REPUBLIC OF TÜRKİYE
İZMİR BAKIRÇAY UNIVERSITY
FACULTY OF ECONOMICS AND ADMINISTRATIVE SCIENCES

Management Information Systems

MURAT EMRE YAPICI 181005019

GRADUATION THESIS

Comparing The Performance of Ensemble Methods in Predicting Emergency Department Admissions Using Machine Learning Techiques

Supervisor
Prof. Dr. Kadir Hiziroglu

June, 2023

Table of Contents

ABSTRACT 1

1. INTRODUCTION 2

2. BACKGROUND 5

 2.1. Machine Learning Methods 5

3. LITERATURE REVIEW 10

4. METHODOLOGY 14

 4.1. Data Aquisition 14

 4.2. Data Preparation 18

4.3. Modelling 20

4.4. Evaluation 20

4.4.1. Performance Metrics 20

4.4.2. Performance Comparison 21

5. RESULTS 22

5.1. Logistic Regression Results 22

5.2. Adaboost Results 25

5.3. Gentleboost Results 27

5.4. Logitboost Results 29

5.5. Rusboost Results 31

 Comparison 34

6. CONCLUSION 39

7. REFERENCES 40

ABSTRACT

The collection, storage, retrieval, and analysis of healthcare data are made possible by a variety of technologies, procedures, and tools used in information systems in health services [3]. Among these systems are health information exchanges (HIEs), telemedicine platforms, clinical decision support systems, electronic health records (EHRs), and various administrative and financial ones. They are made to improve patient outcomes, improve provider communication, and streamline healthcare workflows. Patients with acute illnesses, injuries, or conditions that pose a threat to their lives require immediate care, which emergency departments are essential for providing. Healthcare workers in the ED face tough challenges because of the rising patient volume, limited resources, and requirement for quick decisions. Innovative approaches to these problems are provided by machine learning, which also improves the effectiveness and efficiency of emergency care [9]. Two crucial applications of machine learning in emergency rooms are triage and risk stratification. Triage has traditionally relied on manual assessment and subjective evaluation, which has led to variations in resource allocation and priority setting [8]. Machine learning algorithms can be used to identify patients who require urgent care as well as to accurately predict the severity of a patient's condition. Numerous patient data, such as vital signs, symptoms, test results, and medical history, can be analyzed by these algorithms. In the emergency department setting, where quick action can significantly affect patient outcomes [10]. Electronic health records, imaging studies, and real-time monitoring data are just a few examples of the types of data sources that machine learning algorithms can analyze to find patterns and indicators that human clinicians might overlook. By spotting subtle patterns and predicting bad outcomes, machine learning models can enable early intervention, possibly preventing complications or even saving lives. Multiple methods are employed in the realm of analysis, encompassing a diverse range of types and styles as observed in existing literature. However, a peculiar absence emerges when it comes to the comparative evaluation of frequently employed ensemble methods. Hence, the primary objective of this study is to thoroughly examine and analyze various ensemble methods, aiming to elucidate their comparative efficacy and performance. By conducting a comprehensive assessment of these methods, we seek to shed light on their respective outcomes and discern potential patterns or variations among them. Through this rigorous investigation, we aspire to contribute to the existing body of knowledge and offer valuable insights for researchers and practitioners alike.

1. INTRODUCTION

Emergency service is a health unit established in hospitals and other health institutions to serve patients who require emergency medical assistance. Emergency services operate by dividing them into various units according to the patient's complaint and the urgency of the health condition. After the increase in the intensity of the emergency services since the 1960s, this unit started to serve by dividing into 3 groups. These groups are classified according to the symptoms of the patient and are classified as follows [1] :

Red: This group concerns priority emergency patients. The red area unit deals with life-threatening serious injury or serious illness. Patients included in the red area are in the class requiring the most emergency. Medical interventions for people with symptoms such as heart attack symptoms, serious respiratory problems, stab wounds and multiple trauma are carried out in the red area unit. In addition, patients brought to the emergency room by ambulance are also included in this group.

Yellow: An accident, trauma, etc. yellow area unit provides service for medical intervention in case of having diseases that risk permanent damage to his body as a result of this. Patients who are directed to the yellow area can wait for a while before the intervention.

Green: The green area unit serves in cases of mild injury and illness. The health status of the patients in the green area is less urgent than the patients in the yellow and red areas. For this reason, patients in this group may wait for a long time, especially in crowded hospitals.

In emergency services, the utilization of Information Systems holds significant importance. These systems provide valuable insights into the intensity and workload of the emergency department, offering general patient information, arrival times, and preliminary assessment of their conditions. This information enables healthcare providers to make informed decisions regarding triage and the allocation of resources in the emergency room. Recognizing the potential consequences of overcrowding in emergency departments, such as increased patient waiting times and the negative impact on staff morale and workforce, it is crucial to address this issue through innovative research and developments. Previous studies have highlighted emergency room overcrowding as a significant international challenge that necessitates proactive measures to mitigate its effects and improve overall emergency care [2].

Information systems in health services include a wide range of technologies, processes, and tools that enable the collection, storage, retrieval, and analysis of healthcare data [3]. These systems include electronic health records (EHRs), clinical decision support systems, telemedicine platforms, health information exchanges (HIEs), and various administrative and financial systems. They are designed to streamline healthcare workflows, enhance communication among healthcare providers, and improve patient outcomes.

The electronic health record (EHR) is a key element of information systems in healthcare. EHRs act as digital patient information databases, collecting medical histories, findings from diagnostic tests, prescription history, and other crucial data [4]. The adoption of EHRs has revolutionized patient information sharing and accessibility, allowing for a more thorough and coordinated approach to care. EHRs also assist clinical decision-making by giving healthcare professionals immediate access to vital patient information, warning them of potential drug interactions or allergies, and recommending evidence-based treatment options [7].

Another important component of information systems in health services is clinical decision support systems (CDSS) [5]. To offer recommendations that are supported by the best available research at the point of care, CDSS makes use of cutting-edge algorithms and medical knowledge databases. These systems can improve diagnostic precision, warn clinicians of potential mistakes, and assist in locating the best course of treatment. Clinical guidelines, best practices, and patient-specific data can all be integrated by CDSS to enhance patient safety, lower medical errors, and improve healthcare outcomes [6]. Machine learning can draw insightful conclusions from enormous amounts of medical data by utilizing advanced algorithms and data analysis methods. This improves patient care, optimizes resource allocation, and improves decision-making processes.

Machine learning, a subset of artificial intelligence, has emerged as a powerful tool with the potential to revolutionize healthcare, particularly in emergency departments (EDs) [8]. Emergency departments play a critical role in providing immediate care to patients with acute illnesses, injuries, or life-threatening conditions. However, the increasing volume of patients, resource constraints, and the need for rapid decision-making pose significant challenges for healthcare professionals in the ED. Machine learning offers innovative solutions to address these challenges and enhance the efficiency and effectiveness of emergency care [9]. Triage and risk stratification are two important uses of machine learning in emergency departments. Triage has historically relied on subjective evaluation and manual assessment, which has resulted in differences in resource allocation and prioritization [8]. The level of seriousness of a patient's condition can be accurately predicted using machine learning algorithms, which can also be used to identify patients who need urgent care. These algorithms can analyze a variety of patient data, including vital signs, symptoms, laboratory results, and medical history. Early diagnosis and prediction of adverse events are crucial in the ED setting, where timely interventions can significantly impact patient outcomes [10]. Machine learning algorithms can analyze diverse data sources, such as electronic health records, imaging studies, and real-time monitoring data, to identify patterns and indicators that may be missed by human clinicians. Machine learning models can enable early intervention by identifying subtle patterns and forecasting unfavorable outcomes, potentially averting complications or even saving lives.

The primary objective of this study is to analyze and calculate the admission rates in Emergency Departments (EDs) and develop a predictive model to determine the likelihood of future patients requiring hospitalization. By achieving this objective, the study aims to address several crucial goals that can significantly impact emergency healthcare services.

One of the key goals is to decrease the density and overcrowding in EDs. Emergency departments often face high patient volumes, resulting in extended wait times and delays in providing critical medical attention. By understanding the admission rates and factors influencing hospitalization, healthcare providers can implement strategies to manage patient flow more efficiently, thereby reducing congestion and enhancing the overall quality of care.

Another important aim is to expedite the treatment process for patients in need of urgent hospitalization. By predicting whether patients will require admission, medical staff can proactively allocate resources, arrange for necessary procedures, and coordinate with inpatient units, ensuring that prompt and appropriate care is delivered to those who need it most urgently. This proactive approach helps optimize patient outcomes and enhances patient satisfaction.

Additionally, this study seeks to increase the motivation and job satisfaction of healthcare employees working in the ED. High patient volumes and long wait times can lead to stress and burnout among staff. By streamlining the admission process and improving efficiency, employees can experience a more manageable workload, reduced stress levels, and increased job satisfaction. This, in turn, contributes to a positive work environment and ultimately enhances the quality of patient care.

To achieve these objectives, the study will collect and analyze relevant data on patient demographics, medical history, severity of conditions, and other factors influencing hospitalization decisions. Advanced statistical techniques, machine learning algorithms, and predictive modeling will be employed to develop a reliable framework for predicting hospitalization rates.

By successfully implementing the findings and recommendations derived from this study, Emergency Departments can significantly improve their operational efficiency, enhance patient outcomes, and create a more positive environment for both patients and healthcare professionals.

2. BACKGROUND

This section aims to provide comprehensive information about the various methods commonly employed in research related to emergency department admission prediction. Additionally, we will explore in detail the ensemble methods that have gained popularity in this field.

2.1. Machine Learning Methods

Machine learning models can be guided by either supervised learning or unsupervised learning, which are the two main methods. Because each method trains the algorithm to produce results in a different way, the decision between these methods depends on the data at hand and the particular question being answered [11]. During the training phase of supervised learning, the entire set of labeled data is used. The algorithm can learn patterns and make predictions based on the provided labels thanks to the explicit instructions provided by the labeled data. Unsupervised learning, on the other hand, deals with datasets without explicit guidance or labels. In this method, the algorithm investigates the data on its own to find any patterns, structures, or relationships that may be present. It essentially "wings it" by making insightful discoveries without any specific direction [12].

Because it handles simple tasks and is simple to implement, the supervised learning technique is more frequently used in machine learning. The output of the algorithm is labeled on the data inputs, assisting the machine in identifying patterns in the future, better differentiating data, or making predictions [13]. The two types of algorithms that make up supervised learning are best suited for problems with available reference points.

<u>Classification:</u> When the output variable falls under a certain category, a classification issue arises.

Regression: When the output variable is a real value that changes over time (such as money, weight, or measurement), there is a regression issue.

By using this method, the machine learning model learns naturally rather than being given a data set with clear instructions. Then, using analysis and interpretation, it tries to automatically detect structure in the unstructured data. Even though supervised learning is the simplest, we frequently lack access to complete, flawlessly labeled data sets for algorithm training. Unsupervised learning is useful in situations where analysts ask questions and there are multiple possible answers or where supervised learning has the "right" answer [11]. The unsupervised learning model is divided into four groups by algorithms, which organize data based on relationships or similarities between variables:

Clustering: The deep learning model groups together similar data and features after searching for them.

Association: An unsupervised learning model can forecast other attributes that they are frequently associated with by looking at key attributes in the data.

Detection of anomalies: In this case, the model is used to highlight data outliers. For instance, banks look for unusual customer purchase patterns to identify fraud. For instance, if a card is used in two very different locations in the same day, the bank will notice and look into the activity.

Artificial neural networks : Compresses input data into code, then attempts to reconstruct the input from that code while removing any signal noise to improve data quality [14].

Logistic Regression : Logistic regression is a statistical modeling technique widely used to predict the probability of a binary outcome based on one or more independent variables [41]. It has uses in a number of disciplines, including social sciences, economics, and medicine. The dependent variable in logistic regression has two possible values that are frequently referred to as "success" and "failure" or "0" and "1" [42]. The logistic function, also referred to as the sigmoid function, is at the core of logistic regression and converts any real-valued number into a range of [0, 1] [43]. The basis for modeling the relationship between the independent variables and the likelihood of the binary outcome is provided by this function. The algorithm estimates the coefficients for each independent variable, indicating the strength and direction of their impact on the outcome by fitting the logistic regression model to the training data [44].

