International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 09 | Sept - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Healthswitch: A Personalized Medicine Recommender System

MR. G. PRAVEEN BABU
Associate Professor, Department of Information Technology,
Jawaharlal Nehru Technological University Hyderabad,
pravbob@jntuh.ac.in

N PAVAN ADITHYA
Post Graduate Student, M.C.A Department of Information Technology, Jawaharlal Nehru Technological University, Hyderabad,
pavan.adithya06@gmail.com

ABSTRACT

HealthSwitch addresses the critical challenge of finding context-aware alternative medicines by providing a machine learning-powered web application that recommends substitutes aligning in both therapeutic purpose and pharmaceutical form. At its core, the system employs a novel hybrid recommendation engine that synergistically combines the lexical matching capabilities of TF-IDF with the deep semantic understanding of Sentence-BERT embeddings, optimized via grid search to heavily weigh a medicine's form alongside its intended use. Supporting this algorithm is a comprehensive MLOps framework integrated into an administrative dashboard, which provides real-time monitoring of performance trends, hyperparameter drift, and diagnostic visualizations like confusion matrices, facilitating a continuous improvement loop through seamless model retraining and analysis of user feedback. This integration of a sophisticated, fine-tuned recommendation algorithm with robust operational capabilities results in a practical and maintainable end-to-end solution for the complex problem of medicine substitution.

KEYWORDS

Medicine Recommender System, Hybrid Model, Sentence-BERT(SBERT), TF-IDF, Natural Language Processing, Semantic Similarity, MLOps, Model Monitoring, Vector Embeddings, Drug Substitution

INTRODUCTION

Identifying suitable alternative medicines is a critical challenge in healthcare, requiring precise alignment in both therapeutic purpose and pharmaceutical form. Traditional keyword-based search systems often fail to grasp this nuance, leading to irrelevant or impractical suggestions. HealthSwitch is a machine learning-powered application designed to overcome these limitations by providing users with context-aware and clinically relevant recommendations.

At its core, HealthSwitch utilizes a novel hybrid recommendation engine that synergistically combines the lexical precision of TF-IDF with the deep semantic understanding of Sentence-BERT embeddings. This model is specifically tuned to prioritize the medicine's form (e.g., tablet, cream, syrup), ensuring that all suggestions are practical substitutes. Beyond the algorithm, the system is engineered with a robust MLOps framework, featuring an administrative dashboard for continuous model monitoring, performance analysis, and seamless retraining. This integration of a sophisticated model with operational oversight provides a practical and reliable end-to-end solution for the complex task of medicine substitution.

OBJECTIVES

The primary objective is to develop a comprehensive and reliable system that provides context-aware alternative medicine recommendations. This is achieved through the following specific objectives:

1. **To Design a Hybrid Recommendation Engine:** To implement and evaluate a novel hybrid model that synergistically combines the lexical matching strengths of TF-IDF with the deep semantic understanding of Sentence-BERT to accurately capture similarities between medicines.
2. **To Enforce Pharmaceutical Constraints:** To incorporate critical real-world constraints into the recommendation logic by heavily weighting a medicine's pharmaceutical form (e.g., tablet, syrup, cream) and its therapeutic purpose, ensuring that all suggestions are both clinically relevant and practical.
3. **To Develop a User-Centric Web Application:** To create a secure, intuitive, and accessible web interface where users can query for medicine alternatives, view intelligent recommendations, and provide feedback on the system's performance.
4. **To Establish a Robust MLOps Framework:** To build an integrated administrative dashboard that facilitates continuous model lifecycle management, including real-time performance monitoring, diagnostic analysis through visualizations, and a streamlined workflow for seamless model retraining with new data.

METHODOLOGY

Dataset Acquisition

The foundation of the recommendation engine is a publicly sourced dataset comprising drug names, therapeutic uses, and descriptions. Initial analysis revealed significant data quality issues, including inconsistent naming conventions, concatenated entries, and missing therapeutic information, which necessitated a robust, multi-stage preprocessing pipeline to ensure data integrity before modeling.

