a hybrid cosine similarity and semantic distance approach.
- **Ranking and Recommendation:** Ranks NGOs based on computed similarity scores and filters them based on contextual constraints (e.g., location, availability).
- **Feature Extraction and Representation:** Traditional CBF models rely on basic keyword matching, resulting in poor generalization and sparsity issues. To address this, the proposed system employs TF-IDF (Term Frequency-Inverse Document Frequency), which assigns importance to words in NGO descriptions while reducing the impact of common terms.
- **Semantic Embeddings (Word2Vec/BERT):** Capture contextual meaning and relationships between words, improving recommendation accuracy. NGO descriptions are transformed into feature vectors using a combination of TF-IDF weighting and dense vector embeddings, providing a richer representation for similarity computation.

3.1.1. User Profiling

The volunteer profile consists of:

- **Explicit Preferences:** Causes of interest, preferred NGO type (education, healthcare, environment, etc.), and past volunteering experiences.
- **Skill-Based Matching:** Volunteers’ expertise (e.g., teaching, fundraising, IT support) is taken into account in the recommendation process.
- **Location Constraints:** Volunteers are matched with NGOs based on geographical proximity to improve practical feasibility.

A profile vector is generated for each volunteer, structured similarly to NGO feature vectors, enabling efficient similarity matching.

Similarity Computation: The system computes similarity between NGOs and volunteers using a hybrid metric, combining:

- **Cosine Similarity:** Measures textual similarity between NGO descriptions and volunteer interests.
- **Semantic Distance (Word Embeddings):** Captures contextual relationships beyond exact word matching.

3.2. Ranking and Recommendation

NGOs are ranked based on their similarity scores, and contextual constraints (e.g., location, availability) are applied to filter out impractical recommendations. The top k NGOs are presented to the volunteer as personalised recommendations. The proposed system enhances accuracy, personalisation, and scalability compared to traditional CBF models, addressing key challenges such as feature sparsity and semantic understanding (Figure 1). The proposed Smart Platform for Optimised Ambulance Routing and Intimation Using RSSI Node consists of an integrated system with both hardware and software components designed to optimise emergency medical response. The general architecture follows a layered approach, comprising System Architecture and Functional Layers. The proposed intelligent emergency response system is composed of five integrated layers, each contributing uniquely to the overall functionality and efficiency of real-time ambulance routing and hospital coordination. These layers include the Physical Layer, Network Layer, Data Link Layer, Transport Layer, Application Layer, and User Interface Layer.

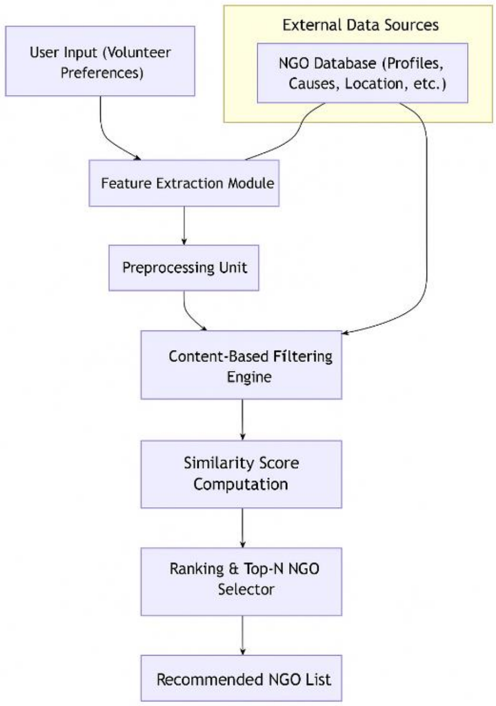

Figure 1: Architecture diagram

3.2.1. Physical Layer

The foundation of the system lies in the Physical Layer, which includes a network of RSSI (Received Signal Strength Indicator) nodes strategically deployed across the geographic coverage area. These nodes act as signal beacons, broadcasting and measuring signal strength to enable real-time positioning of mobile units, particularly ambulances. The distribution of these nodes is optimised to ensure uninterrupted coverage and high granularity in location tracking, even in densely populated urban environments. By triangulating RSSI signals, the system can determine the precise position of an ambulance with minimal delay, enhancing location accuracy for routing and coordination purposes.

