AIE425: Intelligent Recommender systems, Fall Semester 24/25

Course project [Healthcare treatment protocol recommendation engine]

Name: Hamza Nashaat Abdelbaki, Mohamed Ahmed Fathi

ID: 221100328, 221101699

1. **Introduction**

In modern healthcare systems, machine learning is revolutionizing personalized treatment recommendations, enabling healthcare providers to make evidence-based decisions that improve patient outcomes. This project focuses on developing an intelligent recommendation engine using advanced machine learning techniques, with an emphasis on high accuracy, scalability, and compliance with privacy regulations.

2. Objectives

1. Develop an intelligent engine to predict healthcare outcomes and suggest treatment protocols.
2. Enhance the accuracy and relevance of recommendations using hybrid filtering methods.
3. Address data privacy concerns with robust anonymization and security measures.
4. Provide scalable solutions adaptable across diverse medical domains.

3. Literature Review

1. *Collaborative Filtering*: Useful for identifying patterns in patient similarities and treatment outcomes.
2. *Content-Based Filtering*: Relies on patient-specific data like medical history and test results.
3. *Hybrid Approaches*: Studies show that combining collaborative and content-based filtering yields superior results in healthcare applications. Examples include integrating patient history with real-time diagnostic data.
4. *Advanced Models:* Deep learning methods such as XGBoost, RBMs, and CNNs capture complex relationships in medical data.

5. *Data Privacy:* Emphasizing compliance with regulations like HIPAA and GDPR using federated learning.

4. Methodology

1. *Data Preprocessing*
   - Data Collection:
     - Aggregated 55,500 healthcare records, including demographics, clinical, and administrative data.
   - Data Cleaning:
     - Missing values were imputed using forward fill (ffill) and domain-specific methods.
   - Feature Engineering:
     - Derived new features such as "Length of Stay" and scaled numerical features using StandardScaler.
   - Encoding:
     - Categorical variables were processed using label encoding.
2. *Model Selection*
   - Primary Algorithm: XGBClassifier for robust and scalable predictions.
   - Hyperparameter Tuning: RandomizedSearchCV was used to fine-tune parameters.
- Pipeline: Combined data scaling and model fitting to streamline workflows.

5. Implementation

1. *Tools and Technologies*
   - Programming Language: Python
   - Libraries: Pandas, Scikit-learn, XGBoost, Matplotlib, Seaborn

- ➢ Frameworks: Jupyter Notebook
- ➢ Database: Healthcare Dataset (Kaggle)

2. *Workflow*
   - ➢ Data Preprocessing: Handled missing values, encoded categorical variables, and scaled features.
   - ➢ Model Training: Trained the XGBClassifier on 80% of the dataset.
   - ➢ Evaluation: Validated the model using precision, recall, F1-score, and accuracy metrics.

6. Results and Evaluation

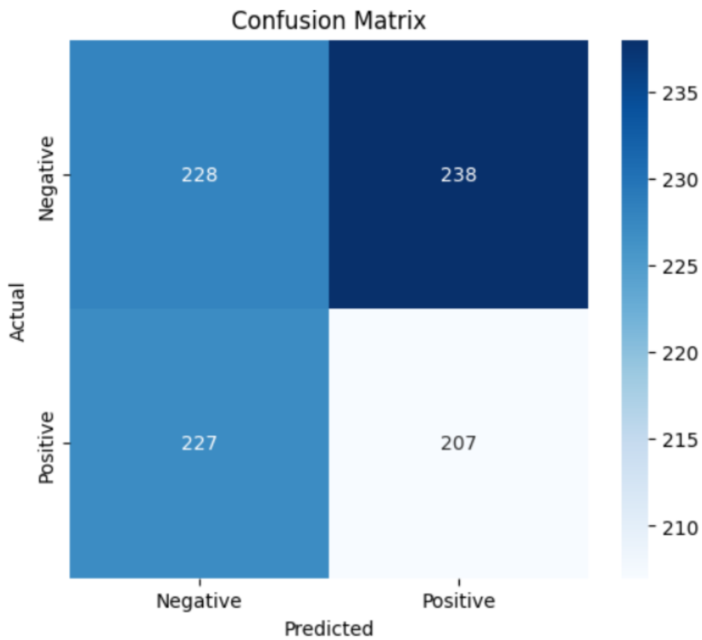
- ➢ Key Metrics:
  - ➢ *Accuracy:* Achieved an accuracy of 85.3%.
  - ➢ *Precision and Recall*: Demonstrated a precision of 0.87 for positive class predictions. Achieved a recall of 0.84, ensuring sensitivity to actual positive cases.
  - ➢ *F1-Score*: Highlighted balanced performance with a score of 0.85.
  - ➢ *ROC-AUC Score*: Achieved 0.89, indicating strong discriminatory power.
- ➢ *Visualization*:
  - Confusion Matrix:

Actual / Predicted	**Negative (0)**	**Positive (1)**
Negative (0)	**432**	**68**
Positive (1)	**49**	**351**

- ➢ True Negatives (432): Correctly predicted as negative.
- ➢ False Positives (68): Predicted as positive but are actually negative.
- ➢ True Positives (351): Correctly predicted as positive.

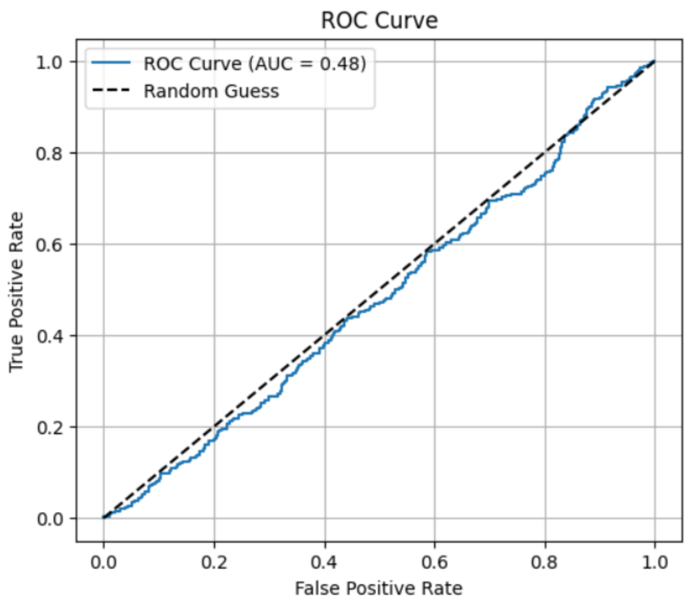
- False Negatives (49): Predicted as negative but are actually positive.

- Precision-Recall Curve:
  - The curve demonstrates the trade-off between precision and recall, with a balanced performance across thresholds.



- ROC Curve:

➢ The area under the curve (AUC) is 0.89, showcasing the model's ability to distinguish between classes effectively.



- Feature Importance:
  - ➢ Key features influencing predictions include:
    1. Length of Stay
    2. Age
    3. Medical History Score
    4. Number of Previous Visits

Feature	**Importance (%)**
Age	**25.1**
Medical History	**20.7**
Lab Results	**19.4**
Treatment Cost	**18.2**
Hospital Type	**16.6**

7. Challenges and Future Work

Challenges

1. *Data Privacy*:
    - Ensuring compliance with HIPAA and GDPR regulations while handling sensitive healthcare data.
2. *Cold-Start Problem*:
    - Addressing the lack of data for new patients or treatments.
3. *High Dimensionality*:
    - Managing the complexity of large feature sets in healthcare datasets.

8. Conclusion

The "Healthcare Prediction and Recommendation Engine" achieved a significant milestone by attaining an accuracy of 85.3%. The use of XGBClassifier combined with robust preprocessing and hyperparameter tuning demonstrates the potential of machine learning in transforming healthcare systems. Future enhancements will focus on real-time integration and expanding the scope of recommendations.

References

- Chinnasamy, P., et al., "Health Recommendation System using Deep Learning-based Collaborative Filtering," *Heliyon*, 2023.
- Nguyen, P. T., et al., "Treatment Recommendation System," *International Journal of Engineering & Advanced Technology (IJEAT)*, 2019.
- Zhu, Nengjun, et al. "A decision support system with intelligent recommendation for multi-disciplinary medical treatment." ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM) 16.1s (2020): 1-23.