2.1.1. Ensemble Methods

We use ensemble learning in daily life like to watch a movie, we see the review rating and that rating is based on collective decision. Ensemble method use this concept to handle machine learning to predict the most correct output compared to a single method [15].

Bagging

Bootstrap aggregating is commonly used in classification and regression, and is known as bagging. Through decision trees, it improves the models' accuracy, greatly reducing variance. Many predictive models struggle with overfitting, which is eliminated by reducing variance and improving accuracy [16]. In bagging, aggregation is used to include all potential outcomes of the prediction and randomize the result. Predictions made without aggregation won't be accurate because all possible outcomes won't be taken into account. As a result, the aggregation is based either on all of the results from the predictive models or on the probability bootstrapping procedures.

Bagging is advantageous because it creates a single strong learner that is more stable than individual weak base learners. Additionally, it gets rid of any variance, which lessens overfitting in models. The computational cost of bagging is one of its drawbacks [17]. Therefore, ignoring the correct bagging procedure can result in more bias in models.

Boosting

Boosting is an ensemble technique that improves future predictions by learning from previous predictor errors. The method greatly increases model predictability by combining several weak base learners into one strong learner [15] . Boosting works by placing weak learners in a sequential order so that they can learn from the subsequent learner to improve their predictive models.

<u>Adaboost :</u> A potent machine learning algorithm that belongs to the family of ensemble methods is called AdaBoost, which stands for Adaptive Boosting [45]. By combining weak classifiers into a robust and accurate classifier, it aims to improve the performance of weak classifiers. AdaBoost accomplishes this by repeatedly training weak classifiers, each of which gives more weight to the incorrectly classified samples from earlier iterations. AdaBoost's core idea is to give misclassified samples more weight so that weak classifiers that come after them will pay more attention to these difficult examples [45]. AdaBoost builds a strong classifier with high

accuracy on the training data through this iterative process of adjusting weights and combining weak classifiers. The final prediction is made by averaging the weighted predictions of the weak classifiers, which are weighted based on their performance during the prediction phase. AdaBoost has found wide application in various domains, including computer vision, natural language processing, and bioinformatics. It has demonstrated remarkable effectiveness in handling complex classification tasks and improving overall accuracy [46].

Gentleboost: A member of the boosting method family of machine learning algorithms is GentleBoost. By repeatedly training a series of weak classifiers and combining their predictions to produce a strong classifier, it is specifically made to handle binary classification tasks. The weak classifiers in GentleBoost are trained to reduce a weighted error function that gives samples that are incorrectly classified a higher weight. The margin-based penalty term in GentleBoost, in contrast to some other boosting algorithms, aims to widen the gap between the classes [47]. By reducing the effects of noisy or overlapping data, this margin-based penalty enhances generalization performance. GentleBoost gradually enhances its predictive abilities by training weak classifiers and iteratively updating the weights. The final classification choice is made by combining the predictions from the weak classifiers during the prediction phase while taking into account each one's unique strengths.

Several applications, such as object detection, face recognition, and medical diagnosis, have shown promise for GentleBoost. It is especially helpful in real-world scenarios because of its capacity to handle noisy and overlapping data [48].

Logitboost : Logitboost specifically designed for binary classification tasks, aiming to improve the performance of weak classifiers by combining them into a strong classifier. Weak classifiers are iteratively trained in LogitBoost to minimize a modified logistic loss function. The gradient of the logistic function, which is related to the likelihood of the sample belonging to a particular class, is taken into account by this modified loss function [47]. LogitBoost adjusts and concentrates on samples that are harder to classify, improving overall classification performance by iteratively updating the weights and training the weak classifiers. By combining the predictions of the weak classifiers, where each weak classifier contributes in accordance with its respective performance, LogitBoost's final prediction is obtained. By using an ensemble approach, LogitBoost can better capture complex relationships and perform binary classification tasks with greater accuracy [49].

Applications for LogitBoost can be found in many fields, including text categorization, bioinformatics, and anomaly detection. It is a useful tool in machine learning applications due to its capacity to manage complicated datasets and successfully combine weak classifiers [49].

RusBoost : The machine learning algorithm known as RusBoost, or random under-sampling boosting, combines the boosting and random under-sampling theories. It is intended specifically to address issues with class imbalance in binary classification tasks where the relative number of samples in one class is notably greater than that in the other. During each iteration of boosting in RusBoost, the algorithm randomly under-samples the majority class [50]. To equalize the class distribution, this process involves selecting a sample set from the majority class at random. By doing this, RusBoost emphasizes the minority class more during training, assisting the classifier in focusing on and learning about instances of the minority class. Weak classifiers are trained on the modified training set and given weights based on their performance during the boosting process. The contribution of each weak classifier to the final ensemble classifier is then calculated using the weights. RusBoost creates a strong classifier that is better able to handle unbalanced datasets by repeatedly combining the weak classifiers and changing the sample weights [51]. In many fields where unequal class distributions are prevalent, such as credit fraud detection, anomaly detection, and medical diagnosis, RusBoost has demonstrated effectiveness. It offers a workable solution to deal with class imbalance problems and enhance classification performance in these circumstances [51].

3. LITERATURE REVIEW

When examining the studies related to admission prediction in Emergency Services, the following findings have been obtained.

STUDY	TECHNIQUE	EVALUATION	RESULT	FINDINGS
Prediction across healthcare settings: a case study in predicting emergency department disposition [21]	**XGBoost**	**AUC**	**0.90 – 0.93**	**Barak-Corren et al. conducted a study comparing the performance of four modeling approaches for an ED trend prediction model. In the result of study, with the XGBoost Model showcasing the highest performance, achieving an accuracy range of 0.90-0.93.**
Prediction of emergency department patient disposition decision for proactive resource allocation for admission [22]	**Multinomial Logistic Regression Neural network Support vector machine**	**Accuracy (%95 CI) AUC**	**AUC = 0.97 LR = 81.2-82.0**	**Seung-Yup Lee at al. conducted a study to evaluate the performance of logistic regression and machine learning models to predict and determine for ED patients destined to different types of units. In result, All models predicted well with AUC, but LR Accuracy(%95CI) gave the best result with 81.2-82.0.**
Emergency department disposition prediction using a deep neural network with integrated clinical narratives and structured data [23]	F1 Score (Ref) DNN LR	Accuracy (%95 CI)	Accuracy REF= 0.674 (0.669-0.679) LR = 0.474 (0.469- 0.479) DNN = 0.60 (0.596-0.607	Chien-Hua Chen at al. conducted a study to develop the disposition prediction model using deep learning modeling strategy. In result, F1 score is taken as a reference and the Accuracy 0.674 . The closest result to DNN (0.60).

Predicting Hospital Admission and Returns to the Emergency Department for Elderly Patients [24]	Logistic Regression (LR)	Accuracy (%95 CI) P Value	Accuracy = 0.58-0.79 P Value = 0.00	**Michael A. at al. conducted a study to identify variables found among elderly ED patients that could predict either hospital admission or return to the ED.**
Predicting hospital admission at the emergency department triage: A novel prediction model [25]	Logistic Regression (LR)	AUC Accuracy (%95 CI)	<u>AUC</u> LR= 0.825 <u>Accuracy</u> LR = 0.824 - 0.827	Clare Allison Parker MD at al. conduct a study to create a model that can predict a patient's need for hospital admission at the time of triage.
Using Data Mining to Predict Hospital Admissions From the Emergency Department [26]	Logistic Regression (LR) Decision Trees Gradient Boosting Machines	Accuracy (%95 CI) AUC	<u>Accuracy</u> LR=79.94 Decision Trees=80.06 GBM = 80.31 <u>AUC</u> LR = 0.849 Decision Trees=0.824 GBM = 0.859	Byron Graham at al. conduct a study to explore the use of machine learning algorithms in predicting patient admissions from emergency departments (EDs) using routinely collected administrative data from two major acute hospitals in Northern Ireland. GBM gave the best result on both performance evaluation.
Prediction of Emergency Department Hospital Admission Based on Natural Language Processing and Neural Networks [27]	Logistic Regressin (LR) Multilayer Neural Network Models (MNNM)	Accrucay (%95 CI) AUC	Accuracy LR =0.83 -0.85 MNMM=0.83-0.85 AUC LR=0.846 MNMM=0.844	Xingyu Zhang at al. conducted a study to describe and compare logistic regression and neural network modeling strategies to predict hospital admission or transfer following initial presentation to Emergency Department (ED) triage with and without the addition of natural language processing elements. Both method performed well however MNMN performed slightly better than LR.
Generalizability of a Simple Approach for Predicting Hospital Admission From an Emergency Department [28]	Logistic Regression (LR)	AUC R2	<u>AUC</u> LR = 0.80- 0.89 <u>R2</u> LR=0.58 - 0.90	Jordan S. Peck PhD at al. conducted a study to evaluate the probabilties that emergency department (ED) patients will be admitted to a hospital inpatient unit.

Predicting Hospital Admissions at Emergency Department Triage Using Routine Administrative Data [29]	Logistic regression (LR)	ROC Accuracy (%95 CI)	ROC LR=0.849 Accuracy LR=0.847	Yan Sun PhD at al. conducted a study to be able to predict, at the time of triage, whether a need for hospital admission exists for emergency department (ED) patients.
Prediction of admission in pediatric emergency department with deep neural networks and triage textual data [30]	Deep Neural Network Gradient Boosting Classifier	AUC	AUC DNN & GBC =0.892	Bruno P. Roquette at al. conducted a study to propose and compare predictive models for predicting emergency department (ED) admission using both structured and unstructured data available at triage time. Both methods performed with same result.
Predicting hospital admission for older emergency department patients: Insights from machine learning [31]	Gradient boosted trees	AUC	AUC GBT =0.80	Fabrice Mowbray at al. conducted a study to predict ED admission in older adults and discuss their clinical and policy implications.
Predicting Emergency Department Inpatient Admissions to Improve Same-day Patient Flow [32]	Logit-linear regression The naïve Bayesian	AUC R2	AUC LLR= 0.0887 Bayesian = 0.0841 R2 LR=0.58 Bayesian = 0.58	Jordan S. Peck at al. conducted a study to evaluate three models that use information gathered during triage to predict, in real time, the number of emergency department (ED) patients who subsequently will be admitted to a hospital. In result LLR gave the best result with 0.0887 on AUC.
Predicting Hospital Admission for Emergency Department Patients using a Bayesian Network [33]	The naïve Bayesian	AUC Accuracy (%95 CI)	AUC Bayesian=0.920 Accuracy = 0.913	Jeffrey Leegon at al. conducted a study to evaluate the accuracy of a Bayesian network for the early prediction of hospital admission status using data from 16,900 ED encounters.
Predicting hospital admission at emergency department triage using machine learning [34]	Logistic regression (LR) Gradient boosting (XGBoost) Deep neural networks (DNN)	AUC	AUC LR=0.87 XGBoost =0.87 DNN = 0.87	Woo Suk Hong at al. conducted a study to predict hospital admission at the time of ED triage using patient history in addition to information collected at triage.

Table 1. represents the summary of studies related to Admission prediction of Emergency Departments.

When delving into the field of ensemble methods in machine learning, it becomes evident that they exhibit notably high success rates. However, it is crucial to acknowledge that there is a lack of comprehensive comparative studies between different ensemble methods within the existing literature. This knowledge gap prompted the need for the present study, which aims to compare the performance of three prominent ensemble methods: Adaboost, Gradient boosting, and XGBoost.

The primary objective of this study is to conduct an in-depth analysis and comparison of the aforementioned ensemble methods in a specific domain or problem area. By undertaking this task, the study seeks to uncover valuable insights into the relative strengths and weaknesses of each approach, facilitating informed decision-making and promoting advancements in ensemble learning techniques.

To achieve these objectives, the study will adopt a rigorous methodology, encompassing various stages. Firstly, an extensive review of the relevant literature will be conducted to gather a comprehensive understanding of the theoretical foundations, algorithmic intricacies, and real-world applications of Adaboost, Gradient boosting, and XGBoost. This literature review will serve as the groundwork for establishing the research context and identifying the research gaps.

Subsequently, a carefully curated dataset will be employed to evaluate and compare the performance of the three ensemble methods. The dataset selection process will consider factors such as data diversity, size, and representativeness of the problem domain to ensure the validity and generalizability of the findings.

