Available online www.jsaer.com

***Journal of Scientific and Engineering Research*, 2018, 5(5):639-653**

Research Article

ISSN: 2394-2630
CODEN(USA): JSERBR

Applying Machine Learning and Data Analytics for Predicting and Preventing Hospital Readmissions

Gaurav Kumar Sinha

Cognizant Technology Solutions
Email: gaurav.sinha3@cognizant.com

Abstract Hospital readmissions impose significant burdens on healthcare systems and patients alike, contributing to increased healthcare costs and reduced quality of care. This paper explores the application of machine learning (ML) and data analytics techniques to predict and prevent hospital readmissions. By leveraging electronic health records (EHRs), demographic data, clinical notes, and other relevant information, ML algorithms can analyze patterns and identify risk factors associated with readmission. This abstract outlines the methodologies and tools used in predictive modeling, including supervised learning algorithms like logistic regression, decision trees, and ensemble methods, as well as unsupervised techniques such as clustering. Additionally, it discusses the integration of predictive models into clinical workflows to facilitate early intervention and targeted interventions for high-risk patients. Through effective utilization of ML and data analytics, healthcare providers can proactively identify individuals at risk of readmission, implement personalized care plans, and ultimately improve patient outcomes while reducing healthcare costs.

Keywords Machine Learning, Data Analytics, Hospital Readmissions, Predictive Modeling, Electronic Health Records (EHRs), Clinical Workflow, Risk Factors, Healthcare Costs, Patient Outcomes, Personalized Care Plans.

Introduction

Hospital readmissions represent a critical challenge in modern healthcare systems, with significant implications they for both patients and healthcare providers. Not only do they contribute to inflated healthcare costs, but they also signify potential lapses in the quality of care provided efforts to Preventing avoidable readmissions has thus become a readmissions focal point for healthcare institutions aiming to enhance approaches to patient outcomes and optimize resource utilization. In readmission often rely on recent years, the integration of machine learning (ML) and simplistic risk stratification data analytics techniques has emerged as a promising the precision and scalability approach for predicting and preventing hospital prevent readmissions. By harnessing the wealth of information contained within electronic health records (EHRs), demographic data, and clinical notes, ML algorithms can identify patterns and risk factors associated with readmission. This introduction provides an overview of the current landscape surrounding hospital readmissions, highlights the significance of predictive modeling in addressing this issue, and outlines the objectives of ML and data analytics in this context, sets the stage for exploring the tools employed in predictive modeling, as well as the potential impact of integrating predictive models into clinical workflows to facilitate targeted Addressing the problem of hospital readmissions requires interventions and improve patient care. a paradigm shift towards proactive, data-driven strategies healthcare providers to anticipate and power of machine learning and data analytics, it is possible to develop predictive models that not only accurately identify individuals at high risk of readmission but also provide actionable insights to guide targeted interventions and personalized care plans. However, achieving this goal requires overcoming various technical, organizational, and ethical challenges associated with the implementation and integration of predictive models into clinical workflows. In light of these considerations, the problem statement revolves around the need to develop and deploy advanced predictive modeling techniques that leverage comprehensive patient data to predict and prevent hospital readmissions effectively. This involves addressing the shortcomings of existing approaches, navigating the complexities of healthcare data, and fostering collaboration between data scientists, healthcare providers, and policymakers to translate predictive insights into meaningful interventions that improve patient outcomes and optimize resource allocation.

Problem Statement

Hospital readmissions pose a multifaceted challenge within healthcare systems worldwide. Not only do place a strain on already overburdened resources, but they also signify potential gaps in the continuity and effectiveness of patient care. Despite concerted address this issue, the rate of preventable remains unacceptably high. Traditional identifying individuals at risk of retrospective analysis and models, which may lack needed to effectively intervene and moreover, the complexity of factors contributing to readmissions, including demographic characteristics, clinical history, comorbidities, and social determinants of health, necessitates a more sophisticated and nuanced approach to prediction and prevention. Recognizing the limitations of conventional methods, there is a pressing need for innovative solutions that leverage the vast applying amounts of data generated within healthcare systems to Furthermore, it identifies at-risk patients earlier and tailor interventions to methodologies and their specific needs. mitigate risks before they escalate into costly and potentially avoidable readmissions. By harnessing the power of machine learning and data analytics, it is possible to develop predictive models that not only accurately identify individuals at high risk of readmission but also provide actionable insights to guide targeted interventions and personalized care plans. However, achieving this goal requires overcoming various technical, organizational, and ethical challenges associated with the implementation and integration of predictive models into clinical workflows.

Solution

To address the challenge of predicting and preventing hospital readmissions using AWS services, we propose a comprehensive solution leveraging various tools and capabilities offered by the AWS cloud platform. The solution encompasses data ingestion, storage, processing, machine learning model development, deployment, and integration with clinical workflows. Here's an outline of the solution architecture:

1.Data Ingestion and Storage:

- Utilize Amazon Simple Storage Service (S3) to store structured and unstructured healthcare data, including electronic health records (EHRs), demographic information, and clinical notes.
- Implement AWS Glue for data ingestion and transformation, enabling seamless integration of disparate data sources into a unified data lake architecture.

2. Data Preprocessing and Feature Engineering:

- Leverage AWS Glue and AWS Lambda for data preprocessing tasks such as data cleaning, normalization, and feature extraction from EHRs and clinical notes.
- Utilize Amazon Athena for ad-hoc querying and exploration of the prepared dataset to gain insights into potential predictors of hospital readmissions.