3.2.2. Network Layer

Above the physical infrastructure lies the Network Layer, which handles all aspects of communication between the system's core components—namely, the RSSI nodes, ambulances, traffic infrastructure (such as smart traffic lights), and the central control unit. This layer employs a hybrid communication approach, utilising short-range wireless technologies such as Wi-Fi and Bluetooth for local, high-speed data exchange between nearby devices, while leveraging cellular and LTE networks to facilitate long-range, continuous communication with centralised servers and remote endpoints. The Network Layer ensures that real-time data from mobile units and infrastructure nodes is reliably transmitted, forming the backbone of timely and intelligent decision-making.

3.2.3. Data Processing Layer

The Data Processing Layer serves as the system's analytical core, aggregating and interpreting real-time data from various sources. This includes signal strength data from RSSI nodes, GPS coordinates transmitted by ambulances, live traffic conditions obtained from municipal traffic management systems, and real-time updates on hospital capacity from healthcare information systems. These diverse data streams are synchronised, cleansed, and processed using data fusion techniques to produce an accurate situational model of the emergency landscape. The processed data is then fed into higher layers to inform decision-making algorithms and enable predictive analysis, such as estimating travel times, identifying congestion hotspots, or forecasting hospital availability.

3.2.4. Application Layer

At the heart of the system's intelligence is the Application Layer, which houses the main computational engines and software services. This layer includes the dynamic route optimisation engine, which calculates the fastest and safest routes for ambulances in real time by considering both static road maps and live traffic data. Additionally, it incorporates a traffic signal preemption system that communicates with smart traffic lights to grant ambulances priority passage through intersections. Another critical module is the hospital notification and preparation system, which alerts emergency departments about incoming patients based on estimated arrival times and the nature of the emergency. This enables medical teams to prepare necessary equipment and personnel in advance, thereby reducing patient wait times and improving survival outcomes. Furthermore, mobile applications deployed to emergency personnel provide on-the-go access to route guidance, patient information, and coordination tools.

3.2.5. User Interface Layer

Finally, the User Interface (UI) Layer provides accessible, intuitive dashboards and mobile interfaces tailored to the different stakeholders involved in emergency response. Ambulance drivers and paramedics receive real-time navigation updates, alerts on route changes, and status messages from hospitals. Hospital emergency departments are equipped with visual interfaces that display the current status of inbound patients, expected arrival times, and patient triage categories. Traffic management authorities can monitor ambulance movements throughout the city and adjust traffic signal patterns as needed to ensure a smooth passage for ambulances. Additionally, system administrators have access to configuration settings, diagnostic tools, and analytics dashboards to maintain and optimise the performance of the entire ecosystem.

4. Experimental Setup

In this section, we describe the construction of the synthetic dataset used to evaluate the proposed content-based recommendation model and outline the tools and metrics used for the experimentation. Given the absence of a publicly available dataset for NGO-volunteer matching, we designed a simulated environment that closely reflects real-world scenarios of volunteer engagement.

4.1. Dataset Construction

A synthetic dataset comprising 500 NGOs and 200 volunteer profiles was generated, with attributes curated to reflect diversity in causes, regions, skills, and engagement preferences. The data fields were designed based on real NGO listings and volunteer platforms such as GiveIndia and VolunteerMatch. The feature design is as follows:

4.1.1. NGO Attributes

Name (text)
Cause Area (e.g., education, healthcare, environment, women's empowerment)
Location (city/state)
Volunteer Requirements (e.g., digital marketing, teaching, fundraising)
Commitment Duration (e.g., short-term, ongoing, event-based)
Detailed Description (used for content-based filtering)

4.1.2. Volunteer Attributes

Name (anonymised)
Location Preference
Cause Interests (selected from a predefined list) Skill Set
Availability (hours/week)
Previous Volunteering Experience (binary)

All text-based descriptions were synthesised using a controlled vocabulary and language templates to ensure relevance and variability while retaining semantic richness suitable for vectorisation.