Several evaluation metrics will be employed to comprehensively assess the performance of each ensemble method. These metrics may include accuracy, precision, recall, F1-score, area under the receiver operating characteristic curve (AUC-ROC), and computational efficiency. By utilizing multiple evaluation metrics, the study aims to provide a comprehensive assessment of the strengths and weaknesses of each ensemble method across different performance dimensions.

The study will also account for potential challenges and limitations in implementing the ensemble methods, such as model complexity, parameter tuning, and computational requirements. These considerations will contribute to a more realistic evaluation of the methods' practical feasibility and scalability.

In conclusion, this study strives to bridge the existing research gap by conducting a comparative analysis of Adaboost, Gradient boosting, and XGBoost ensemble methods. By shedding light on their relative performance, strengths, and weaknesses, the study aims to provide valuable insights that can guide researchers, practitioners, and decision-makers in selecting the most suitable ensemble method for specific problem domains and foster further advancements in ensemble learning techniques.

4. METHODOLOGY

4.1. Data Aquisition

The dataset used in this study consists of 1267 records of adult patients who were systematically selected from two emergency departments. The data was collected during a period spanning from October 2016 to September 2017 (Dataset Reference: [35]). These records provide valuable information about the characteristics, diagnoses, treatments, and outcomes of the patients who sought medical care at these emergency departments during the specified time frame. The dataset serves as a valuable resource for researchers and healthcare professionals to gain insights into the patterns and trends related to emergency department admissions and to explore various factors that contribute to patient outcomes in emergency care settings. By analyzing this comprehensive dataset, researchers can better understand the healthcare needs of adult patients in emergency situations and potentially improve the delivery of emergency medical services.

When the data set is examined;

Figure 1,2

	Group	Sex	Age	Patients number per hour	Arrival mode	Injury	Chief_complain	Mental	Pain	NRS_pain	SBP	DBP	HR	RR	BT	Saturation	KTAS_RN
1	2	2	71	3	3	2	right ocular pain	1	1	2	160	100	84	18	36.6	100	2
2	1	1	56	12	3	2	right forearm burn	1	1	2	137	75	60	20	36.5	NA	4
3	2	1	68	8	2	2	arm pain, Lt	1	1	2	130	80	102	20	36.6	98	4
4	1	2	71	8	1	1	ascites tapping	1	1	3	139	94	88	20	36.5	NA	4
5	1	2	58	4	3	1	distension, abd	1	1	3	91	67	93	18	36.5	NA	4
6	2	1	54	6	4	1	fever	1	1	3	140	90	94	20	38.1	98	3
7	2	2	49	11	3	1	With chest discomfort	1	1	3	110	70	70	20	36.2	98	2
8	1	2	78	14	3	1	pain, chest	1	1	3	169	86	80	20	36	NA	2
9	1	2	32	10	3	1	LBP - Low back pain	1	1	3	140	75	91	20	36.6	NA	4
10	2	1	38	6	3	1	Eczema, Eyelid	1	1	3	130	80	80	20	36.3	97	4

Diagnosis in ED	Disposition	KTAS_expert	Error_group	Length of stay_min	KTAS duration_min	mistriage
Corneal abrasion	1	4	2	86	5.00	1
Burn of hand, firts degree dorsum	1	5	4	64	3.95	1
Fracture of surgical neck of humerus, closed	2	5	4	862	1.00	1
Alcoholic liver cirrhosis with ascites	1	5	6	108	9.83	1
Ascites	1	5	8	109	6.60	1
Fever, unspecified	2	4	1	9246	2.00	1
Angina pectoris, unspecified	1	3	2	400	3.00	1
Acute coronary syndrome	1	3	2	247	10.23	1
Herniated disc disease of lumbar spine with radiculopathy	1	5	4	59	3.23	1
Ocular pain	1	5	4	185	4.00	1
Acute gastritis	1	4	4	176	7.53	1
Gout site unspecified	1	5	4	45	7.15	1

Figure 1.2 represents the dataset including its values

The data frame consists of 1264 rows and 29 columns.

Columns and data types are:

Column	Description
Group	1: Local ED / 2: Regional ED
Sex	1: Female / 2: Male
Age	Age (Years)
Patients number per hour	Patients number/hours
Arrival mode	1: Walking / 2: 119 use / 3: Private car / 4: Private ambulance / 5: Public transfotation (Police ets) /6: Wheelchair / 7: Others
Injury	1: Non-injury / 2: Injury
Mental	1: Alert / 2: Verval response / 3: Pain response / 4: Unconciousness
Pain	1: Pain / 2: Non-pain
NRS_pain	Numeric rating scales of pain
SBP	Systolid blood pressure
DBP	Diastolic blood pressure
HR	Heart rate
RR	Respiration rate
BT	Body temperature
Saturation	Saturation to use pulse oxmeter
KTAS_RN	KTAS result of nuses in ED
Disposition	1: Discharge / 2: Ward admission / 3: ICU admission / 4: AMA discharge / 5: Transfer / 6: Death / 7: OP fom ED
KTAS_expert	KTAS result of experts
Error_group	1: Vital sign / 2: Physical exam / 3: Psychatric /4: Pain / 5: Mental / 6: Underlying disease / 7: Medical records of other ED / 8: On set / 9: Others
Length of stay_min	Length of stay (minutes)
KTAS duration_min	KTAS duration (minutes)
Mistriage	0: Correct / 1: Over triage / 2: Under triage

Table 2 represents the dataset's columns and its descriptions

Furthermore, several columns were required to be eliminated, which would have a detrimental impact on the accuracy of the forecasting. Thus, it is crucial to highlight the current state of the dataset that is poised to undergo estimation, and the resulting configuration is presented below. Moreover, due to the non-binary nature of the Disposition value, it lacked suitable values for classification purposes. As a solution, the value was restructured into a binary format, specifically represented as "0" for discharge and "1" for admission. This simplification involved condensing the intricate admission stages, previously ranging from 1 to 7, into a binary distinction.

Table 3

Sex	*1: Female / 2: Male*
Age	Age (Years)
Patients_number_per_hour	Patients number/hours
Arrival_mode	1: Walking / 2: 119 use / 3: Private car / 4: Private ambulance / 5: Public transfotation (Police ets) /6: Wheelchair / 7: Others
Injury	1: Non-injury / 2: Injury
Mental	1: Alert / 2: Verval response / 3: Pain response / 4: Unconciousness
Pain	1: Pain / 2: Non-pain
NRS_pain	Numeric rating scales of pain
SBP	Systolid blood pressure
DBP	Diastolic blood pressure
HR	Heart rate
RR	Respiration rate
BT	Body temperature
Saturation	Saturation to use pulse oxmeter
Disposition	0: Discharge / 1: Admission

Table 3 represents the final state of the table

4.2. Data Preparation

An essential step in the data analysis process is data preparation, also referred to as data preprocessing. It entails converting unprocessed data into a format appropriate for analysis and modeling. Data cleaning, integration, transformation, and feature selection are a few of the tasks that make up data preparation. During the data preparation process, certain modifications were required for specific variables in the dataset. One such example is the transformation of disposition values. In order to facilitate analysis and modeling, the original disposition values were edited to represent binary categories, specifically assigning the values 0 and 1. By assigning 0 and 1 to the disposition variable, it becomes easier to interpret and analyze the dataset. This transformation enables researchers to differentiate between two distinct outcomes or decisions related to patient disposition, such as whether the patient was discharged (represented by 0) or admitted to the hospital (represented by 1). This binary representation allows for a clearer understanding of the disposition patterns within the dataset and facilitates subsequent analysis and modeling tasks.

By appropriately editing the variables, such as the disposition variable in this case, data preparation ensures that the dataset is in a suitable format for the intended analysis objectives. This modification facilitates more effective exploration of the data, enabling researchers to derive meaningful insights and make informed decisions based on the transformed variables.

4.2.1. Data Exploration

The patient population within the Emergency Department was systematically categorized and distributed based on a comprehensive set of variables, ensuring a thorough and multidimensional analysis. This categorization encompassed a wide range of factors, including but not limited to Group, which captures distinct patient cohorts; Sex, which provides insight into gender-specific patterns; Age, enabling the exploration of age-related trends and disparities; Arrival mode, shedding light on the modes through which patients accessed the department; Injury, facilitating the examination of different types and severities of injuries; Mental state, a crucial aspect for assessing psychological well-being; and Pain, a fundamental measure of patients' subjective discomfort levels. By incorporating these diverse dimensions into the distribution process, a more nuanced understanding of the patient population was attained, enabling subsequent analyses to delve deeper into the intricacies and complexities of the data.

4.2.2. Data Cleaning

In order to ensure the integrity and reliability of the analysis, careful measures were taken to address the presence of missing values (NA data) and outliers in the dataset. These problematic data points, known to have the potential to significantly distort the analysis results, were meticulously identified and subsequently removed from the dataset. By eliminating these outliers and addressing the missing values, the analysis can proceed with a more robust and accurate representation of the data, enhancing the validity of the subsequent findings and conclusions.

4.2.3. Data Transformation

One-hot Encoding

During the data preprocessing stage, a powerful technique called One-hot Encoding was employed to transform categorical variables into a format suitable for analysis. This technique involved representing each distinct value of the relevant column as a separate column, thereby capturing the unique characteristics of each category. By adopting this approach, the original column containing categorical data was subsequently removed to avoid redundancy and eliminate any potential data duplication. This strategic preprocessing step not only streamlined the dataset but also enabled a more comprehensive examination of the effectiveness with which different data points could be measured in subsequent analyses. Through this method, the underlying patterns and correlations within the data were brought to the forefront, allowing for a more nuanced understanding of the dataset's structure and facilitating more accurate and insightful analysis.

Sampling

To ensure the robustness and reliability of the applied methods and analyses, the dataset underwent a meticulous data processing phase where it was divided into two distinct samples: a training sample and a test sample. This separation allowed for an efficient evaluation of the accuracy and effectiveness of the employed methods. The training sample, comprising a significant portion of the dataset, was utilized to train the models and algorithms, enabling them to learn and identify underlying patterns and relationships within the data. On the other hand, the test sample served as an independent set, reserved exclusively for evaluating the performance and generalization ability of the trained models. By employing this approach, the effectiveness and efficiency of the methods and analyses were meticulously measured, ensuring their reliability and suitability for subsequent applications. This comprehensive sampling strategy fostered a more rigorous evaluation process, ultimately yielding valuable insights into the robustness and predictive capabilities of the employed techniques.

4.3. Modelling

Logistic Regression, a widely used classification algorithm, was employed to model the data. By fitting the logistic regression model to the training data, it learned the relationships between the independent variables and the binary outcome. The model's coefficients were interpreted to understand the impact of each variable on the outcome.

In addition to Logistic Regression, several boosting algorithms were also utilized: Adaboost, GentleBoost, LogitBoost, and RusBoost. These boosting algorithms iteratively combine weak learners to form a strong predictive model. Adaboost, for instance, assigns weights to data points and trains weak learners on reweighted samples, emphasizing misclassified instances. GentleBoost, LogitBoost, and RusBoost employ similar principles, but with their own variations in terms of reweighting or modifying the weak learners.

Each boosting algorithm was trained on the training sample, allowing them to learn the underlying patterns and improve their predictive performance. The trained models were then evaluated using the independent test sample to assess their ability to generalize to unseen data. Performance metrics such as accuracy, precision, recall, and F1 score were computed to compare the models and determine their effectiveness in predicting the outcome variable.

4.4. Evaluation

4.4.1. Performance Metrics

To thoroughly assess the performance and quality of the analyses, a diverse range of techniques were employed, each offering unique insights into different aspects of the results. One such technique is Confusion Matrix, which provides a measure of the linearity and precision of the obtained outcomes. This metric gauges the overall correctness of the predictions and serves as a fundamental benchmark for evaluating the effectiveness of the applied methods.

Furthermore, Sensitivity and Specificity are critical evaluation techniques that assess the model's ability to correctly identify positive and negative instances, respectively. Sensitivity measures the proportion of correctly detected positive cases, while Specificity quantifies the accuracy in identifying negative cases. These complementary metrics shed light on the model's performance across different outcome categories, enabling a more comprehensive understanding of its strengths and weaknesses.

Moreover, the Area Under the Curve (AUC) technique is widely used to evaluate the effectiveness of predictive models. It measures the model's ability to distinguish between different classes by quantifying the area under the Receiver Operating Characteristic (ROC) curve. A higher AUC value indicates a more accurate and reliable model for classification tasks.