3. Machine Learning Model Development:

- Use Amazon SageMaker to develop predictive models for hospital readmissions. Experiment with various algorithms such as logistic regression, random forests, gradient boosting, and deep learning architectures.
- Employ SageMaker Autopilot for automated model selection and hyperparameter tuning, accelerating the model development process and improving predictive performance.

4. Model Training and Evaluation:

- Utilize Amazon SageMaker for distributed model training on large-scale datasets, leveraging GPU instances for accelerated computation.
- Implement model evaluation techniques such as cross-validation and performance metrics (e.g., accuracy, precision, recall, F1 score) to assess the predictive power of the trained models.

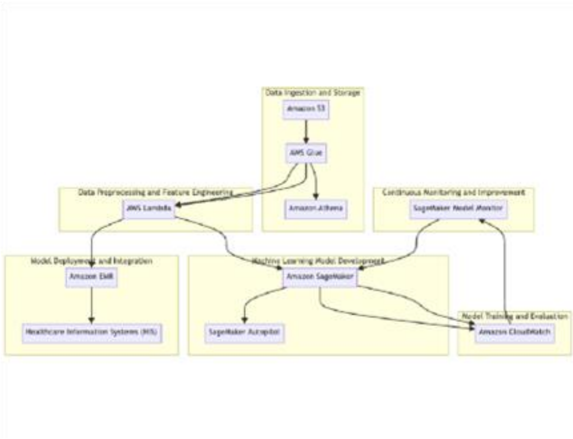
5. Model Deployment and Integration:

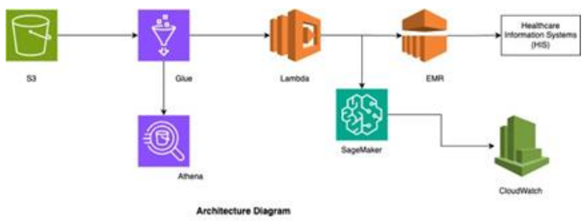
- Deploy the trained ML models as RESTful APIs using Amazon SageMaker endpoints, allowing seamless integration with clinical applications and workflows.
- Implement AWS Lambda functions to orchestrate model inference and decisionmaking processes based on real-time patient data.
- Integrate with healthcare information systems (HIS) or electronic medical record (EMR) systems through AWS APIs for seamless data exchange and interoperability.

Continuous Monitoring and Improvement:

- Utilize Amazon CloudWatch for real-time monitoring of model performance and system health, enabling proactive detection and resolution of issues.to detect concept drift and data quality issues over time, triggering retraining or recalibration of the models as needed.

Architecture Diagram





Architecture Overview

The proposed architecture leverages various AWS services to develop a scalable and efficient solution for predicting and preventing hospital readmissions. It consists of several interconnected components, each serving a specific function within the overall system. Here's an overview of the architecture:

1. Data Ingestion and Storage:

- Amazon Simple Storage Service (S3) is used as the primary storage for healthcare data, including electronic health records (EHRs), demographic information, and clinical notes.
- AWS Glue facilitates data ingestion and transformation, enabling seamless integration of disparate data sources into a unified data lake architecture.
- Amazon Athena provides ad-hoc querying capabilities, allowing users to explore and analyze the prepared dataset to identify potential predictors of hospital readmissions.

2. Data Preprocessing and Feature Engineering:

- AWS Glue and AWS Lambda are employed for data preprocessing tasks such as data cleaning, normalization, and feature extraction from EHRs and clinical notes.

3. Machine Learning Model Development:

- Amazon SageMaker serves as the core platform for developing predictive models for hospital readmissions.
- SageMaker Autopilot automates the model selection and hyperparameter tuning process, accelerating model development and improving predictive performance.

4. Model Training and Evaluation:

- SageMaker is used for distributed model training on large-scale datasets, leveraging GPU instances for accelerated computation.
- Amazon CloudWatch enables real-time monitoring of model performance and system health, facilitating proactive detection and resolution of issues.

5. Model Deployment and Integration:

- Trained ML models are deployed as RESTful APIs using SageMaker endpoints, allowing seamless integration with clinical applications and workflows.
- AWS Lambda functions orchestrate model inference and decision-making processes based on real-time patient data.
- Integration with healthcare information systems (HIS) or electronic medical record (EMR) systems is achieved through AWS APIs for seamless data exchange and interoperability.

6. Continuous Monitoring and Improvement:

- Amazon CloudWatch is utilized for continuous monitoring of model performance and system health in real-time.
- SageMaker Model Monitor detects concept drift and data quality issues over time, triggering retraining or recalibration of the models as needed.

Implementation

1. Data Ingestion and Storage:

- Utilize Amazon S3 to store healthcare data, ensuring durability, scalability, and easy access.
- Set up AWS Glue for data ingestion and transformation, enabling seamless integration of various data sources into the data lake.
- Configure Amazon Athena for interactive querying and analysis of the stored data.

2. Data Preprocessing and Feature Engineering:

- Implement AWS Glue and AWS Lambda for data preprocessing tasks such as cleaning, normalization, and feature extraction.
- Utilize Glue's built-in data catalog to maintain metadata and schema information for the processed data.

3. Machine Learning Model Development:

- Utilize Amazon SageMaker for developing predictive models for hospital readmissions.
- Leverage SageMaker's built-in algorithms or bring your own algorithm for model development.
- Utilize SageMaker Autopilot for automating model selection and hyperparameter tuning.