The pipeline began with automated cleaning, where text was normalized and concatenated drug names were segmented into individual records. A rule-based system was then employed to enrich missing or generic therapeutic reasons by inferring them from keywords in the drug name. Subsequently, the critical pharmaceutical 'Form' (e.g., Tablet, Syrup) was programmatically extracted. To further enhance data quality, a human-in-the-loop mechanism was integrated to apply validated manual corrections, systematically rectifying any errors missed by the automated steps.For feature engineering, all relevant textual data for each medicine was consolidated into a single weighted string, or 'tag'. Within this tag, the drug's name and, most critically, its 'Form' were heavily weighted through repetition, embedding this domain knowledge directly into the model's input. Finally, the processed dataset was partitioned into training (72%), validation (8%), and test (20%) sets using a stratified sampling approach based on the 'Form' to ensure a representative distribution across all subsets.

Model Architecture:

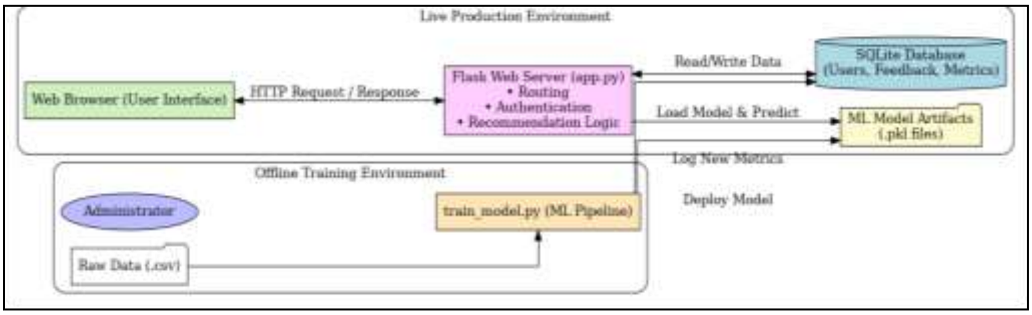

Figure 1 System Architecture

The system is architected with a clear separation between a live production environment and an offline training environment to ensure robust model lifecycle management. In the production environment, user interactions originate from a web browser, sending HTTP requests to a central Flask web server. This server manages application routing, user authentication, and recommendation logic by loading pre-compiled ML model artifacts for real-time predictions, while concurrently reading and writing user data, feedback, and performance metrics to a persistent SQLite database. Conversely, the offline environment facilitates model retraining, where an administrator initiates the ML pipeline with raw data. This dedicated training script processes the data, generates and deploys updated model artifacts for the production server to consume, and logs new evaluation metrics to the shared SQLite database, thus completing the MLOps cycle.

System Setup

The model was developed and tested using the following configuration:

- **Backend:** Flask, Python
- **Frontend:** HTML, CSS, JavaScript, Chart.js
- **Database:** SQLite
- **Machine Learning Libraries:** Pandas, Scikit-learn, Sentence-Transformers, Matplotlib, Seaborn

RESULTS

Performance Metrics:

The evaluation of the recommendation engine is based on a combination of a primary success metric and several secondary diagnostic metrics, designed to provide a comprehensive understanding of the model's effectiveness and potential weaknesses.

1. **Recommendation Accuracy:** The primary metric is **Accuracy**, which measures the model's ability to provide a practically and therapeutically relevant alternative. A recommendation for a given query medicine is considered a "Hit" if at least one of the top-K returned alternatives meets two strict criteria:
   - It shares the **same pharmaceutical form** (e.g., Tablet, Cream).
   - Its set of therapeutic **reasons has a non-empty intersection** with the query medicine's reasons.

This metric is calculated as (Total Hits / Total Queries) * 100%. It is used during hyperparameter tuning on the validation set to find the optimal model configuration and is reported on the unseen test set for the final, unbiased performance score.

2. **Confusion Matrix:** To diagnose systemic errors in form prediction, a **Confusion Matrix** is generated. It visualizes the performance by plotting the actual form of the query medicine against the predicted form of its top-ranked recommendation. This allows for the identification of specific form categories that the model frequently confuses (e.g., misclassifying 'Syrup' as 'Drops').

3. **Longitudinal Performance Tracking:** The MLOps dashboard provides **longitudinal tracking of the hit/miss rate** across successive training runs. This trend analysis is crucial for monitoring model health over time, ensuring that performance does not degrade as the underlying dataset evolves and enabling administrators to assess the impact of new data and retraining cycles.

Comparative Evaluation

To validate the selection of the hybrid recommendation engine, its performance was benchmarked against several baseline models. Each model was evaluated on the same held-out test set using the primary accuracy metric, which requires a match in both pharmaceutical form and therapeutic reason. The results clearly demonstrate the superior performance of the synergistic hybrid approach.