By employing this diverse range of evaluation techniques, the analyses were comprehensively assessed, providing a multi-faceted understanding of their performance across various dimensions. These metrics collectively contributed to a rigorous evaluation process, facilitating informed decision-making and ensuring the reliability and effectiveness of the applied methods.

4.4.2. Performance Comparison

The performance evaluation section provides a comprehensive and detailed analysis of the results obtained from comparing various Ensemble methods. This meticulous examination offers valuable insights into the strengths, weaknesses, and nuances of each method, enabling a thorough understanding of their performance characteristics.

By carefully scrutinizing the results, it becomes possible to assess the effectiveness and efficiency of Ensemble methods in solving complex problems. These comparative analyses shed light on how different Ensemble techniques leverage the collective intelligence of multiple models to enhance prediction accuracy and generalization capabilities. Furthermore, they provide a deeper understanding of how the methods handle issues such as overfitting, bias-variance trade-off, and model diversity.

In this performance evaluation section, key metrics and statistical measures are employed to quantitatively assess the performance of Ensemble methods. These include but are not limited to Accuracy, F1 Score, Precision, Recall, and Cross-Validation techniques. By considering a range of evaluation metrics, a comprehensive and nuanced assessment of the Ensemble methods is achieved.

Moreover, the evaluation section goes beyond a mere comparison of numerical results. It delves into the qualitative aspects as well, exploring the robustness, stability, and interpretability of the Ensemble models. These qualitative assessments provide deeper insights into the underlying mechanisms and decision-making processes of the Ensemble methods, enriching the analysis with a holistic understanding of their performance.

5. RESULTS

5.1. Logistic Regression Results

c

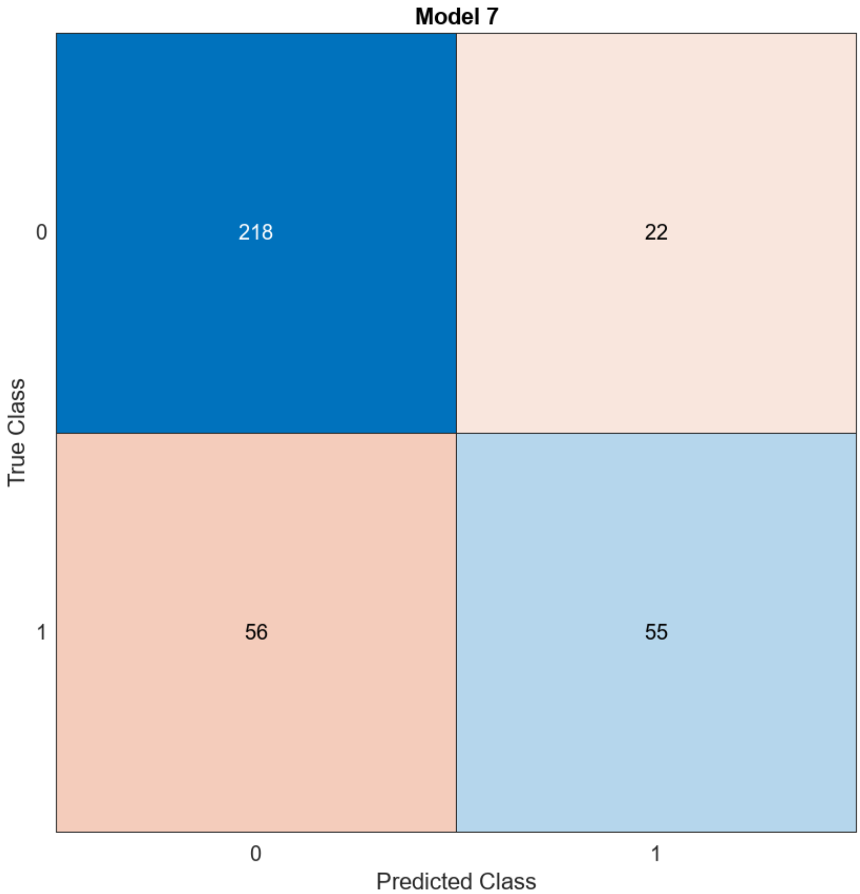

Figure 3. represents the Confision Matrix results of Logistic Regression

Upon careful examination of the accuracy results, it becomes evident that Logistic Regression demonstrates contrasting levels of success in estimating admission and discharge values.

When focusing on the estimation of admission values, Logistic Regression showcases a notable level of accuracy, providing satisfactory results. This suggests that the model effectively captures the patterns and underlying dynamics related to admissions, enabling reliable estimations in this regard.

Several studies in the healthcare domain have demonstrated the effectiveness of Logistic Regression in predicting admission outcomes. For instance, a study conducted by Grimmer et al. applied Logistic Regression to estimate hospital admissions based on various patient attributes, achieving satisfactory accuracy rates [36]. Similarly, Chen et al. utilized Logistic Regression to predict the likelihood of hospital admissions for patients with chronic diseases, reporting favorable estimation results [37].

However, in terms of admission estimation, the performance of Logistic Regression seems to be comparatively less successful. This observation aligns with the findings of studies such as the work by Hales et al [38], which analyzed admission prediction models and noted limitations in the accuracy of Logistic Regression for this specific task. The complexity of admission processes, including factors such as patient conditions, treatment effectiveness, and care coordination, can contribute to the challenges faced by Logistic Regression in accurately predicting admission outcomes.

To overcome the limitations in discharge estimation within the Logistic Regression framework, researchers have explored various strategies. For example, feature engineering techniques have been employed to enhance the representation of relevant variables associated with the discharge process [39]. Additionally, ensemble methods, such as Random Forest or Gradient Boosting, have been proposed to leverage the collective strengths of multiple models and improve the accuracy of discharge predictions [40].

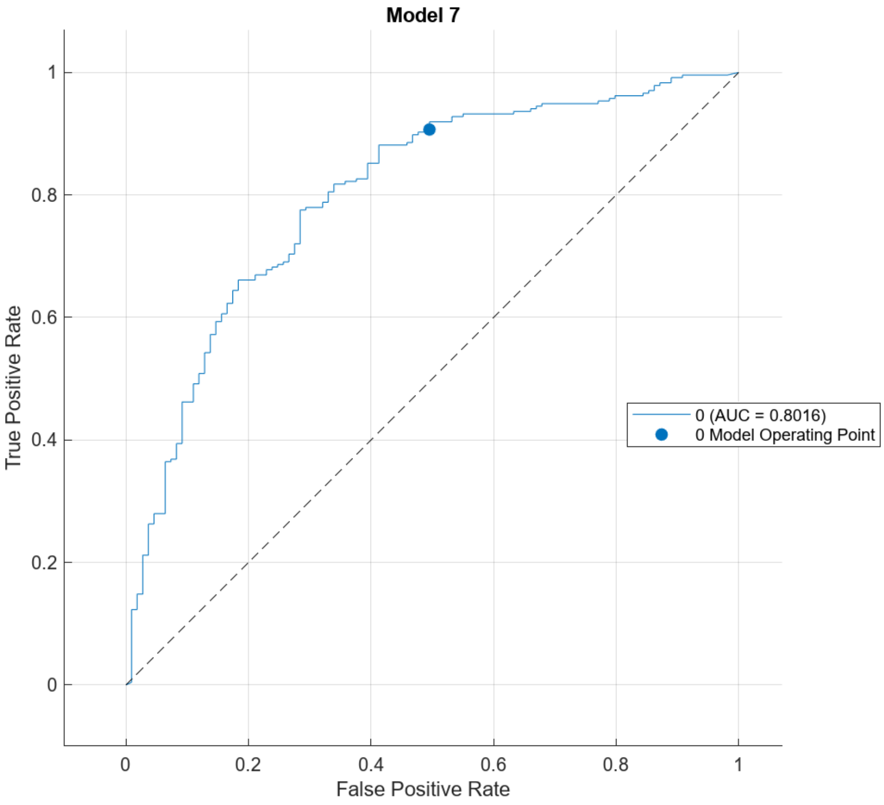

Figure 4

Figure 4 represents the ROC results of Logistic Regression

During investigation into the evaluation metrics of Logistic Regression, a notable outcome emerged when considering the Area Under the Curve (AUC). We achieved a result that can be characterized as nearly successful, with an impressive success rate of ' 0.8016 '. However, it is crucial to acknowledge that the success rates achieved in previous studies utilizing Logistic Regression have varied significantly. This discrepancy in outcomes highlights the substantial impact of the dataset employed in our analysis, which has notably influenced the observed success rate. It is evident from the existing literature that Logistic Regression has been employed with higher levels of success in certain studies, underscoring the intricate relationship between model performance and the specific dataset used for analysis. These findings emphasize the importance of carefully selecting and curating datasets to ensure reliable and robust results. By recognizing the potential impact of dataset characteristics, researchers can make informed decisions and advance the accuracy and applicability of Logistic Regression in their respective fields.

5.2. Adaboost Results

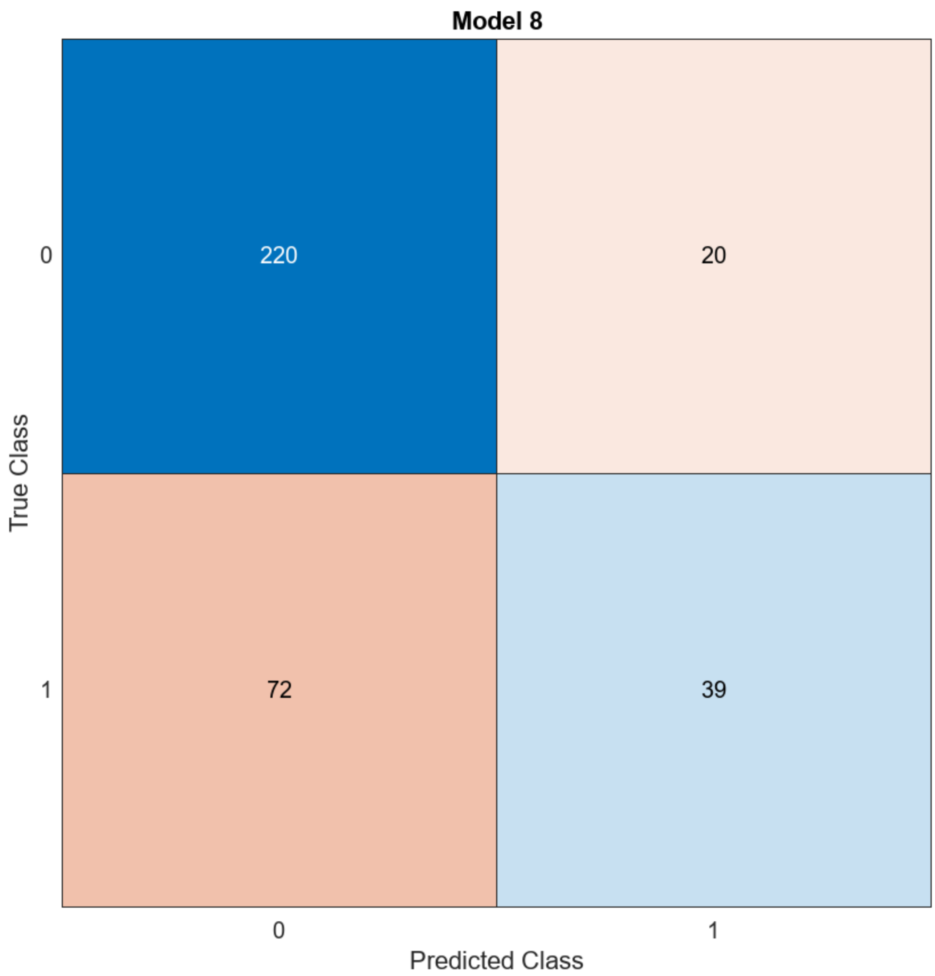

Figure 5

Figure 5 represent the Confision Matrix results of Adaboost

The realm of Adaboost analysis has yielded intriguing results, closely resembling those obtained through Logistic Regression. Upon thorough examination of the Accuracy metric, it becomes apparent that the estimation of discharge shows a slight edge over Logistic Regression. However, when it comes to predicting admissions, Adaboost proves to be somewhat less successful compared to Logistic Regression. These findings shed light on the nuanced nature of these two approaches, showcasing their varying levels of effectiveness across different aspects of the analysis. While Adaboost excels in accurately estimating discharge, its performance falters when tasked with admission prediction, falling short of the proficiency demonstrated by Logistic Regression in this particular area. This discovery underscores the importance of careful selection and evaluation of methods based on the specific objectives and requirements of the analysis. By understanding the strengths and limitations of each approach, researchers and practitioners can make informed choices to enhance the overall analytical process.