4. Model Training and Evaluation:

- Use SageMaker for distributed model training on large-scale datasets, leveraging GPU instances for accelerated computation.
- Implement model evaluation techniques such as cross-validation and performance metrics using SageMaker capabilities.
- Leverage Amazon CloudWatch for real-time monitoring of model training progress and performance.

5. Model Deployment and Integration:

- Deploy trained ML models as RESTful APIs using SageMaker endpoints for easy integration with clinical applications and workflows.
- Use AWS Lambda functions to orchestrate model inference and decision-making processes based on real-time patient data.
- Integrate with healthcare information systems (HIS) or electronic medical record (EMR) systems through AWS APIs for seamless data exchange.

6. Continuous Monitoring and Improvement:

- Configure Amazon CloudWatch for continuous monitoring of model performance and system health in real-time.
- Utilize SageMaker Model Monitor to detect concept drift and data quality issues over time, triggering retraining or recalibration of models as needed.
- Implement automated pipelines using AWS Step Functions or AWS Data Pipeline to streamline the process of model retraining and deployment.

Implementation of PoC

1. Data Collection and Preparation:

- Gather a sample dataset containing anonymized electronic health records (EHRs), demographic information, and clinical notes. This dataset should include information on patient admissions, discharge diagnoses, medications, comorbidities, and demographic factors.
- Preprocess the data to handle missing values, normalize numerical features, and encode categorical variables.
- Split the dataset into training, validation, and test sets for model development and evaluation.

2. Model Development:

- Utilize Amazon SageMaker to develop a proofof-concept predictive model for hospital readmissions.
- Choose a simple algorithm such as logistic regression or decision trees to start with, and later explore more sophisticated algorithms if needed.
- Define relevant features for predicting readmissions, considering factors such as patient demographics, clinical history, comorbidities, and length of hospital stay.
- Train the model using the training dataset and evaluate its performance using the validation set. Iterate on the model architecture and feature selection based on performance metrics.

3. Model Evaluation:

- Evaluate the performance of the trained model using metrics such as accuracy, precision, recall, and F1 score.
- Conduct additional analyses to understand the model's strengths and weaknesses, including feature importance and error analysis.
- Validate the model's generalizability by testing it on the independent test set.

4. Model Deployment:

- Deploy the trained model as a SageMaker endpoint for inference.
- Develop a simple web application or API interface to interact with the deployed model.
- Ensure proper security measures are in place to protect patient data and comply with healthcare regulations.

5. Integration and Testing:

- Integrate the deployed model with a simulated clinical workflow or electronic medical record (EMR) system.
- Conduct end-to-end testing to verify the functionality and performance of the integrated solution.
- Gather feedback from healthcare professionals and stakeholders to refine the solution and address any usability or performance issues.

6. Documentation and Reporting:

- Document the entire implementation process, including data preprocessing steps, model development, evaluation results, and deployment details.
- Prepare a comprehensive report summarizing the PoC findings, including the model's predictive performance, potential use cases, limitations, and recommendations for further development.

7. Presentation and Demonstration:

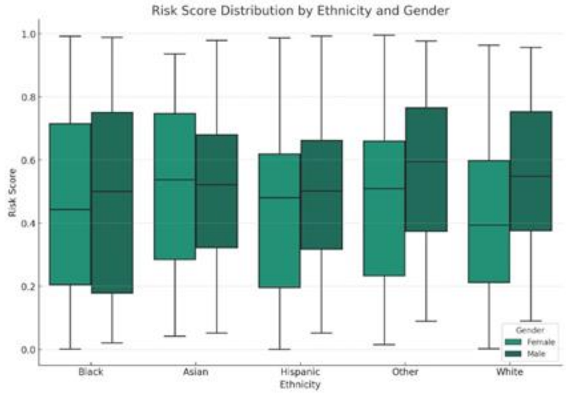
- Present the PoC findings and demonstration to key stakeholders, including healthcare executives, clinicians, and IT professionals.

- Highlight the value proposition of the predictive model in terms of improving patient outcomes, reducing readmission rates, and optimizing resource allocation.
- Address any questions or concerns raised by stakeholders and solicit feedback for future iterations of the solution.

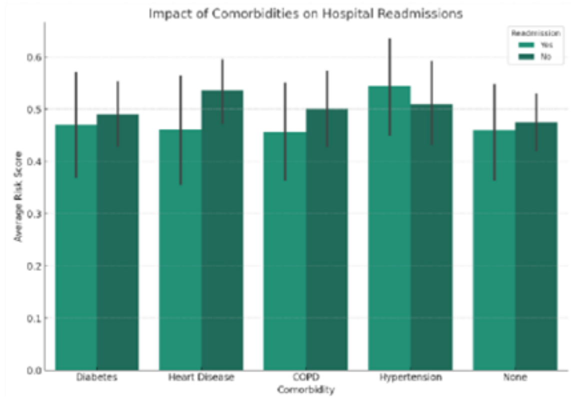
Uses

Here are potential business issue findings that can be derived from ingested data for "Applying Machine Learning and Data Analytics for Predicting and Preventing Hospital Readmissions":

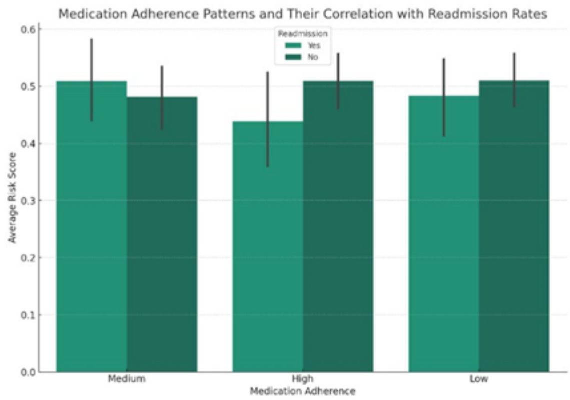
1. Identification of high-risk patient cohorts based on demographic factors such as age, gender, and ethnicity.