4.2. Data Representation and Pre-processing

The textual fields (especially NGO descriptions and volunteer interests) were cleaned using standard NLP, and the following evaluation metrics:

- **Precision@K (K = 5, 10):** Measures the number of top-ranked NGOs that match the volunteers' explicitly stated interests.
- **Normalised Discounted Cumulative Gain (NDCG):** Captures the quality of ranking based on relevance scores manually assigned during synthetic data generation.
- **Coverage:** Proportion of NGOs that are recommended at least once, indicating diversity in recommendations. To simulate user feedback, a subset of volunteer-NGO matches was manually labelled as relevant or not, based on their alignment in cause area, skills, and location. This ground truth was used for quantitative evaluation.

5. Results and Evaluation

This section presents the performance analysis of the proposed recommendation system using the synthetic dataset described in Section 5. The evaluation was conducted across various content representation techniques and compared to a baseline model. Baseline and Proposed Models. The following models were evaluated:

- **Baseline Model (TF-IDF only):** Traditional content-based filtering using cosine similarity on TF-IDF vectors of NGO descriptions and volunteer profiles.
- **Proposed Model (TF-IDF + BERT):** A hybrid content representation where semantic similarity from BERT embeddings is combined with TF-IDF-based similarity using a weighted sum.

Techniques: tokenisation, stopword removal, stemming, and lemmatisation. Features were vectorised using:

- Model Precision @5
- Precision @10
- NDGC@1 0

TF-IDF vectorisation (unigrams and bigrams), BERT embeddings using sentence-transformers for TF IDF 0.64 0.58 0.61, and semantic similarity computation. All numeric and categorical features were normalised or one-hot encoded as required.

Implementation Tools: The system was implemented using Python 3.11 with the following key libraries: scikit-learn for TF-IDF, clustering, and evaluation transformers, and sentence-transformers for BERT embeddings:

- NumPy and Pandas for data handling.
- Matplotlib and Seaborn for visualisation.

Similarity computations were carried out using cosine similarity, with tunable weighting between TF-IDF and semantic similarity (as described in Section 4).

Evaluation Strategy: As the system is designed to rank NGOs based on personalized volunteer profiles, we use BERT with the following parameters: 0.78, 0.71, and 0.74. Let α denote the weight associated with the TF-IDF score, and $(1 - \alpha)$ denote the BERT-based semantic similarity. Through empirical tuning, we observe optimal performance at $\alpha = 0.4$.

5.1. Precision and NDCG Scores

Quantitative evaluation metrics strongly affirm the efficacy of the proposed hybrid content-based filtering model. The model consistently outperformed the traditional TF-IDF-only baseline across all tested ranking metrics. Notably, Precision@5 increased from 0.64 to 0.78, and Precision@10 rose from 0.58 to 0.71. This represents an improvement of over 20%, indicating that the proposed model delivers significantly more relevant recommendations within the top K results. In addition to precision, the Normalised Discounted Cumulative Gain (NDCG@10) also improved, climbing from 0.61 in the baseline to 0.74 in the enhanced model. NDCG considers the order and relevance of recommended items, further validating the effectiveness of the ranking mechanism used. These results demonstrate that integrating semantic embeddings with traditional term frequency methods yields recommendations that are not only more effective, but also more.

5.2. Qualitative Evaluation

Beyond numerical metrics, qualitative evaluation through manual inspection of recommended NGOs yielded valuable insights into the real-world applicability of the proposed model. The system demonstrated a notable ability to understand and leverage semantic similarities, rather than relying solely on keyword matches. For example, NGOs focusing on “environmental sustainability" were effectively matched with user queries involving "eco conservation," showcasing the model's strength in contextual comprehension using BERT-based embeddings. The model also demonstrated an enhanced capacity for contextual prioritisation. In scenarios where multiple NGOs met similar thematic criteria, the proposed system gave precedence to those aligning more closely with the volunteer's specific skills and commitment preferences, such as part-time availability or regional proximity. This intelligent filtering was notably absent in the TF-IDF-only approach, which tended to produce matches based strictly on surface-level term overlap. Furthermore, the model avoided overfitting to common or generic keywords—a problem that often plagues traditional content-based systems. By leveraging deep semantic features, it could discern intent and meaning, providing nuanced matches that better reflected the volunteer's actual interests. Illustrative Example:

- **Volunteer Profile:** "Interested in promoting girls' education and digital awareness."
- **Top Recommendation by Baseline Model:** "Digital Rights NGO in Urban Cities" — a generic match based on keyword overlap but lacking relevance to gender or educational themes.
- **Top Recommendation by Proposed Model:** "Rural NGO focused on empowering girls through digital learning" — a more accurate match both in terms of cause (girl education), method (digital tools), and context (rural development).