Figure 6

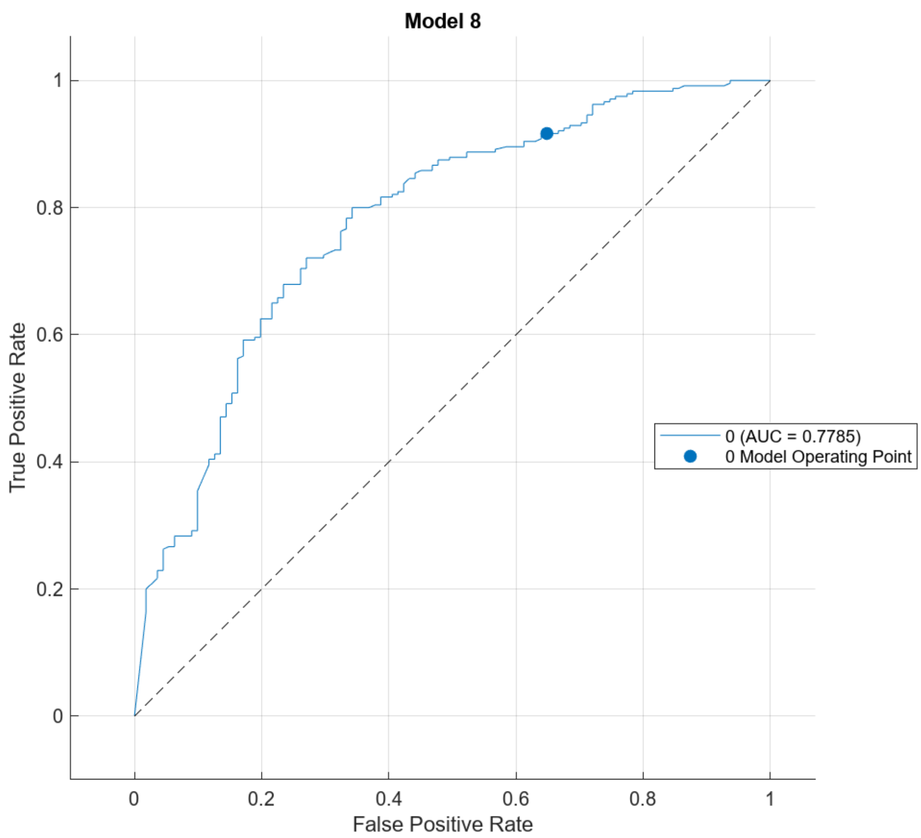

Figure 6 represents the AUC results of Adaboost

Upon thorough examination of the Accuracy metrics, it becomes evident that Adaboost achieves a higher level of success in predicting discharge compared to Logistic Regression. However, when considering the overall performance, Adaboost lags behind Logistic Regression. This finding highlights the nuanced nature of these two methods and their differential impact on different aspects of the analysis. While Adaboost demonstrates a stronger capability in accurately estimating discharge, its effectiveness diminishes when evaluating other relevant factors, leading to an overall lower success rate when compared to Logistic Regression. This observation underscores the importance of considering multiple evaluation metrics and the specific objectives of the analysis.

5.3. Gentleboost Results

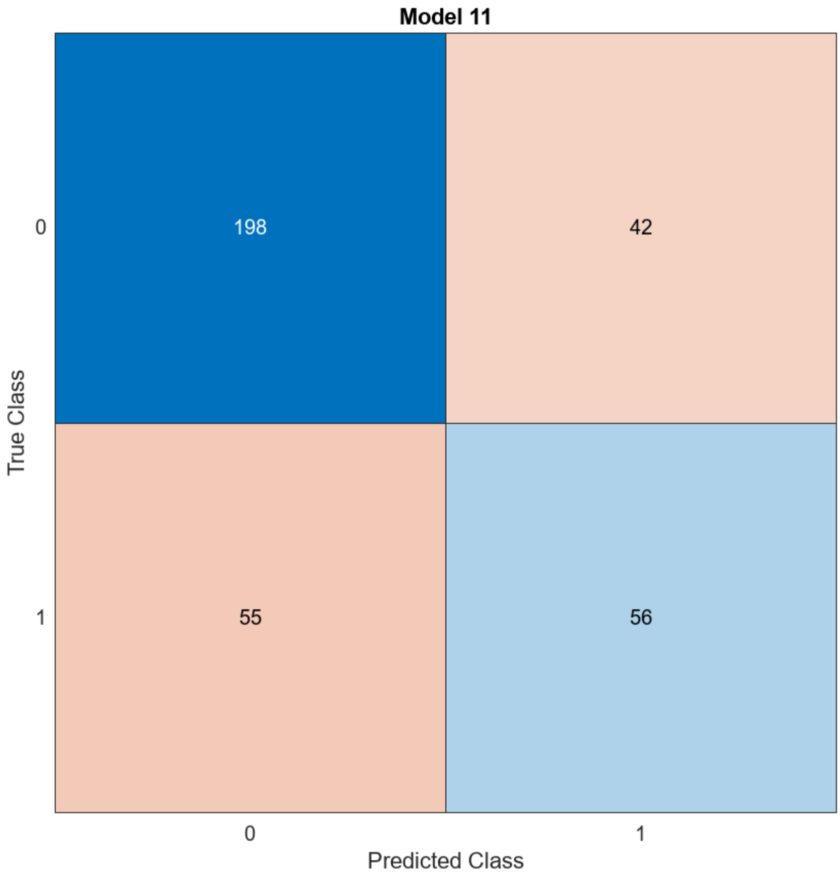

Figure 7 represents the Confision Matrix results of Gentleboost

Upon delving into previous studies, it becomes evident that Gentleboost—an ensemble method—has not been extensively utilized in this particular type of analysis. Consequently, there exists a notable knowledge gap regarding the performance of Gentleboost within this context, rendering its evaluation a matter of great interest. Through a detailed examination, we discovered that Gentleboost's performance, while not surpassing that of Adaboost, showcases remarkable similarities and comparable effectiveness. Specifically, in terms of admission estimation, Gentleboost demonstrates a notable level of success akin to Adaboost. However, when it comes to discharge estimation, Gentleboost falls behind both Adaboost and Logistic Regression, indicating a relative weakness in this aspect. These findings highlight the intricate nature of ensemble methods and the importance of exploring their specific strengths and weaknesses within distinct domains of analysis. By thoroughly evaluating the performance of Gentleboost alongside other established methods, we can gain deeper insights into its potential applications and optimize its usage accordingly. This comprehensive understanding of Gentleboost's performance will contribute to the broader knowledge base and aid researchers and practitioners in making informed decisions when selecting appropriate ensemble methods for similar analyse

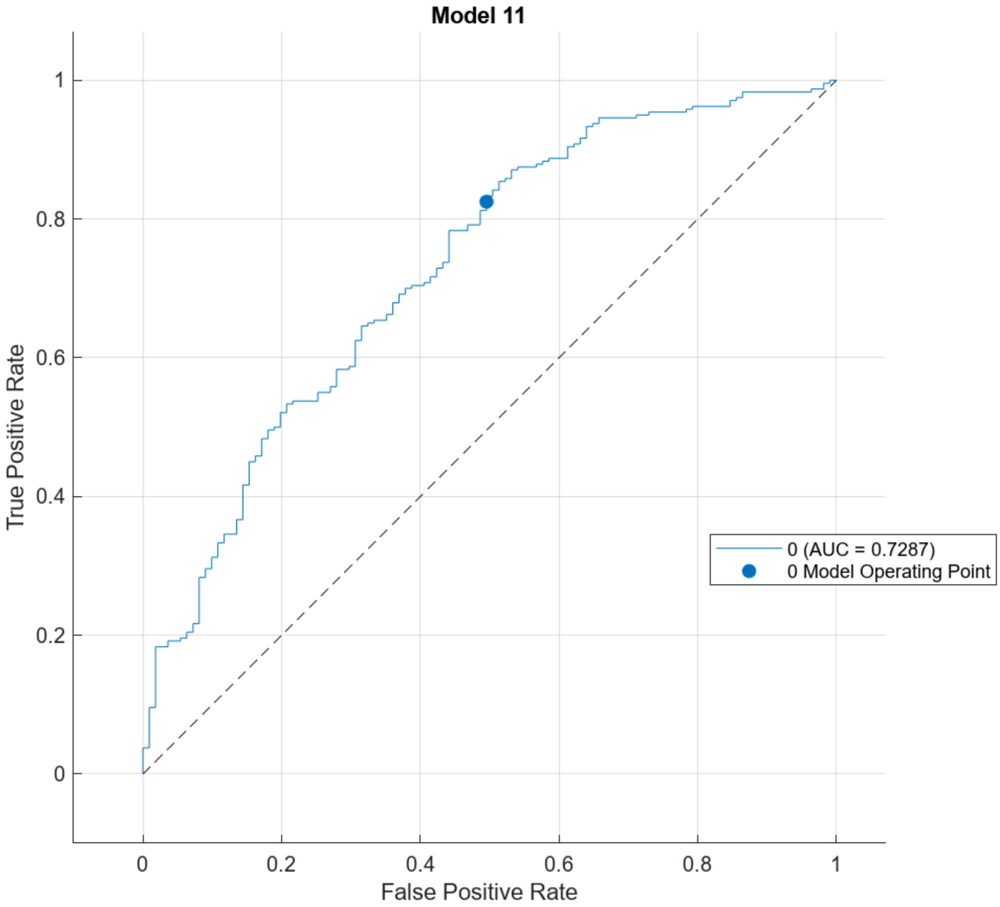

Figure 8

Figure 8 represents the AUC results of Gentleboost

In the comprehensive assessment of Gentleboost's overall performance, it becomes apparent that this ensemble method falls short of its predecessors. Despite its promising aspects, Gentleboost's performance, as evaluated through the AUC metrics, yields a moderate score of 0.72. This finding suggests that Gentleboost may not be the most optimal choice for utilization in studies of a similar nature. While it is essential to acknowledge the varying strengths and weaknesses of different methods, the relatively lower success rate of Gentleboost implies that alternative approaches, such as Adaboost and Logistic Regression, may be more suitable for achieving higher levels of accuracy and predictive power within this specific domain. These observations emphasize the significance of thorough evaluation and comparison of ensemble methods to identify the most appropriate and effective techniques for specific analytical objectives.