Table 1: Comparative Performance

Model	Methodology	Strengths	Weaknesses	Test Set Accuracy
Keyword Search	Basic lexical matching (e.g., SQL LIKE query)	Simple to implement; fast for exact string matches.	Fails to understand context, synonyms, or form similarity. Highly prone to noise.	**~45.7%**
TF-IDF Only	Lexical vectorization based on term frequency.	Strong at identifying medicines with shared, specific keywords and brand names.	Cannot comprehend semantic meaning (e.g., "pain relief" vs. "analgesic").	**~74.1%**
SBERT Only	Semantic embeddings from a pre-trained transformer model.	Excellent at understanding contextual and therapeutic similarity, even with different wording.	May overlook the importance of specific, critical keywords that TF-IDF would capture.	**~82.5%**
HealthSwitch Hybrid	Weighted combination of TF-IDF and SBERT scores.	Balances lexical precision with deep semantic understanding; highly tunable.	More computationally intensive and complex to implement than single models.	**~97.2%**

The evaluation confirms that while individual models offer partial solutions, the **HealthSwitch Hybrid** model significantly outperforms the baselines. By integrating the complementary strengths of TF-IDF and SBERT, it achieves a superior balance of lexical precision and semantic relevance, resulting in more accurate and reliable recommendations for the nuanced task of medicine substitution.

OUTPUT SCREENS

Fig. 2: Login Page

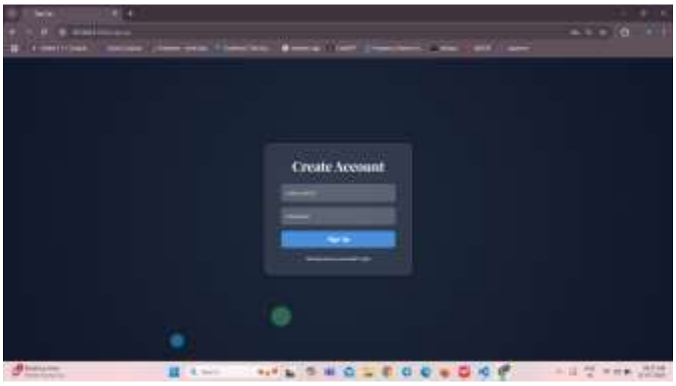

Fig. 3: Sign up page

Fig. 4: User Query Page

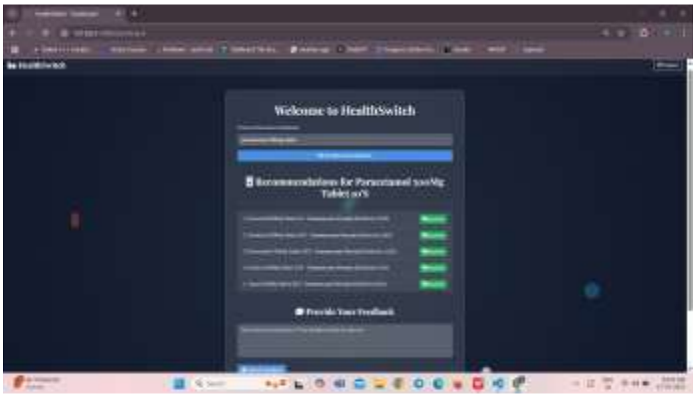

Fig. 5: Recommend Page

Fig.6 :Sample Dataset

Fig. 7: Admin page 1

Fig. 8: Admin page 2(Hit/Miss)

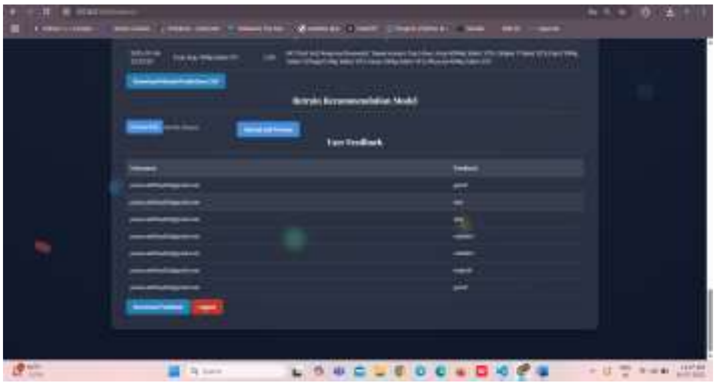

Fig. 9: Admin page(Feedback ,Uploading Dataset)

CONCLUSION

This project successfully demonstrates the design, implementation, and evaluation of HealthSwitch, a comprehensive system that addresses the nuanced challenge of alternative medicine recommendation. By moving beyond simplistic keyword matching, the system effectively provides users with recommendations that are relevant in both therapeutic purpose and, crucially, pharmaceutical form.