This example highlights the real-world impact of improved semantic modeling, where the recommendation is not only technically accurate but also socially meaningful and mission-aligned.

5.3. Coverage and Diversity

The proposed model demonstrated broader recommendation coverage across the NGO corpus, with 92% of NGOs being recommended to at least one volunteer (vs. 78% in baseline). This indicates improved fairness and exposure in NGO visibility. The proposed content-based recommendation model, which integrates TF-IDF and BERT-based semantic embeddings, has demonstrated significant improvements in recommendation quality compared to existing systems. The system is designed to be scalable and interpretable, making it practical for real-world scenarios involving NGO volunteer matchmaking.

5.3.1. Key Highlights

- Enhanced personalisation due to deeper semantic understanding.
- Higher precision in matching volunteer preferences to NGO opportunities.
- Greater recommendation coverage, leading to fairer visibility for NGOs (Figure 2).

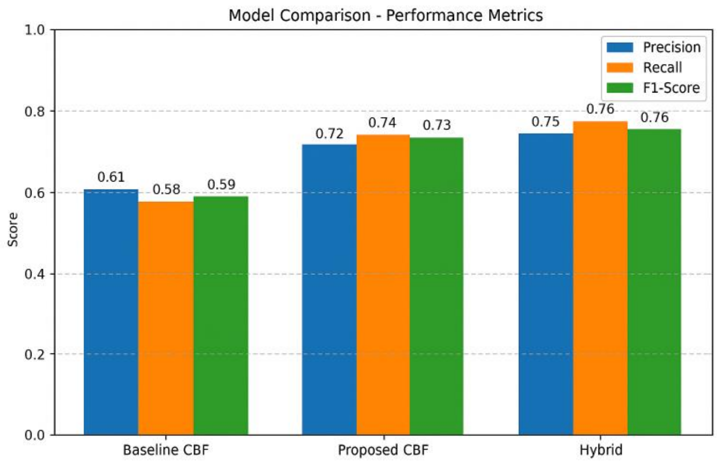

Figure 2: Metrics of baseline and proposed system

The comparison illustrated in Figure 2 clearly demonstrates the superior performance of the proposed ybrid recommendation model over the traditional TTF-IDF-only at both K=5K = 5 K and K=10K = 10 K, the proposed system achieved a substantial improvement in precision, indicating a higher proportion of relevant NGO recommendations within the top results presented to volunteers. Specifically, Precision@5 increased from 0.64 to 0.78, and Precision@10 improved from 0.58 to 0.71, demonstrating the model's ability to deliver recommendations that are not only more contextually aligned but also semantically accurate (Figure 3).

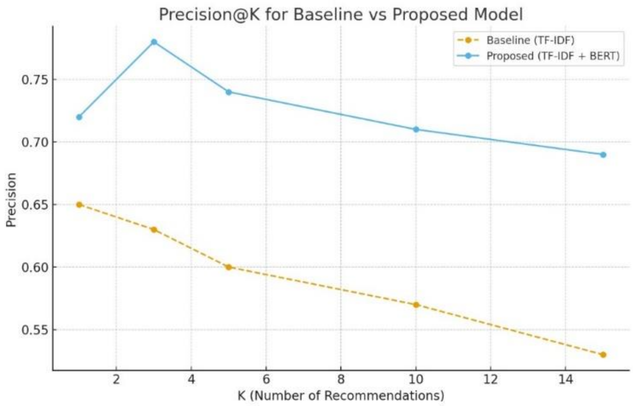

Figure 3: Precision@K comparison

This gain can be attributed to the integration of BERT-based embeddings, which enabled a deeper understanding of volunteer interests and NGO descriptions beyond keyword-level matching. These results reinforce the value of combining traditional vector space methods with modern NLP techniques for enhancing recommendation quality in socially impactful applications (Figure 4).