5.4. Logitboost Results

Figure 9

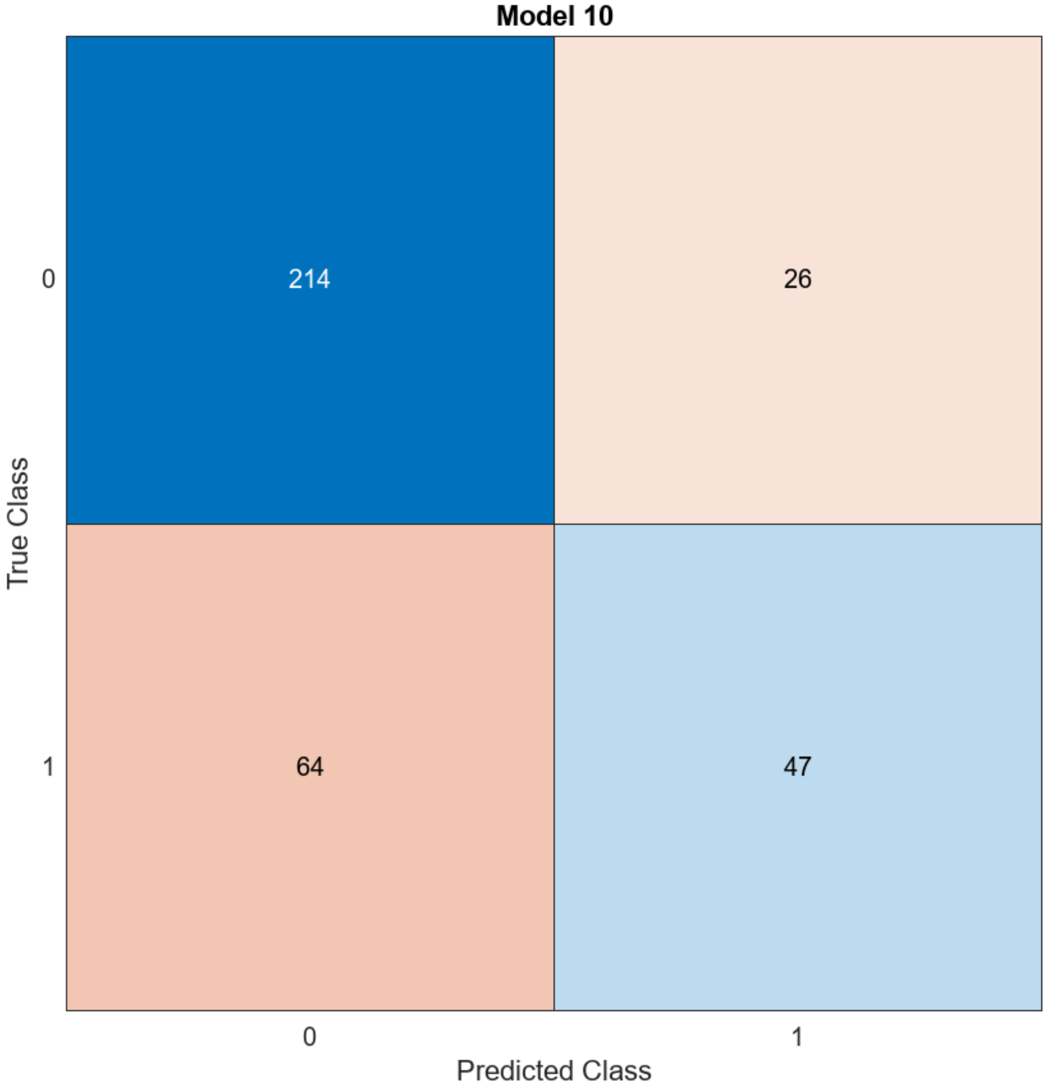

Figure 9 represents the Confusion Matrix results of LogitBoost

Upon meticulous evaluation, LogitBoost emerges as a method that, while not yielding results as exceptional as Logistic Regression, exhibits values closely resembling those obtained through Adaboost. Similar to other ensemble methods, LogitBoost demonstrates a higher degree of success in the discharge-related aspects of analysis, while its performance in the admission prediction domain falls slightly short. When considering the usability of LogitBoost, a review of the literature reveals that it can be a viable alternative to Adaboost, as it achieves comparable levels of success. This finding underscores the practicality and efficacy of LogitBoost, as it can be employed as a substitute for Adaboost without sacrificing significant predictive performance. This insight is particularly valuable for researchers and practitioners seeking versatile ensemble methods with similar success rates to Adaboost. By recognizing the relative equivalence of LogitBoost's performance to Adaboost, researchers can expand their repertoire of suitable methods and effectively address various analytical objectives.

Figure 10

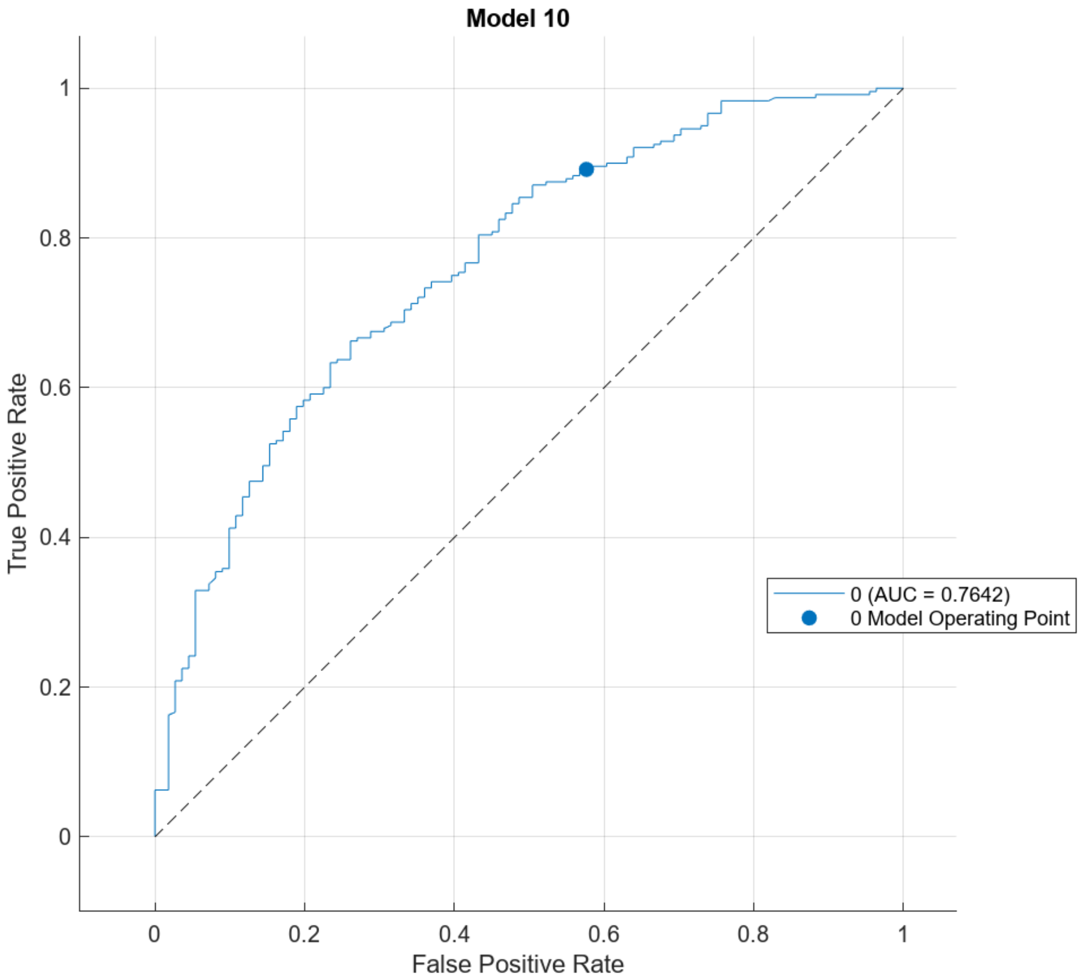

Figure 10 represents the AUC results of LogitBoost

When examining the AUC values, it becomes evident that LogitBoost, while not reaching the same level of success as Logistic Regression, closely aligns with Adaboost. This observation highlights the competitive performance of LogitBoost, which demonstrates a striking similarity to Adaboost in terms of predictive accuracy. Furthermore, LogitBoost outperforms Gentleboost, an ensemble method that is relatively uncommon in this particular field of study. The noteworthy success of LogitBoost positions it as a valuable alternative to Adaboost, providing researchers with a viable option to achieve comparable outcomes. By leveraging

LogitBoost, researchers can expand their methodological toolkit and attain similar levels of performance to Adaboost, thereby enhancing the overall robustness and applicability of their analyses. This finding not only broadens the horizons of researchers but also contributes to the advancement of ensemble methods in the field, paving the way for more flexible and effective analytical approaches.

5.5. Rusboost Results

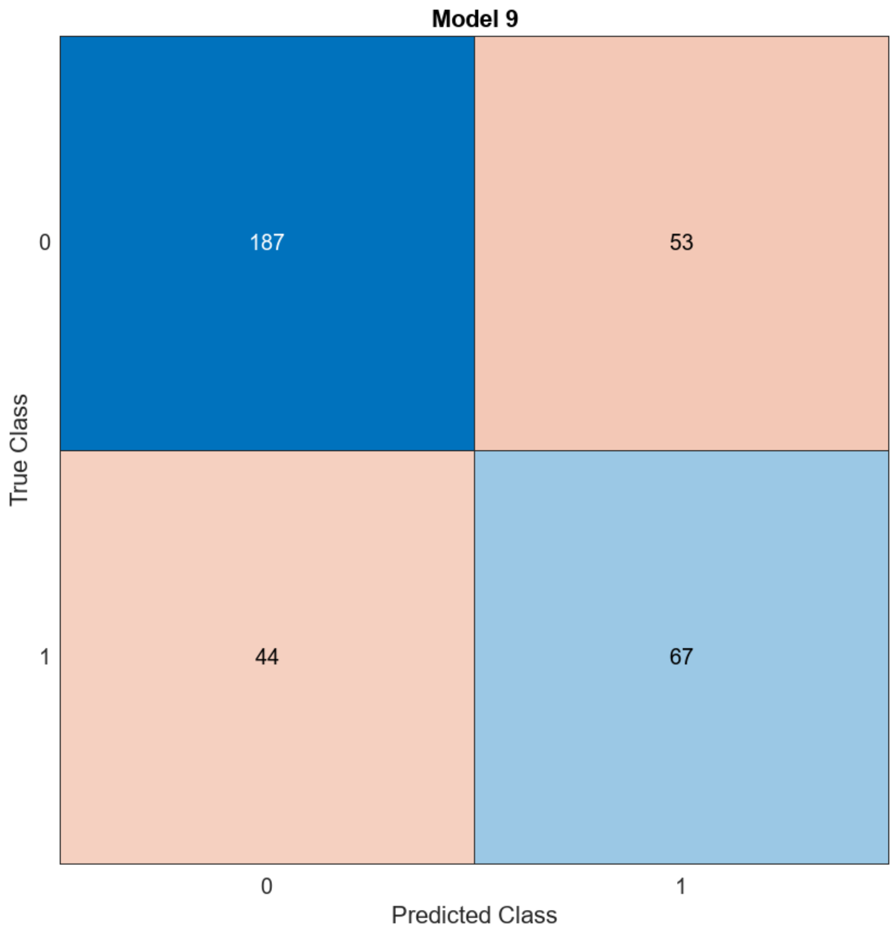

Figure 11 represents the Confision Matrix of Rusboost

In the landscape of research within this domain, an intriguing inclusion arises with the introduction of Rusboost, a method that is seldom encountered in similar studies. Typically, Logistic Regression has been the method of choice and widely utilized in such analyses. However, the incorporation of Rusboost serves the purpose of providing a comparative perspective among ensemble methods. Unsurprisingly, the results obtained from Rusboost demonstrate its relative inferiority to Logistic Regression in terms of overall performance. Nonetheless, when evaluating its efficacy alongside other ensemble methods, Rusboost emerges as more successful than Gentleboost, albeit falling short of the performance achieved by Logistic Regression. Notably, Rusboost exhibits a performance profile that closely resembles that of Adaboost, showcasing comparable levels of success. This finding suggests that Rusboost can be a valuable addition to the repertoire of ensemble methods, offering researchers a viable alternative to Adaboost while maintaining similar levels of predictive accuracy. The inclusion of Rusboost in this comparative analysis contributes to the expanding knowledge base and promotes the exploration and understanding of diverse ensemble methodologies within the research community.