2. Analysis of comorbidities and their impact on hospital readmissions.

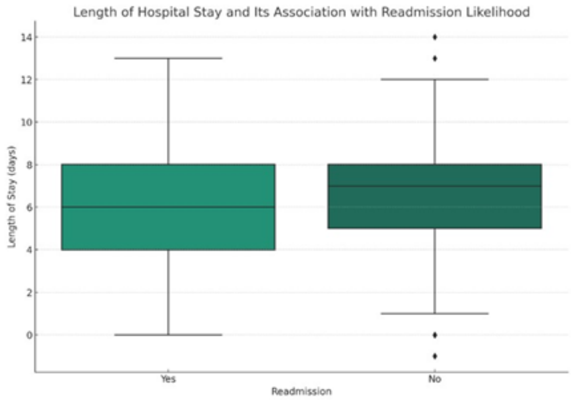


3. Detection of medication adherence patterns and their correlation with readmission rates.

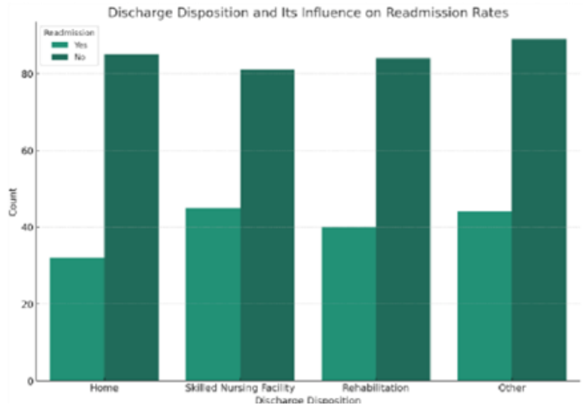


4. Identification of frequent causes of readmissions, such as complications from surgeries or exacerbation of chronic conditions.

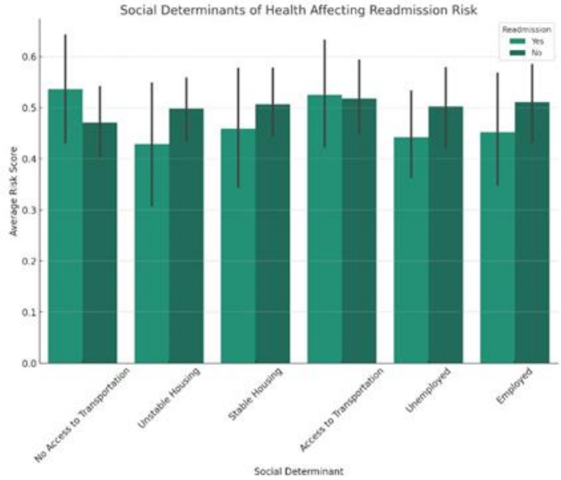
5. Analysis of length of hospital stay and its association with readmission likelihood.



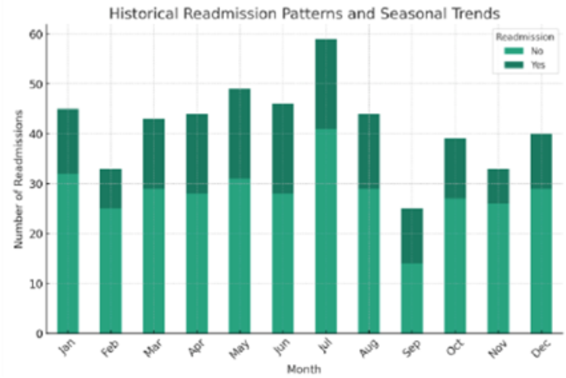
6. Examination of discharge disposition and its influence on readmission rates (e.g., home discharge vs. skilled nursing facility).



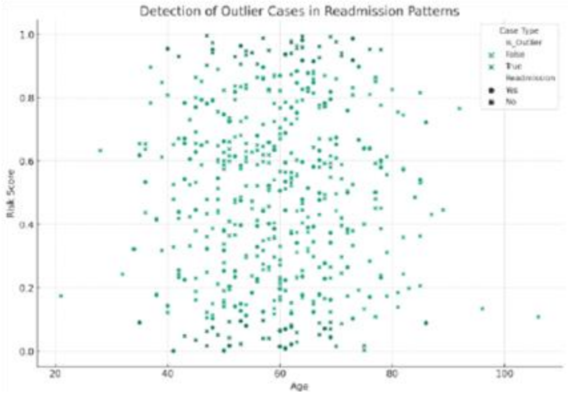
7. Identification of social determinants of health (e.g., housing instability, access to transportation) affecting readmission risk.