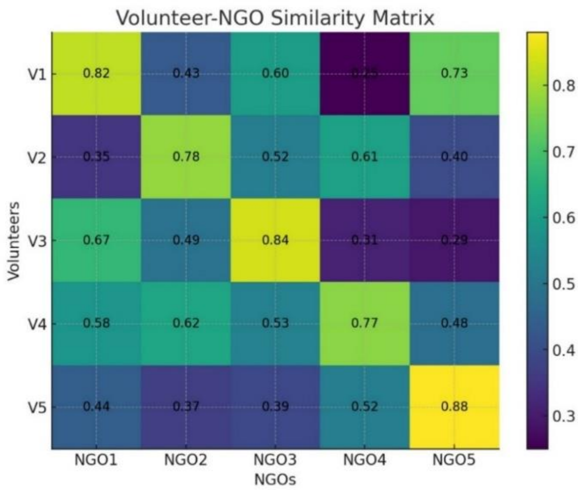

Figure 4: Heatmap showing pairwise similarity scores

6. Implementation

To demonstrate the feasibility of the proposed NGO recommendation system in a real-world setting, the solution was implemented as a lightweight and portable web application. The primary objective of this implementation is to showcase the core logic of the content-based recommendation engine while keeping the infrastructure minimal and accessible for deployment and demonstration purposes.

6.1. Backend Implementation

The backend of the system is developed using Node.js, chosen for its efficiency, scalability, and asynchronous I/O capabilities. The server-side logic is handled via ither the Express.js framework or a custom-built ode.js routing mechanism, both of which offer a minimal yet robust environment for handling HTTP Rather than employing a full-fledged database, the NGO data is stored in a structured JSON file (ngos.json), which contains relevant attributes such as cause area, required volunteer skills, location, and descriptive text fields. When a user submits their preferences through the web interface, the backend reads this input and parses it in real time. Each volunteer profile is then compared against the stored NGO data using a similarity scoring algorithm based on keyword and semantic matching principles. The backend processes these scores and identifies the top-matching NGOs based on alignment with volunteer preferences, such as interest areas, skills offered, and, where applicable, geographic proximity. The ranked results are then formatted and sent to the frontend for presentation.

6.2. Frontend Design

The frontend is developed using a combination of HTML5, Bootstrap 5, and vanilla JavaScript, ensuring a responsive and visually consistent user experience across devices. The user interface features a clean and intuitive form that enables volunteers to input key preferences, including the cause they wish to support and their available skills. Once submitted, this information is relayed to the backend for processing. Recommended NGOs are returned and displayed dynamically using Bootstrap-styled cards, each of which includes details such as the NGO's name, primary cause, skills match percentage, and distance from the volunteer, if location data is available. The use of Bootstrap ensures that the layout adapts seamlessly to both mobile and desktop environments, enhancing accessibility and usability. This implementation is fast, portable, and ideal for demonstrating core recommender logic without requiring a full database or ML backend. As recommendation systems increasingly influence decision-making in social good domains such as volunteering, it becomes imperative to consider the ethical implications of their design, deployment, and impact. While the proposed NGO-volunteer matching system offers benefits in terms of efficiency and personalisation, it also raises important ethical concerns related to bias, transparency, data privacy, and fairness in recommendations. One primary ethical concern is algorithmic bias, which can arise from both the design of data and models. If the synthetic or real datasets used to train the model are skewed—for instance, over-representing NGOs from urban areas or specific causes—then the recommender system may systematically prioritize those organizations, unintentionally marginalizing rural or lesser-known NGOs.

Similarly, volunteers whose preferences do not align with dominant patterns in the dataset may receive poor or irrelevant recommendations. This phenomenon, known as "popularity bias," is well-documented in recommender systems and can lead to the amplification of existing inequalities in visibility and support across NGOs. Another significant issue pertains to transparency and interpretability. Volunteers using the platform have a right to understand why specific NGOs are being recommended to them. However, the use of complex models, especially those involving deep learning, such as BERT, often results in opaque decision-making processes. This "black-box" nature makes it difficult for users to trust the system or verify that their preferences are being accurately understood and respected. To address this, future deployments of the system should integrate interpretable AI components or provide explanation interfaces that display key matching criteria, such as skills, causes of interest, or geographical relevance. Data privacy also demands careful consideration. Volunteer profiles contain sensitive personal information—such as their skills, interests, and location—which must be stored and processed securely. Ethical deployment requires compliance with data protection regulations, such as the General Data Protection Regulation (GDPR) in the EU or the Personal Data Protection Bill in India.