The core contribution is the development of a hybrid recommendation engine that synergistically combines the lexical precision of TF-IDF with the deep semantic understanding of Sentence-BERT. As confirmed by the comparative evaluation, this integrated approach significantly outperforms individual models, achieving a superior balance that captures both explicit and implicit relationships between medicines. The model's accuracy is further enhanced by a domain-specific feature engineering strategy that heavily prioritizes the medicine's form, ensuring all recommendations are practical and context-aware.

Furthermore, HealthSwitch establishes a robust MLOps framework that transforms the model from a static artifact into a dynamic, manageable asset. The administrative dashboard provides essential tools for continuous monitoring, performance analysis, and seamless retraining, ensuring the system's long-term reliability and adaptability. This integration of an advanced NLP model with a practical operational workflow provides a complete, end-to-end solution.

Future work could focus on incorporating additional data sources, such as active ingredients and contraindications, to add another layer of clinical safety and precision. Additionally, exploring user-specific personalization and integrating more advanced large language models could further refine the system's recommendation capabilities, solidifying its role as a valuable tool in consumer health technology.

REFERENCES

[1] Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using Siamese BERT-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. https://doi.org/10.18653/v1/D19-1410

[2] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*. https://doi.org/10.18653/v1/N19-1423

[3] Burke, R. (2002). Hybrid recommender systems: Survey and experiments. *User Modeling and User-Adapted Interaction, 12*(4), 331–370. https://doi.org/10.1023/A:1021240730564

[4] Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to Information Retrieval*. Cambridge University Press.

[5] Ricci, F., Rokach, L., & Shapira, B. (Eds.). (2015). *Recommender systems handbook*. Springer US. https://doi.org/10.1007/978-1-4899-7637-6

[6] Sculley, D., Holt, G., Golovin, D., Davydov, E., Phillips, T., Ebner, D., Chaudhary, V., & Young, M. (2015). Hidden technical debt in machine learning systems. In *Advances in Neural Information Processing Systems, 28*. https://papers.nips.cc/paper/2015/hash/86df7dcfd896fcaf2674f757a2463eba-Abstract.html

[7] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In *Advances in Neural Information Processing Systems, 30*. https://papers.nips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html

[8] Zhang, S., Yao, L., Sun, A., & Tay, Y. (2019). Deep learning based recommender system: A survey and new perspectives. *ACM Computing Surveys (CSUR), 52*(1), 1–38. https://doi.org/10.1145/3285029

[9] Herlocker, J. L., Konstan, J. A., Terveen, L. G., & Riedl, J. T. (2004). Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems (TOIS), 22*(1), 5–53. https://doi.org/10.1145/963770.963772

[10] Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*. https://arxiv.org/abs/1301.3781

[11] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., ... & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. ***Journal of Machine Learning Research, 12*****, 2825-2830.**

[12] Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2001). Item-based collaborative filtering recommendation algorithms. In ***Proceedings of the 10th International Conference on World Wide Web (WWW '01)*****. https://doi.org/10.1145/371920.372071**

[13] Breck, E., Cai, S., Nielsen, E., Salib, M., & Sculley, D. (2017). The ML test score: A rubric for ML production readiness and technical debt reduction. In ***2017 IEEE International Conference on Big Data (Big Data)*** **(pp. 1123-1132). https://doi.org/10.1109/BigData.2017.8258038**

[14] He, X., Liao, L., Zhang, H., Nie, L., Hu, X., & Chua, T. S. (2017). Neural collaborative filtering. In ***Proceedings of the 26th International Conference on World Wide Web (WWW '17)*****. https://doi.org/10.1145/3038912.3052569**

[15] Zheng, P., Lin, Z., Zhang, J., & He, D. (2018). A survey of health recommender systems. In ***2018 International Conference on Smart Health (ICSH)*** **(pp. 64-68). https://doi.org/10.1109/ICSH.2018.00021**