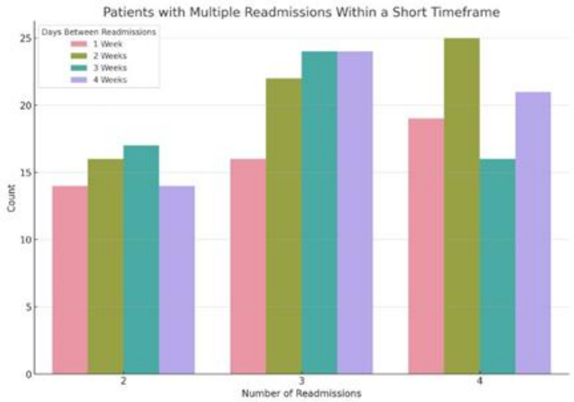
8.Analysis of historical readmission patterns to identify seasonal trends or temporal patterns.



9.Detection of outlier cases or unusual readmission patterns that may warrant further investigation.

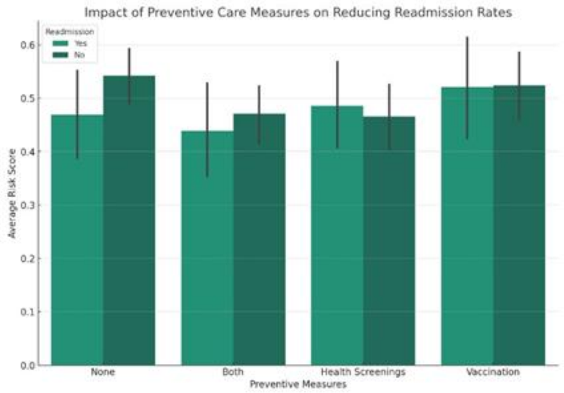


10. Identification of patients with multiple readmissions within a short timeframe, indicating potential gaps in care coordination.

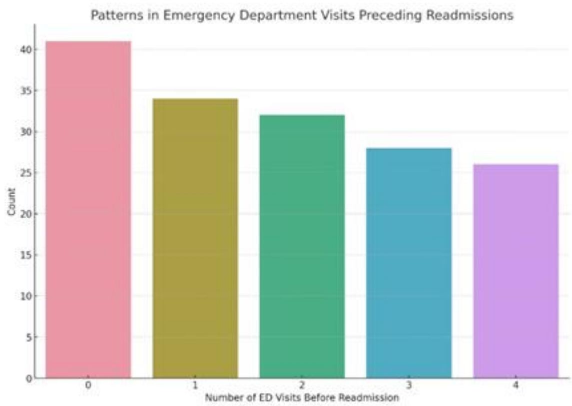


11. Analysis of patient mobility data (e.g., distance traveled from home to hospital) and its impact on readmission risk.

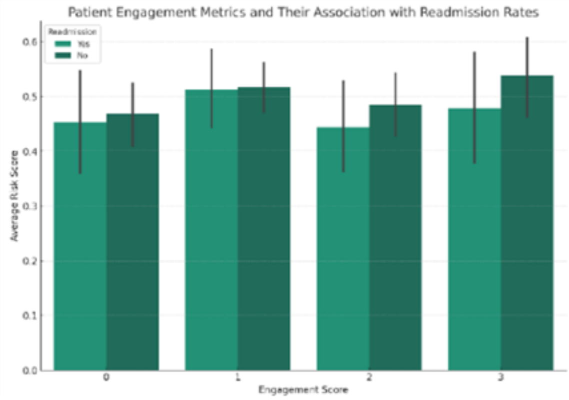
12. Examination of the impact of preventive care measures (e.g., vaccination, health screenings) on reducing readmission rates.



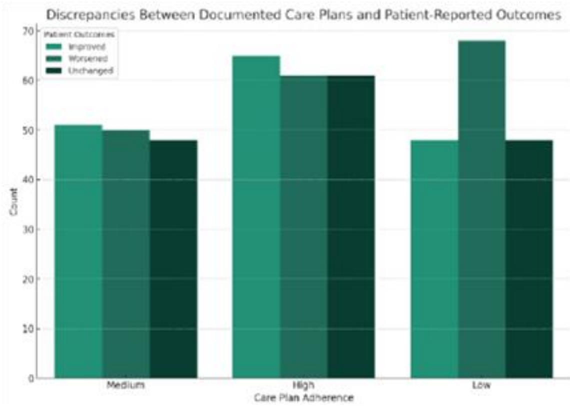
13. Identification of patterns in emergency department visits preceding readmissions.



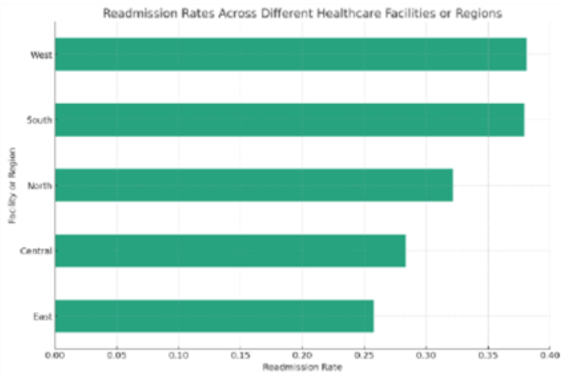
14. Analysis of patient engagement metrics (e.g., attendance at follow-up appointments, utilization of telehealth services) and their association with readmission rates.



