Homework 3

Dt. 14th December 2023
BA878 Machine Learning and Data Infrastructure in Healthcare Fall 2023
Boston University Questrom School of Management
Prateek Naharia, nahariap@bu.edu

Problem 0. Real World Impact

A practical use case for your project on predicting hospital admissions could be its application in managing patient flow during seasonal flu outbreaks. Specifically, it could serve as a predictive tool in hospitals to forecast admission rates during peak flu seasons. This tool would enable hospitals to optimize staffing, bed allocation, and resource management in anticipation of increased admissions. By predicting potential surges in patient numbers, hospitals can ensure adequate care and reduce the strain on healthcare systems during these critical periods. This specific use case illustrates how your project could actively contribute to improving patient treatment and hospital preparedness.

Problem 1. Dealing With Data

a. If the model relies on real-time data for immediate decision-making, such as in emergency situations, delays in data availability could significantly impact its effectiveness. However, if the model is used for long-term planning or trend analysis, a delay might be less critical. To mitigate the impact of delays, the model could be designed to incorporate data as it becomes available, rather than relying on a complete dataset. Additionally, implementing techniques for dealing with incomplete or partially updated datasets could help maintain the model's accuracy and relevance. This might involve using statistical methods to estimate missing data points or developing algorithms that can adapt to data with varying degrees of freshness.
b. To handle outliers, which are data points significantly deviating from the distribution, use statistical methods like Z-scores or visual tools like scatter plots for detection. Assess whether these outliers are errors or valid extremes. Based on this, either remove, adjust, or include them carefully in the analysis to avoid skewing results.
c. Handling data drift, or changes in data distribution over time is not explicitly detailed in the project. However, in a real-world application, addressing data drift would involve regularly updating the model with new data, monitoring model performance over time, and retraining the model when significant changes in data distribution are detected. This process ensures that the model remains accurate and relevant to current trends and patterns in the data. Regular model evaluations and updates are essential for maintaining its effectiveness in a dynamic environment like a hospital emergency department.

Problem 2. Performance Evaluation

- Post-deployment, the performance of the project can be evaluated using metrics from the classification report and the confusion matrix.
- Real-time evaluation would monitor metrics like precision, recall, and f1-score to ensure the model adapts to new data.
- From A/B testing perspective - Evaluation post deployment in addition to the existing metrics, we could set up a control group where the traditional method of admission prediction is used and compare it against a test group using our model.
- We would measure outcomes like the accuracy of predictions, patient wait times, hospital resource utilization, and overall patient outcomes.
- Continuous monitoring and adjustments based on these metrics and user feedback will guide iterative improvements and maintain the model's relevance and accuracy.

Problem 3. Social Implications of Model Deployment

a. Stakeholders for this project include healthcare providers, patients, hospital administrators, and insurance companies. Providers and patients would benefit from more efficient and accurate care delivery. Administrators could see improved resource allocation and operational efficiency. Insurance companies might leverage predictive data for policy adjustments.
b. IRB involvement depends on the data's nature and use. If the project uses identifiable patient data or impacts patient care, IRB approval is typically required to ensure ethical considerations and patient privacy are upheld. If data is anonymized and used for operational improvements without direct patient impact, IRB approval might not be necessary.

Problem 4. Paper Submission

"**Journal of Medical Internet Research**" or "**Journal of the American Medical Informatics Association**," given their focus on healthcare and informatics, would be a good candidate for my paper.
The journals are receptive to healthcare technologies and machine learning.
Preparing the project for publication would involve a thorough literature review, ensuring the methodology meets academic standards, potentially conducting additional experiments for robustness, and framing the results within the broader context of current healthcare challenges and technological advancements.
https://amia.org/news-publications/journals/jamia
https://www.jmir.org/