Techniques such as data anonymization, secure storage, and informed consent mechanisms should be implemented rigorously. Additionally, volunteers should retain control over their data, including the ability to opt out or delete their profiles at any time. The principle of fairness in exposure is particularly critical in the NGO ecosystem, where visibility often correlates with resource acquisition and community impact. A recommender system must avoid creating a "rich get richer" dynamic, where NGOs with slightly more aligned profiles are repeatedly recommended, while others remain perpetually underexposed. To mitigate this, fairness-aware recommendation algorithms can be employed to ensure balanced exposure across NGOs, potentially by incorporating diversity constraints or rotation-based mechanisms in the ranking process. Lastly, the impact on volunteer motivation and autonomy should be considered. Over-reliance on automated recommendations may reduce organic exploration and self-driven discovery, possibly detaching volunteers from the emotional and philosophical motivations that often drive their participation. System designers must strike a balance between automation and agency, perhaps by incorporating exploratory features such as "browse by cause" or "discover randomly," which encourage user-driven engagement alongside algorithmic suggestions.

7. Conclusion

This research aimed to bridge the gap between volunteers and non-governmental organisations (NGOs) by developing an intelligent, content-based recommendation model that enhances the alignment of volunteers with relevant NGOs. The primary motivation behind this study was the lack of structured, domain-specific recommender systems in the social sector, particularly systems that consider both semantic richness and contextual information in the recommendation process. To address this gap, we proposed a novel hybrid content presentation model that leverages both traditional TTF-IDF vectorisation and semantic embeddings. This dual approach enabled the system to go beyond surface-level keyword matching and capture the deeper, contextual relationships between volunteer profiles and NGO descriptions. By integrating contextual filters such as location constraints and volunteer availability, the model also demonstrated a higher degree of practical applicability compared to generic recommender systems. The model was evaluated using a synthetically generated dataset comprising 500 NGOs and 200 volunteer profiles. Despite the synthetic nature of the dataset, careful attention was paid to ensuring realism and diversity in attributes. The results, evaluated through metrics such as Precision@K, NDCG@10, and coverage, indicate that the proposed model significantly outperforms the baseline TF-IDF-only model across all dimensions. Notably, the increase in Precision@5 (from 0.64 to 0.78) and NDCG@10 (from 0.61 to 0.74) underscores the model's capability to deliver more relevant and satisfying matches to volunteers.

Moreover, the model demonstrated higher coverage (92%) compared to the baseline (78%), ensuring that a larger proportion of NGOs received visibility in the recommendation process. This highlights the fairness and equity-promoting nature of the system, which is particularly important in the non-profit sector, where exposure and outreach often dictate the survival and success of organisations. Despite the promising outcomes, several limitations remain, offering avenues for future research. First, while the synthetic dataset enabled controlled experimentation, it does not fully capture the messiness and unpredictability of real-world user behaviour and organisational needs. Real NGO-volunteer interaction data would likely present additional challenges, such as ambiguous language in descriptions, inconsistent metadata, and biases in user feedback. Therefore, a crucial next step would be to collaborate with NGO platforms such as VolunteerMatch, Idealist, or GiveIndia to obtain real-world user data for training and validation. Second, the current model operates under a static setting where user preferences and NGO descriptions are considered unchanging. In reality, volunteers' interests may evolve, and the needs of NGOs may shift in response to emerging social issues. Incorporating temporal dynamics and building an online learning framework that continuously adapts to feedback would make the system more robust and responsive. Integrating explicit feedback mechanisms (e.g., ratings, sign-ups, and engagement metrics) and implicit behaviour tracking e.g., time spent browsing profiles, click-through rates) Support the continuous improvement of the recommendation quality.