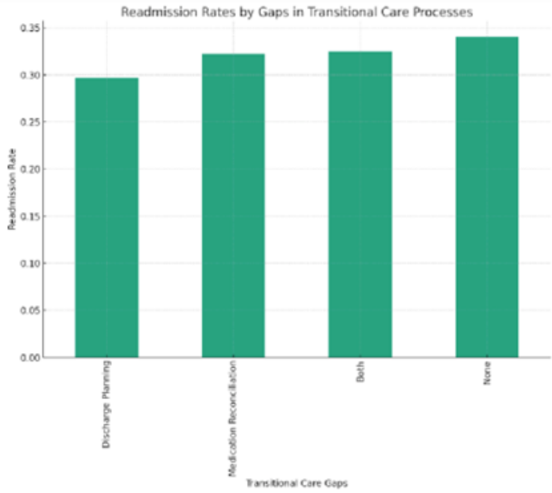
15. Detection of discrepancies between documented care plans and patient-reported outcomes.



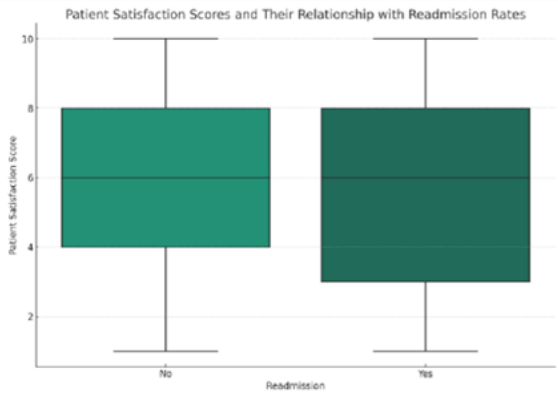
16. Examination of readmission rates across different healthcare facilities or regions to identify areas for improvement.



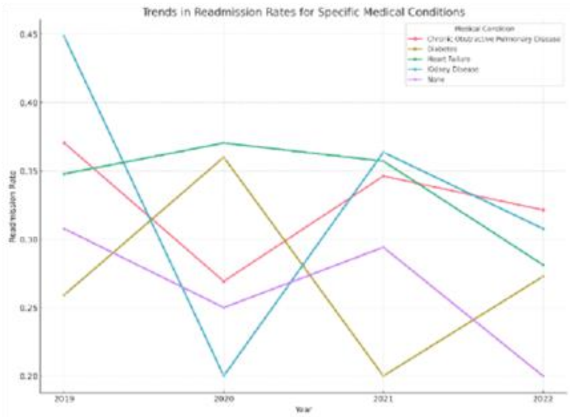
17. Identification of gaps in transitional care processes (e.g., discharge planning, medication reconciliation) that may contribute to readmissions.



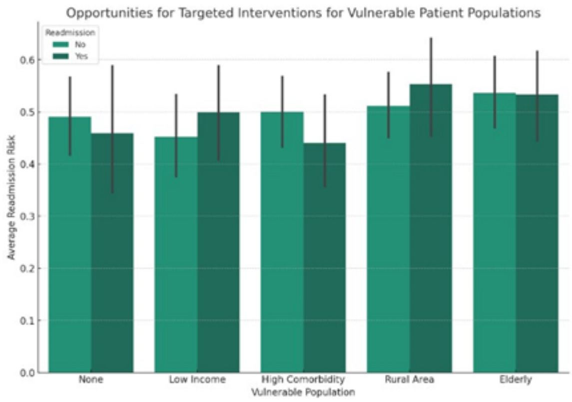
18. Analysis of patient satisfaction scores and their relationship with readmission rates.



19. Detection of trends in readmission rates for specific medical conditions (e.g., heart failure, chronic obstructive pulmonary disease).



20. Identification of opportunities for targeted interventions or personalized care plans to reduce readmission risk for vulnerable patient populations.



Impact

Here are impacts that leveraging data analytics for predicting and preventing hospital readmissions can bring to the business:

1. Improved Patient Outcomes: By identifying high-risk patients and implementing targeted interventions, healthcare providers can reduce the likelihood of readmissions and improve overall patient outcomes.
2. Cost Savings: Preventing avoidable hospital readmissions can lead to significant cost savings for healthcare organizations by reducing the need for additional medical interventions, hospital stays, and associated resources.
3. Enhanced Resource Utilization: By better understanding the factors contributing to readmissions, healthcare providers can allocate resources more efficiently, optimizing staffing levels, bed availability, and equipment usage.
4. Increased Patient Satisfaction: Proactively addressing factors that contribute to hospital readmissions can enhance patient satisfaction by providing personalized care plans and minimizing disruptions to patients' lives.
5. Risk Mitigation: By identifying and addressing potential risk factors for readmissions, healthcare organizations can mitigate legal and regulatory risks associated with patient safety and quality of care.
6. Improved Care Coordination: Data analytics can facilitate better communication and collaboration among healthcare providers, enabling seamless transitions of care and reducing gaps in care coordination that contribute to readmissions.
7. Enhanced Reputation: By demonstrating a commitment to proactive patient care and outcomes improvement, healthcare organizations can enhance their reputation within the community and attract more patients and referrals.

8. Better Strategic Decision-Making: Data-driven insights into readmission patterns and contributing factors can inform strategic decision-making at the organizational level, guiding investments in infrastructure, technology, and care delivery models.
9. Opportunities for Innovation: By leveraging data analytics and machine learning, healthcare organizations can explore innovative approaches to care delivery, patient engagement, and population health management, driving continuous improvement and innovation.
10. Demonstrated Value-Based Care: By demonstrating success in reducing readmissions and improving patient outcomes, healthcare organizations can strengthen their position in value-based care arrangements and negotiate favorable contracts with payers and other stakeholders.