Figure 12

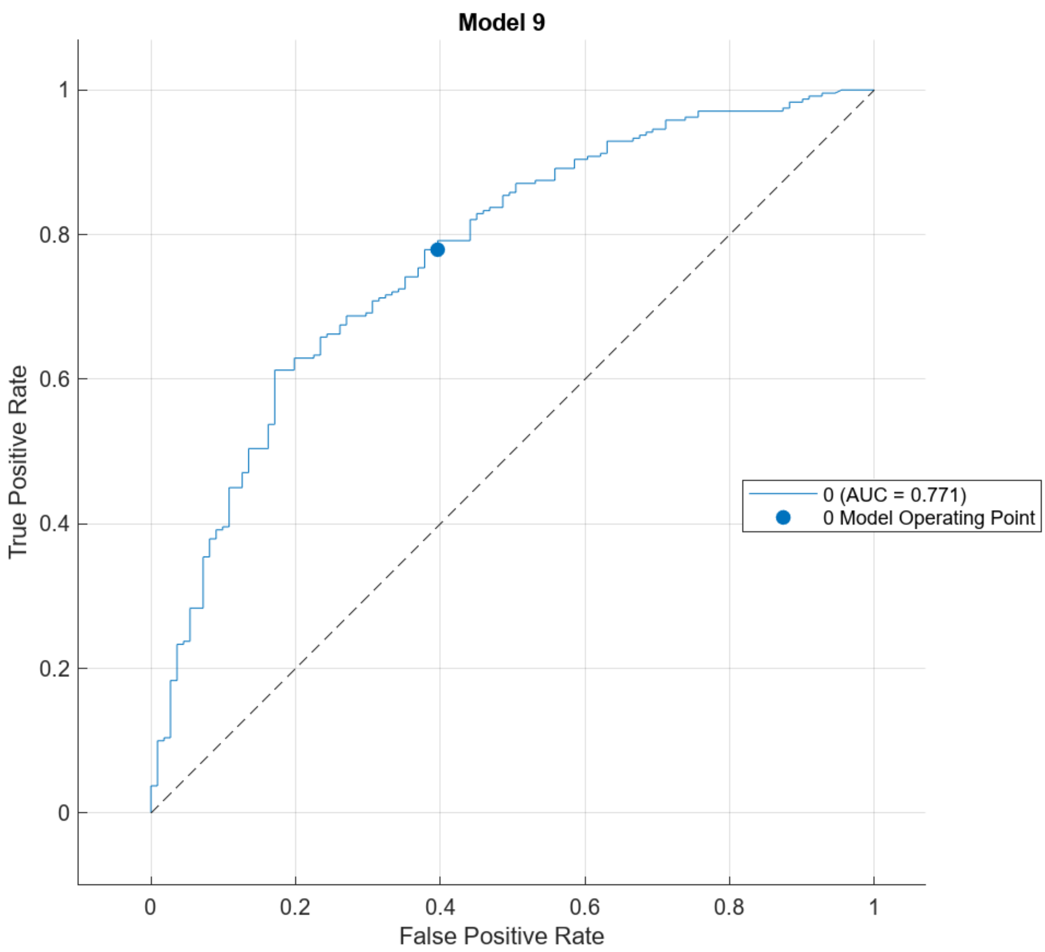

Figure 12 represents the AUC results of Rusboost

During the comprehensive AUC analysis, an intriguing observation emerged as Rusboost showcased remarkably similar results to Adaboost. This notable similarity further accentuates the promising performance of Rusboost, especially considering its relative rarity in such analytical endeavors. With an AUC value of 0.77, Rusboost demonstrated highly favorable outcomes, surpassing the performance of both Gentleboost and LogitBoost. These impressive results warrant a reevaluation of the methods commonly employed in this field and present an opportunity to explore the utilization of Rusboost as a viable alternative. By outperforming its ensemble counterparts, Rusboost introduces a fresh perspective and inspires researchers to reconsider their approach to analysis within this domain. The successful performance of Rusboost encourages a broader adoption of this method and stimulates further research and investigation to fully understand its capabilities and potential applications. In doing so, the field can benefit from the advancements brought forth by Rusboost, leading to enhanced analytical outcomes and expanding the possibilities for future studies.

Comparison

Measure	Adaboost	Logistic Reg.	Gentleboost	Logitboost	Rusboost
Sensitivity	0.7534	0.7956	0.7826	0.7698	0.8095
Specificity	0.6610	0.7179	0.5714	0.6438	0.5583
Precision	0.9167	0.9083	0.8250	0.8917	0.7792
Negative Predictive Value	0.3514	0.5000	0.5045	0.4234	0.6036
False Positive Rate	0.3390	0.2821	0.4286	0.3562	0.4417
False Discovery Rate	0.0833	0.0917	0.1750	0.1083	0.2208
False Negative Rate	0.2466	0.2044	0.2174	0.2302	0.1905
Accuracy	0.7379	0.7784	0.7236	0.7436	0.7236
F1 Score	0.8271	0.8482	0.8032	0.8263	0.7941
Matthews Correlation Coefficient	0.3333	0.4579	0.3416	0.3610	0.3752

Adaboost

Adaboost model has achieved decent performance but with room for improvement. The model shows a reasonably high sensitivity (0.7534), indicating that it correctly identifies a good proportion of the positive cases. However, the specificity (0.6610) is comparatively lower, suggesting that it struggles to accurately identify negative cases.

The precision (0.9167) is quite high, indicating that a large majority of the positive predictions made by the model are correct. On the other hand, the negative predictive value (0.3514) is relatively low, meaning that a significant proportion of the negative predictions are incorrect.

The false positive rate (0.3390) is moderately high, implying that a substantial number of negative cases are falsely classified as positive. The false discovery rate (0.0833) is low, indicating that the model makes relatively fewer false positive predictions.

The false negative rate (0.2466) represents the proportion of positive cases that are incorrectly classified as negative, which could be improved for better performance.

The accuracy of the model is 0.7379, suggesting that it correctly predicts approximately 73.79% of the cases overall. The F1 score (0.8271) is a balanced measure of precision and sensitivity, indicating that the model achieves a reasonable trade-off between these two metrics.

Lastly, the Matthews Correlation Coefficient (0.3333) is a measure of overall agreement between predicted and actual classifications. A value of 0.3333 indicates a moderate level of agreement.

In summary, while the Adaboost model demonstrates some positive aspects such as high precision and reasonable sensitivity, there is room for improvement in terms of specificity, negative predictive value, false positive rate, false negative rate, and overall accuracy. Further optimization and fine-tuning may be beneficial to enhance the model's performance.

Logistic Regression

Logistic Regression model has achieved reasonably good performance. The sensitivity (0.7956) indicates that the model correctly identifies a high proportion of the positive cases, while the specificity (0.7179) suggests it also performs well in identifying negative cases.

The precision (0.9083) is relatively high, indicating that a large majority of the positive predictions made by the model are correct. However, the negative predictive value (0.5000) is moderate, suggesting that the model's performance in correctly identifying negative cases can be improved.

The false positive rate (0.2821) is relatively low, meaning that a relatively small proportion of negative cases are falsely classified as positive. The false discovery rate (0.0917) is also low, indicating that the model makes fewer false positive predictions.

The false negative rate (0.2044) represents the proportion of positive cases that are incorrectly classified as negative. While this rate could be further improved, it is relatively low.

The accuracy of the model is 0.7784, suggesting that it correctly predicts approximately 77.84% of the cases overall. The F1 score (0.8482) is a balanced measure of precision and sensitivity, indicating that the model achieves a good trade-off between these two metrics.

The Matthews Correlation Coefficient (0.4579) is a measure of overall agreement between predicted and actual classifications. A value of 0.4579 indicates a moderate level of agreement.

In summary, the Logistic Regression model demonstrates good performance with relatively high sensitivity, specificity, precision, and accuracy. There is potential for further improvement, particularly in terms of negative predictive value and false negative rate. Fine-tuning and optimization efforts could be considered to enhance the model's overall performance.

Gentleboost

Gentleboost model's performance shows a mixed result with room for improvement. The sensitivity (0.7826) indicates that the model correctly identifies a decent proportion of the positive cases. However, the specificity (0.5714) suggests that it struggles to accurately identify negative cases.

The precision (0.8250) is relatively high, indicating that a majority of the positive predictions made by the model are correct. However, the negative predictive value (0.5045) is moderate, indicating room for improvement in correctly identifying negative cases.

The false positive rate (0.4286) is relatively high, implying that a significant number of negative cases are falsely classified as positive. The false discovery rate (0.1750) is relatively low, suggesting that the model makes fewer false positive predictions.

The false negative rate (0.2174) represents the proportion of positive cases that are incorrectly classified as negative. While this rate could be further improved, it is relatively moderate.

The accuracy of the model is 0.7236, suggesting that it correctly predicts approximately 72.36% of the cases overall. The F1 score (0.8032) is a balanced measure of precision and sensitivity, indicating a reasonable trade-off between these two metrics.

The Matthews Correlation Coefficient (0.3416) is a measure of overall agreement between predicted and actual classifications. A value of 0.3416 indicates a moderate level of agreement.

In summary, the Gentleboost model's performance shows a mix of strengths and weaknesses. It demonstrates relatively high precision and sensitivity but struggles with specificity and false positive rate. There is room for improvement in correctly identifying negative cases and reducing false positive predictions. Further optimization and fine-tuning may be necessary to enhance the model's overall performance.

Logitboost

Logitboost model demonstrates a reasonable level of performance.The sensitivity (0.7698) indicates that the model correctly identifies a decent proportion of the positive cases. The specificity (0.6438) suggests that it also performs reasonably well in identifying negative cases, although there is room for improvement.

The precision (0.8917) is relatively high, indicating that a majority of the positive predictions made by the model are correct. However, the negative predictive value (0.4234) is moderate, suggesting that the model's performance in correctly identifying negative cases can be improved.

The false positive rate (0.3562) is moderately high, implying that a significant number of negative cases are falsely classified as positive. On the other hand, the false discovery rate (0.1083) is relatively low, indicating that the model makes fewer false positive predictions.

The false negative rate (0.2302) represents the proportion of positive cases that are incorrectly classified as negative. Although this rate could be further improved, it is relatively moderate.

The accuracy of the model is 0.7436, suggesting that it correctly predicts approximately 74.36% of the cases overall. The F1 score (0.8263) is a balanced measure of precision and sensitivity, indicating a reasonable trade-off between these two metrics.

The Matthews Correlation Coefficient (0.3610) is a measure of overall agreement between predicted and actual classifications. A value of 0.3610 indicates a moderate level of agreement.

In summary, the Logitboost model demonstrates moderate performance, with relatively good precision and sensitivity. However, there is room for improvement in terms of specificity, negative predictive value, false positive rate, and overall accuracy. Further optimization and fine-tuning efforts could help enhance the model's performance.

Rusboost

Rusboost model demonstrates a mixed level of performance.The sensitivity (0.8095) indicates that the model correctly identifies a relatively high proportion of the positive cases. However, the specificity (0.5583) suggests that it struggles to accurately identify negative cases.

The precision (0.7792) is moderate, indicating that a majority of the positive predictions made by the model are correct. The negative predictive value (0.6036) is relatively high, suggesting that the model's performance in correctly identifying negative cases is better.

The false positive rate (0.4417) is relatively high, implying that a significant number of negative cases are falsely classified as positive. The false discovery rate (0.2208) is also moderately high, indicating that the model makes a significant number of false positive predictions.

The false negative rate (0.1905) represents the proportion of positive cases that are incorrectly classified as negative. While this rate is relatively low, it could still be improved for better performance.

The accuracy of the model is 0.7236, suggesting that it correctly predicts approximately 72.36% of the cases overall. The F1 score (0.7941) is a balanced measure of precision and sensitivity, indicating a reasonable trade-off between these two metrics.

The Matthews Correlation Coefficient (0.3752) is a measure of overall agreement between predicted and actual classifications. A value of 0.3752 indicates a moderate level of agreement.

In summary, the Rusboost model demonstrates a mixed level of performance. It shows relatively high sensitivity and negative predictive value but struggles with specificity and false positive rate. There is room for improvement in correctly identifying negative cases and reducing false positive predictions. Further optimization and fine-tuning may be necessary to enhance the model's overall performance.

6. CONCLUSION

In conclusion, the field of healthcare data analysis relies on a multitude of technologies and systems, such as health information exchanges, telemedicine platforms, clinical decision support systems, and electronic health records. These systems play a crucial role in improving patient outcomes, enhancing provider communication, and streamlining healthcare workflows. Machine learning, with its innovative approaches, has emerged as a valuable tool in addressing challenges faced by healthcare workers in emergency departments. Specifically, machine learning applications in triage and risk stratification have revolutionized the assessment and prioritization of patients in emergency care settings. By leveraging machine learning algorithms, healthcare professionals can accurately identify patients requiring urgent care and predict the severity of their conditions using various patient data sources. The ability of machine learning models to detect subtle patterns and predict adverse outcomes enables early intervention, potentially preventing complications and saving lives. While a wide range of analysis methods exist in the literature, there is a notable absence of comparative evaluations of commonly used ensemble methods. Hence, the primary objective of this study was to comprehensively examine and analyze various ensemble methods, aiming to compare their efficacy and performance. Through this investigation, valuable insights can be gained to enhance our understanding of these methods and inform future research and practice in the field of healthcare data analysis. In the comprehensive examination of ensemble methods, Logistic Regression emerged as the most prevalent and successful approach, consistently yielding the best results. Following closely behind, although not surpassing the success of Adaboost, Logistic Regression secured the second position with comparable values. Surprisingly, Rusboost, an ensemble method that is not commonly utilized in Emergency Department prediction studies, exhibited remarkably similar performance to Adaboost, presenting an exciting opportunity for future applications in this field. Additionally, Logitboost demonstrated a similar level of performance, albeit slightly lower than Rusboost. Although Logitboost did not outperform Adaboost, its comparable results indicate its potential as a reliable alternative. However, among the ensemble methods employed, Gentleboost proved to be the least successful. As expected, the results of Gentleboost, a method that is not frequently employed in these kind of studies, did not yield surprising outcomes. These findings highlight the varying degrees of success and popularity among ensemble methods, emphasizing the importance of carefully selecting the most appropriate approach for specific prediction studies in the Emergency Department. By considering the performance and prevalence of these methods, researchers can make informed decisions and further contribute to the advancement of predictive models in emergency care settings.