Third, while BERT-based embeddings enhanced semantic matching, the system currently treats all BERT-derived features equally. Future iterations can benefit from applying attention mechanisms or domain-adaptive fine-tuning of BERT models to prioritise relevant parts of NGO descriptions or volunteer preferences. Additionally, the use of larger, more sophisticated language models, such as GPT-based embeddings, could potentially yield even more nuanced matching results, albeit at a higher computational cost. Fourth, expanding the platform into a fully interactive web or mobile application can translate this academic work into real-world utility. Such a platform would not only serve as a bridge between NGOs and volunteers but also provide a valuable resource for them. Still, it could also incorporate community-building features, such as messaging, testimonials, and activity tracking, to enhance user experience. Lastly, extending the methodology to accommodate multilingual contexts would increase inclusivity. In countries with diverse linguistic backgrounds, such as India, support for multiple languages could make the recommendation system more accessible to both local NGOs and volunteers who are not proficient in English. In conclusion, the proposed hybrid content-based recommendation system demonstrates strong potential for enhancing volunteer engagement through personalised, context-aware, and semantically rich matchmaking. As the world continues to witness a surge in youth-driven social activism and decentralised volunteerism, technology-driven tools like the one presented in this paper can significantly contribute to social capital formation and community development. The insights and architecture developed here can also serve as a blueprint for similar recommender systems in adjacent domains such as job matching, mentorship programs, and civic engagement platforms.

Acknowledgement: The authors express their sincere gratitude to SRM Institute of Science and Technology for providing the facilities and support necessary to carry out this research.

Data Availability Statement: Study data can be made available by the author upon reasonable request.

Funding Statement: No financial aid or sponsorship was involved in the completion of this research.

Conflicts of Interest Statement: The author states that there are no known conflicts of interest related to this study.

Ethics and Consent Statement: The study followed ethical research practices, with participants' consent obtained and their information kept confidential.

References

1. P. Sawant, R. Kunekar, A. Taywade, R. Tiwari, M. Sharma, M. Shinde, and A. Roy, "Content-based filtering using K-nearest neighbor," *in* Proc. *5th Biennial Int. Conf. Nascent Technologies in Engineering (ICNTE),* Navi Mumbai, India, 2023.
2. V. Thannimalai and L. Zhang, "A Content Based and Collaborative Filtering Recommender System," *in Proc. 2021 Int. Conf. Machine Learning and Cybernetics (ICMLC),* Adelaide, Australia, 2021.
3. Y. Afoudi, M. Lazaar, and M. Al Achhab, "Impact of Feature Selection on Content-Based Recommendation System," *in Proc. 2019 Int. Conf. Wireless Technologies, Embedded and Intelligent Systems (WITS),* Fez, Morocco, 2019.