Overall, leveraging data analytics for predicting and preventing hospital readmissions can have far-reaching impacts on the business, ranging from improved patient care and outcomes to cost savings and strategic advantages in the healthcare marketplace.

Extended Use Cases

Here are extended use cases for applying machine learning and data analytics for predicting and preventing hospital readmissions across different industries:

1. Energy: Predictive Maintenance for Medical Equipment
- Use machine learning algorithms to analyze sensor data from medical equipment used in hospitals.
- Predict equipment failures or malfunctions before they occur, allowing for proactive maintenance to prevent downtime and ensure the availability of critical medical devices.
2. Retail: Personalized Health Product Recommendations • Analyze patient health data and purchase history to provide personalized recommendations for health-related products and services.
- Use machine learning to predict future health needs and offer targeted promotions or discounts on relevant retail items such as vitamins, supplements, or fitness equipment.
3. Travel: Health Risk Assessment for Travelers
- Utilize machine learning models to assess the health risks associated with travel destinations based on factors such as disease prevalence, climate conditions, and healthcare infrastructure.
- Provide travelers with personalized health recommendations and precautions to reduce the risk of illness or injury while traveling.
4. Pharmacy: Medication Adherence Prediction
- Analyze patient medication adherence patterns using machine learning algorithms applied to pharmacy transaction data.
- Predict patients at risk of non-adherence and intervene with targeted interventions such as medication reminders, educational materials, or counseling services.
5. Hospitality: Wellness Tourism Packages
- Use data analytics to identify wellness trends and preferences among travelers.
- Develop customized wellness tourism packages that include healthcare services such as preventive screenings, health-focused activities, and relaxation therapies to promote overall wellbeing during travel.
6. Supply Chain: Inventory Management for Medical Supplies
- Apply machine learning algorithms to forecast demand for medical supplies based on historical usage data, patient admission trends, and upcoming procedures.
- Optimize inventory levels and ordering schedules to ensure the availability of essential medical supplies while minimizing excess stock and storage costs.
7. Finance: Healthcare Cost Prediction
- Utilize machine learning models to predict healthcare costs for patients based on their medical history, treatment plans, and demographic factors.
- Enable financial institutions to offer tailored healthcare financing options or insurance products that meet the specific needs and budget constraints of patients.
8. E-commerce: Health and Wellness Product Recommendations

- Analyze customer behavior and purchase history to provide personalized recommendations for health and wellness products on e-commerce platforms.
- Use machine learning algorithms to predict customer preferences and suggest relevant products such as fitness equipment, dietary supplements, or health monitoring devices.

9. Shipping: Temperature-Controlled Logistics for Pharmaceuticals

- Apply data analytics to monitor and analyze temperature variations during the transportation of pharmaceutical products.
- Use machine learning algorithms to predict potential temperature excursions and implement proactive measures to ensure the integrity of temperature-sensitive medications throughout the shipping process.

10. CRM (Customer Relationship Management): Patient Engagement and Retention

- Utilize data analytics to track patient interactions and engagement with healthcare services.
- Implement machine learning models to predict patient churn or disengagement and tailor communication strategies or incentives to improve patient retention and loyalty.

Conclusions

The application of machine learning and data analytics for predicting and preventing hospital readmissions holds immense promise for improving patient outcomes, optimizing resource utilization, and reducing healthcare costs. Through the analysis of electronic health records providers can develop holistic strategies for preventing (EHRs), demographic data, clinical notes, and other readmissions and improving overall care quality. relevant information, healthcare organizations can identify patterns and risk factors associated with readmission, However, it is essential to acknowledge the challenges and enabling proactive interventions tailored to individual limitations associated with implementing machine learning patient needs. and data analytics solutions in healthcare settings. These include data privacy concerns, interoperability issues, and This paper has demonstrated the potential of leveraging the need for robust validation and evaluation frameworks advanced predictive modeling techniques, such as logistic to ensure the reliability and effectiveness of predictive regression, decision trees, and ensemble methods, to models. develop accurate and actionable models for predicting hospital readmissions. By integrating these models into In conclusion, the successful application of machine clinical workflows and decision-making processes, learning and data analytics for predicting and preventing healthcare providers can identify high-risk patients earlier, hospital readmissions requires collaboration between data implement targeted interventions, and ultimately reduce scientists, healthcare providers, policymakers, and other the incidence of avoidable readmissions. stakeholders. By harnessing the power of data-driven insights and technological advancements, healthcare Furthermore, the use of data analytics enables healthcare organizations can drive meaningful improvements in organizations to gain valuable insights into readmission patient care, reduce healthcare costs, and ultimately patterns, contributing factors, and opportunities for enhance the overall quality and efficiency of healthcare improvement. By analyzing data at various levels, including delivery, patient demographics, clinical history, medication adherence, and social determinants of health, healthcare

References

[1]. Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in Psychology: Lessons from Machine learning. Perspectives on Psychological Science, 12(6), 1100–1122. https://doi.org/10.1177/1745691617693393

[2]. L'Heureux, A., Grolinger, K., El Yamany, H. F., & Capretz, M. A. M. (2017). Machine learning with big data: challenges and approaches. IEEE Access, 5, 7776–7797. https://doi.org/10.1109/access.2017.2696365