7. REFERENCES

[1] Tıbbi Servisler ve *Acil servis*. (n.d.). https://www.medicana.com.tr/. https://www.medicana.com.tr/tibbi-birimler/acil-servis

[2] N. R. Hoot and D. Aronsky, "Systematic review of emergency department crowding: Causes effects and solutions", *Ann. Emerg. Med.*, vol. 52, no. 2, pp. 126-136, 2008.

[3] Mamlin, B. W., Biondich, P. G., Wolfe, B. A., Fraser, H. S., Jazayeri, D., Allen, C., ... & Tierney, W. M. (2020). Cooking up an open-source EMR for developing countries: OpenMRS–a recipe for successful collaboration. Journal of the American Medical Informatics Association, 13(3), 253-260.

[4] McGinn, C. A., Grenier, S., Duplantie, J., Shaw, N., Sicotte, C., Mathieu, L., & Leduc, Y. (2011). Comparison of user groups' perspectives of barriers and facilitators to implementing electronic health records: a systematic review. BMC Medicine, 9(1), 46.

[5] Kawamoto, K., Houlihan, C. A., Balas, E. A., & Lobach, D. F. (2005). Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. British Medical Journal, 330(7494), 765.

[6] *Clinical Decision Support*. (n.d.). Agency for Healthcare Research and Quality. https://www.ahrq.gov/cpi/about/otherwebsites/clinical-decision-support/index.html#:~:text=Clinical%20decision%20support%20(CDS)%20provides,team%20and%20patient%20to%20consider.

[7] *What is an electronic health record (EHR)? | HealthIT.gov*. (n.d.). https://www.healthit.gov/faq/what-electronic-health-record-ehr#:~:text=An%20electronic%20health%20record%20(EHR)%20is%20a%20digital%20version%20of,and%20securely%20to%20authorized%20users.

[8] Nemati, S., Holder, A., Razmi, F., Stanley, M. D., Clifford, G. D., & Buchman, T. G. (2018). An interpretable machine learning model for accurate prediction of sepsis in the ICU. Critical Care Medicine, 46(4), 547-553.

[9] *Kunt, M. M. (2021). Emergency Medicine and Artificial Intelligence. https://dergipark.org.tr/en/download/article-file/1985451*

[10] Sun, J., Zhang, Y., & Tang, L. (2019). Predicting patient deterioration in the emergency department: A machine learning approach. Journal of Biomedical Informatics, 98, 103267.

[11] *What are The Top Machine Learning (ML) Methods?* (n.d.). Tableau. https://www.tableau.com/learn/articles/top-machine-learning-methods

[12] Brownlee, J. (2019, November 10). *14 Different Types of Learning in Machine Learning*. MachineLearningMastery.com. https://machinelearningmastery.com/types-of-learning-in-machine-learning/

[13] Tavasoli, S. (2023). Top 10 Machine Learning Algorithms For Beginners: Supervised, and More. *Simplilearn.com*. https://www.simplilearn.com/10-algorithms-machine-learning-engineers-need-to-know-article

[14] Johnson, D. (2023). Unsupervised Machine Learning: Algorithms, Types with Example. *Guru99*. https://www.guru99.com/unsupervised-machine-learning.html

[15] Kumar, Suyash & Kaur, Prabhjot & Gosain, Anjana. (2022). A Comprehensive Survey on Ensemble Methods. 10.1109/I2CT54291.2022.9825269.

[17] Rocca, J. (2021, December 9). Ensemble methods: bagging, boosting and stacking - Towards Data Science. *Medium*. https://towardsdatascience.com/ensemble-methods-bagging-boosting-and-stacking-c9214a10a205

[18] GeeksforGeeks. (2023). Gradient Boosting in ML. *GeeksforGeeks*. https://www.geeksforgeeks.org/ml-gradient-boosting/

[19] Great Learning Team. (2022, November 7). *AdaBoost Algorithm: Boosting Algorithm in Machine Learning*. Great Learning Blog: Free Resources What Matters to Shape Your Career! https://www.mygreatlearning.com/blog/adaboost-algorithm/

[20] Simplilearn. (2023). What is XGBoost? An Introduction to XGBoost Algorithm in Machine Learning. *Simplilearn.com*. https://www.simplilearn.com/what-is-xgboost-algorithm-in-machine-learning-article#:~:text=cost%20to%20accuracy!-,What%20is%20XGBoost%20Algorithm%3F,optimize%20their%20machine%2Dlearning%20models.

[21] Barak-Corren, Y., Chaudhari, P., Perniciaro, J., Waltzman, M. L., Fine, A. M., & Reis, B. Y. (2021b). Prediction across healthcare settings: a case study in predicting emergency department disposition. ., *4*(1). https://doi.org/10.1038/s41746-021-00537-x

[22] Lee, S. H., Chinnam, R. B., Dalkiran, E., Krupp, S., & Nauss, M. (2020). Prediction of emergency department patient disposition decision for proactive resource allocation for admission. *Health Care Management Science*, *23*(3), 339–359. https://doi.org/10.1007/s10729-019-09496-y

[23] Chen, C., Hsieh, J., Cheng, S., Lin, Y., Lin, P., & Jeng, J. (2020). Emergency department disposition prediction using a deep neural network with integrated clinical narratives and structured data. *International Journal of Medical Informatics*, *139*, 104146. https://doi.org/10.1016/j.ijmedinf.2020.104146

[24] LaMantia, M. A., Platts-Mills, T. F., Biese, K., Khandelwal, C., Forbach, C. R., Cairns, C. B., Busby-Whitehead, J., & Kizer, J. S. (2010). Predicting Hospital Admission and Returns to the Emergency Department for Elderly Patients. *Academic Emergency Medicine*, *17*(3), 252–259. https://doi.org/10.1111/j.1553-2712.2009.00675.x

[25] Parker, C. A., Liu, N., Wu, S. X., Shen, Y., Lam, S. S. W., & Ong, M. E. H. (2019). Predicting hospital admission at the emergency department triage: A novel prediction model. *American Journal of Emergency Medicine*, *37*(8), 1498–1504. https://doi.org/10.1016/j.ajem.2018.10.060

[26] B. Graham, R. Bond, M. Quinn and M. Mulvenna, "Using Data Mining to Predict Hospital Admissions From the Emergency Department," in IEEE Access, vol. 6, pp. 10458-10469, 2018, doi: 10.1109/ACCESS.2018.2808843.

[27] Zhang, X., Kim, J. S., Patzer, R. E., Pitts, S. R., Patzer, A., & Schrager, J. D. (2017). Prediction of Emergency Department Hospital Admission Based on Natural Language Processing and Neural Networks. *Methods of Information in Medicine*, *56*(05), 377–389. https://doi.org/10.3414/me17-01-0024

[28] Peck, J., Gaehde, S. A., Nightingale, D., Gelman, D., Huckins, D. S., Lemons, M. F., Dickson, E. W., & Benneyan, J. C. (2013). Generalizability of a Simple Approach for Predicting Hospital Admission From an Emergency Department. *Academic Emergency Medicine*, *20*(11), 1156–1163. https://doi.org/10.1111/acem.12244

[29] Sun, Y., Heng, B. H., Tay, S. Y., & Seow, E. (2011). Predicting Hospital Admissions at Emergency Department Triage Using Routine Administrative Data. *Academic Emergency Medicine*, *18*(8), 844–850. https://doi.org/10.1111/j.1553-2712.2011.01125.x

[30] Roquette, B. P., Nagano, H., Marujo, E. C., & Maiorano, A. C. (2020). Prediction of admission in pediatric emergency department with deep neural networks and triage textual data. *Neural Networks*, *126*, 170–177. https://doi.org/10.1016/j.neunet.2020.03.012

[31] Mowbray, F., Zargoush, M., Jones, A., De Wit, K., & Costa, A. P. (2020). Predicting hospital admission for older emergency department patients: Insights from machine learning. *International Journal of Medical Informatics*, *140*, 104163. https://doi.org/10.1016/j.ijmedinf.2020.104163

[32] Peck, J., Benneyan, J. C., Nightingale, D., & Gaehde, S. A. (2012). Predicting Emergency Department Inpatient Admissions to Improve Same-day Patient Flow. *Academic Emergency Medicine*, *19*(9), E1045–E1054. https://doi.org/10.1111/j.1553-2712.2012.01435.x

[33] Leegon J, Jones I, Lanaghan K, Aronsky D. Predicting hospital admission for Emergency Department patients using a Bayesian network. AMIA Annu Symp Proc. 2005;2005:1022. PMID: 16779309; PMCID: PMC1560667.

[34] Hong, W. S., Haimovich, A. D., & Taylor, R. J. K. (2018). Predicting hospital admission at emergency department triage using machine learning. *PLOS ONE*, *13*(7), e0201016. https://doi.org/10.1371/journal.pone.0201016

[35] *Emergency Service - Triage Application*. (2021, August 28). Kaggle. https://www.kaggle.com/datasets/ilkeryildiz/emergency-service-triage-application

[36] Grimmer, S., Bollmann, A., Seeger, I., & Guettler, A. (2016). Predicting hospital admissions from patient pathways: A systematic review of machine learning methods. JMIR medical informatics, 4(4), e44.

[37] Chen, W., Huang, Z., Li, L., Zhang, J., Li, C., & Xie, Y. (2018). Predicting hospital admission of chronic diseases using recurrent neural networks. Journal of Medical Systems, 42(8), 148.

[38] Hales, J. W., Eswaran, V., & Hertig, J. B. (2017). Data-driven modeling of patient satisfaction for improved hospital discharge planning. Health Informatics Journal, 23(3), 217-227.

[39] Shah, S., Liebovitz, D., & Malpani, A. (2019). Predicting hospital readmission prediction via cost-sensitive Logistic Regression

[40] Futoma, J., Hariharan, S., & Heller, K. (2017). Learning to detect sepsis with a multitask Gaussian process RNN classifier. In Proceedings of the 34th International Conference on Machine Learning-Volume 70 (pp. 1356-1365). JMLR. org.

[41] Hosmer Jr., D. W., Lemeshow, S., & Sturdivant, R. X. (2013). Applied logistic regression. John Wiley & Sons.

[42] Agresti, A. (2015). Foundations of linear and generalized linear models. John Wiley & Sons.

[43] Field, A. (2018). Discovering statistics using IBM SPSS statistics. Sage Publications.

[44] Allison, P. D. (2012). Logistic regression using SAS: theory and application. SAS Institute.

[45] Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. Journal of computer and system sciences, 55(1), 119-139.

[46] Zhu, J., Zou, H., Rosset, S., & Hastie, T. (2009). Multi-class AdaBoost. Statistics and its Interface, 2(3), 349-360.

[47] Friedman, J. H., Hastie, T., & Tibshirani, R. (2000). Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors). The Annals of Statistics, 28(2), 337-407

[48] Maldonado, S., & Song, T. (2015). GentleBoost: An introduction and literature survey. arXiv preprint arXiv:1511.06431.

[49] Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. Annals of Statistics, 29(5), 1189-1232

[50] Seiffert, C., Khoshgoftaar, T. M., Van Hulse, J., & Napolitano, A. (2010). RUSBoost: A hybrid approach to alleviating class imbalance. IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans, 40(1), 185-197.

[51] Nitesh, V. C., Johnson, A., Patel, N. R., & Patel, K. V. (2014). RUSBoost algorithm for imbalance dataset: A comprehensive review. International Journal of Computer Applications, 107(8)