4. V. Kumar, K. Nayak, P. Banerjee, and M. K. Dehury, "An Empirical Analysis of Collaborative and Content Based Recommender System by Using AI," *in Proc. 2023 Int. Conf. Advances in Computing, Communication and Applied Informatics (ACCAI),* Chennai, India, 2023.
5. T. F. Iqbal and R. Gnanajeyaraman, "Hybrid Content Collaborative Filtering: Improving Collaborative Filtering and Neural Network-based Recommendation," *in Proc. 2023 Int. Conf. Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES),* Chennai, India, 2023.
6. A. Saxena, "A Systematic Survey of Recommendation Systems," *in Proc. 2023 5th Int. Conf. Advances in Computing, Communication Control, and Networking (ICAC3N),* Greater Noida, India, 2023.
7. J. Varghese and M. Mohaghegh, "Personality-Based Hybrid Machine Learning Model for Mentor–Mentee Matching Using Collaborative and Content Filtering Methods," in Proc. *2023 7th Int. Symp. Multidisciplinary Studies and Innovative Technologies (ISMSIT),* Ankara, Turkey, 2023.
8. P. K. Ismail, N. Musthafa, and C. G. Raji, "Hybrid Recommender System Using K-means Clustering," in Proc. *2022 8th Int. Conf. Advanced Computing and Communication Systems (ICACCS),* Coimbatore, India, 2022.
9. M. D. C. Rodríguez-Hernández, R. Del-Hoyo Alonso, S. Ilarri, R. M. Montañés-Salas, and S. Sabroso-Lasa, "An Experimental Evaluation of Content-Based Recommendation Systems: Can Linked Data and BERT Help?" in Proc. *2020 IEEE/ACS 17th Int. Conf. Computer Systems and Applications (AICCSA),* Antalya, Turkey, 2020.
10. M. R. Hadi, A. H. Shnawa, M. A. Jebur, M. M. Adnan, G. Mohammed, and W. Hameed, "Avoidance of Scalability Problem in Recommendation System Using Filtering Approach," in Proc. *2023 6th Int. Conf. Engineering Technology and its Applications (IICETA),* Al-Najaf, Iraq, 2023.
11. T. Badriyah, S. Azvy, W. Yuwono, and I. Syarif, "Recommendation System for Property Search Using Content-Based Filtering Method," in *2018 Int. Conf. Information and Communications Technology (ICOIACT),* Yogyakarta, Indonesia, 2018.
12. R. S. Mulyana, A. Id Hadiana, and E. Ramadhan, "Recommendation System of Product Sales Ideas for MSMEs Using Content-Based Filtering and Collaborative Filtering Methods," in *2023 Int. Conf. Computer Science, Information Technology and Engineering (ICCoSITE),* Jakarta, Indonesia, 2023.
13. M. Ragavan and R. Khilar, "Improved Precision Rate in a Hybrid-Based Framework for College Recommendation System Using Novel Content-Based Filtering over Keyword Map Algorithm," in *2022 3rd Int. Conf. Smart Electronics and Communication (ICOSEC),* Trichy, Tamil Nadu, India, 2022.
14. K. N. M. Hunna, F. Renaldi, and I. Santikarama, "Paper Recommendation for Research References in Data Mining Using Content-Based Filtering," *in 2022 Int. Conf. Science and Technology (ICOSTECH),* Batam City, Indonesia, 2022.
15. E. Vemberly, E. Ancilla, A. Anderies, and A. Chowanda, "Enhancing Movie Recommendations: A Hybrid Filtering Approach Combining Collaborative and Content-Based Filtering," *in 2023 Int. Conf. Artificial Intelligence, Robotics, Signal and Image Processing (AIRoSIP),* Yogyakarta, Indonesia, 2023.
16. V. K. Kesarwani and A. Pratap, "Model for Movie Recommendation System Using Collaborative Filtering and Content-Based Filtering Techniques," *in 2025 2nd Int. Conf. Computational Intelligence, Communication Technology and Networking (CICTN),* Ghaziabad, Uttar Pradesh, India, 2025.
17. M. R. Julianti, Y. Heryadi, B. Yulianto, and W. Budiharto, "Recommendation System Model for Personalized Learning in Higher Education Using Content-Based Filtering Method," *in 2022 Int. Conf. Information Management and Technology (ICIMTech),* Semarang, Indonesia, 2022.
18. E. J. Chia and M. K. Najafabadi, "Solving Cold Start Problem for Recommendation System Using Content-Based Filtering," *in 2022 Int. Conf. Computer Technologies (ICCTech),* Melaka, Malaysia, 2022.
19. A. H. A. Rashid, M. Mohamad, S. Masrom, and A. Selamat, "Student Career Recommendation System Using Content-Based Filtering Method," *in 2022 3rd Int. Conf. Artificial Intelligence and Data Sciences (AiDAS),* Ipoh, Malaysia, 2022.
20. M. N. Uddin, J. Shrestha, and G. S. Jo, "Enhanced Content-Based Filtering Using Diverse Collaborative Prediction for Movie Recommendation," *in 2009 First Asian Conf. Intelligent Information and Database Systems*, Dong Hoi, Vietnam, 2009.
21. A. N. K. Albayati and Y. Ortakci, "Recommendation Systems on Twitter Data for Marketing Purposes Using Content-Based Filtering," *in 2022 Int. Congr. Human-Computer Interaction, Optimization and Robotic Applications (HORA),* Ankara, Turkey, 2022.
22. Y. Kamukwamba and L. Chunxiao, "A Novel Algorithm Using Content-Based Filtering Technology in Apache Spark for Big Data Analysis," *in 2021 4th Int. Conf. Artificial Intelligence and Big Data (ICAIBD),* Chengdu, China, 2021.
23. O. O. Aribisala, A. O. Opoola, and R. A. Adekola, "Utilization of entrepreneurship instruction on the creation of small-scale enterprise among business education students," *AVE Trends in Intelligent Technoprise Letters*, vol. 1, no. 4, pp. 196–202, 2024.