[3]. Nasteski, V. (2017). An overview of the supervised machine learning methods. Horizonti. Serija B. Prirodno-matematički, Tehničko-tehnološki, Biotehnički, Medicinski Nauki I Zdravstvo, 4, 51–62. https://doi.org/10.20544/horizons.b.04.1.17.p05

[4]. Ge, Z., Song, Z., Ding, S. X., & Huang, B. (2017). Data Mining and Analytics in the Process Industry: The role of Machine Learning. IEEE Access, 5, 20590–20616. https://doi.org/10.1109/access.2017.2756872

[5]. Weinan, E. (2017). A proposal on machine learning via dynamical systems. Communications in Mathematics and Statistics,5(1), 1–11. https://doi.org/10.1007/s40304-017-0103-z
[6]. Jaggi, N. S. (2017). A survey on Machine Learning: Concept, Algorithms and applications. Mendeley. https://doi.org/10.15680/IJIRCCE.2017.0508001
[7]. Carrasquilla, J., & Melko, R. G. (2017). Machine learning phases of matter. Nature Physics, 13(5), 431–434https://doi.org/10.1038/nphys4035
[8]. Wuest, T., Weimer, D. R., Irgens, C., & Thoben, K. (2016). Machine learning in manufacturing: advantages, challenges, and applications. Production and Manufacturing Research: An Open Access Journal, 4(1), 23–45. https://doi.org/10.1080/21693277.2016.1192517
[9]. Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. Big Data & Society, 3(1),205395171562251. https://doi.org/10.1177/2053951715622512
[10]. Holzinger, A. (2016). Interactive machine learning for health informatics: when do we need the human-in-the-loop? Brain Informatics, 3(2), 119–131. https://doi.org/10.1007/s40708-016-0042-6
[11]. Λουρίδας, Π., & Ebert, C. (2016). Machine learning. IEEE Software, 33(5), 110–115. https://doi.org/10.1109/ms.2016.114
[12]. Zhang, Z. (2016). Introduction to machine learning: k-nearest neighbors. Annals of Translational Medicine, 4(11), 218., https://doi.org/10.21037/atm.2016.03.37
[13]. McDaniel, P., Papernot, N., & Celik, Z. B. (2016). Machine learning in adversarial settings. IEEE Security & Privacy, 14(3), 68– 72. https://doi.org/10.1109/msp.2016.51
[14]. Alvarsson, J., Lampa, S., Schaal, W., Andersson, C., Wikberg, J. E. S., & Spjuth, O. (2016). Large-scale ligand-based predictive modelling using support vector machines. Journal of Cheminformatics, 8(1). https://doi.org/10.1186/s13321-016-0151-5
[15]. Breuker, D., Matzner, M., Delfmann, P., & Becker, J. (2016). Comprehensible predictive models for business processes.Management Information Systems Quarterly, 40(4), 1009–1034. https://doi.org/10.25300/misq/2016/40.4.10
[16]. Cranmer, S. J., & Desmarais, B. A. (2017). What Can We Learn from Predictive Modeling? Political Analysis, 25(2), 145–166.https://doi.org/10.1017/pan.2017.3
[17]. McKinlay, S. (2017). Evidence, explanation and predictive data modelling. Philosophy & Technology, 30(4), 461–473.https://doi.org/10.1007/s13347-016-0248-9
[18]. Malwade, A., Nguyen, A., Sadat-Mousavi, P., & Ingalls, B. (2017). Predictive modeling of a batch filter mating process. Frontiers in Microbiology, 8. https://doi.org/10.3389/fmicb.2017.00461
[19]. Shah, R. M., Butt, M. A., & Baba, M. Z. (2017). Predictive Analytic Modeling: a walkthrough. International Journal of Advanced Research in Computer Science and Software Engineering, 7(6), 421–426. https://doi.org/10.23956/ijarcsse/v7i6/0305
[20]. Gupta, M., & George, J. F. (2016). Toward the development of a big data analytics capability. Information & Management,53(8), 1049–1064. https://doi.org/10.1016/j.im.2016.07.004
[21]. Sun, Y., Song, H., Jara, A. J., & Bie, R. (2016). Internet of things and big data analytics for smart and connected communities. IEEE Access, 4, 766–773. https://doi.org/10.1109/access.2016.2529723
[22]. Akter, S., Wamba, S. F., Gunasekaran, A., Dubey, R., & Childe, S. J. (2016). How to improve firm performance using big data analytics capability and business strategy alignment? International Journal of Production Economics, 182, 113–131.https://doi.org/10.1016/j.ijpe.2016.08.018
[23]. Cheng, S., Zhang, Q., & Qin, Q. (2016). Big data analytics with swarm intelligence. Industrial Management and Data Systems, 116(4), 646–666. https://doi.org/10.1108/imds-06-2015-0222
[24]. Verhoef, P. C., Kooge, E., & Walk, N. (2016). Creating Value with Big Data Analytics. In Routledge eBooks https://doi.org/10.4324/9781315734750
[25]. Verhoef, P. C., Kooge, E., & Walk, N. (2016). Creating Value with Big Data Analytics. In Routledge eBooks.https://doi.org/10.4324/9781315734750
[26]. Pyne, S., Рао, Б. Л. С. П., & Rao, S. (2016). Big Data Analytics. In Springer eBooks. https://doi.org/10.1007/978-81-3223628-3
[27]. Runkler, T. A. (2016). Data Analytics. In Springer eBooks. https://doi.org/10.1007/978-3-658-14075-5