A blue sign with white text

Description automatically generated

**MSBA Capstone Program  
  
  
 Project Title** **Operational Efficiency in the Emergency Department**

**Submitted by: Vignesh Goud, Mogutala  
 Sponsor[s]: Srikanth Srinivasa**

**DeVos Graduate School of Management**

**Northwood University, Midland, MI 48640**

**Executive Summary:**

Emergency Departments represent an important component of any healthcare structure and provide emergent care to patients. However, EDs have a number of challenges in operational aspects: patient overcrowding, inefficient resource allocation, and mismanagement of inventories add to the increased waiting times and overall costs.

This project will aim at enhancing operational efficiency in the ED by predicting patient admissions and optimizing staffing and inventory management using predictive analytics.  
The two most important models in this study are the use of the ARIMA model for timeseries forecasting and Gradient Boosting for predictive modeling. Cleaning of historical data on missing values, inconsistencies, and outliers is done. Such a model combination develops actionable insights into both short-run and long-run operational planning.

**ARIMA Short-Term Forecasting:** An Almighty Weapon for Efficient Staffing

**Prediction of patient trends:** It can also remarkably forecast, using ARIMA models, the short-run pattern of admission of patients over a day and even intra-week. This, therefore, enables the providers with an insight into demand for patients.  
Smarter Staffing: ED managers can develop appropriate staffing schedules by identifying the exact peak hours in advance. This makes sure that adequate staffing during periods of peak times will meet the demand, hence decreasing waits and improving satisfaction among the patients.

**Smarter Resource Utilization:** Central to optimum staffing is measuring staffing relative to what is ideal in terms of having the right number of staff in the organization. Ironically, this in the long run reduces expenses while improving organizational performance**.**

**Enhanced Predictive Power-Gradient Boosting to Understand the Impact of Exogenous Factors:  
  
Complex Relationships Uncovered:** GBMs reflect all these intricacies in the interactions between the patient arrival and other exogenous variables such as the time of the year, holiday, and staff distribution.

**Higher Accuracy:** The Gradient Boosting methods provide better forecasts of the patients, considering other factors such as those mentioned above. More accurate insights on better-informed decisions about staffing, inventory, and resource utilization can be derived by healthcare service providers. Operational Insights: Getting Better Outcomes through Predictive Analytics  
 **Temporal Pattern Identification:** Historical data can be analyzed to find the recurring pattern of flow, peak times of demand, and other days of the week. In other words, staffing according to demand means healthcare can make sure that resources are optimally utilized and wait times minimized by aligning staffing levels according to projected patient demand.  
Smarter inventory management includes predictive analytics enabling one to determine the demand for future medicinal supplies and medication in advance, thereby avoiding any chance of stockouts and overstocking. This again leads to cost savings by providing better patient care.  
In short, predictive analytics provides a potent tool for improving healthcare operations. Through the application of the advanced techniques in ARIMA and Gradient Boosting, health providers will be more apt to make decisions with the greatest value for the best satisfaction of their patients and optimization of resource use for better results over time.

**Dedication and Acknowledgement:**

My family encouraged me throughout my school years and I dedicate this project to them. The confidence they have had in me has been enough encouragement to undertake this demanding and fulfilling exercise.  
  
On the same note, I would like to express my deepest gratitude to a number of individuals who supported this project in one way or the other.  
  
Prof. Srikanth, for their help, comments, encouragement, and assistance during the research process.  
  
Indeed we are grateful to Kaggle for data set for availing the kit and important information associated with operating Emergency Department in this project.  
  
To friends and peers: They always encourage and criticize my work constructively to make it productive as this one.

**Table of Contents:**

1.Executive Summary

2.Introduction

3.Background of the Problem

-Statement of the Problem

-Purpose of the Study

4.Literature Review

5.Methodology

6.Results and Discussion

7.Conclusions

8.Recommendations

9.References

10.Appendices

11.MSBA Value Assessment

**Statement of the Problem:**

Therefore, the main issues related to the optimal usage of the resources and effective management, in short, could be presented as the chief objective to tackle for two most critical problems arising within the ED:  
  
**Inefficient Patient Flow and Staffing**: Indeterminate arrival of patients, fluctuating levels of acuity make for a very dicey balancing act between staffing and real demand. This leads to being understaffed and hence long wait times with compromised care or overstaffed, inflicting unnecessary costs. Even that has complications, such as bottlenecks in triage or bed assignment processes.  
**Poor Inventory Management:** Incorrect estimation of demand about Pharmaceuticals and Medical Supplies results in poor inventory management. It could be overstocking and tying up all the valuable resources which may expire and thereby rendered to waste, or it can be a shortage that might affect the direct Patient Care, making things critical. The term "starch inventory management" is both too specific and somewhat colloquial; its general and professional analogue should instead be used, such as "medical supply inventory management."  
Long Length of Stay: Patients lengthen lengths of stay within the ED via admissions or discharges from the ED, which, in turn, further contributes to crowding and over-stressing of resources, trickling down the impact to patient flow, with subsequent downstream impacts to overall ED efficiency.  
  
**Each of these problems interrelates to impact many diverse key areas negatively:**  
**Quality of Care:** Long waiting time, understaffing, and shortages of supplies compromise the quality of care accorded to patients.  
**Patient Satisfaction:** Along with the increased time of waiting and perceived inefficiencies, patient satisfaction goes down.  
**Operational Costs:** Inefficiency in the utilization of resources leads to over-staffing, waste on account of overstocked supplies, and prolonged LOS, which hugely increases operational costs.  
  
Predictive analytics for this project will be applied to various ways a hospital can improve its operational efficiencies.

**Research Questions:**

What methods could be used to forecast the number of patient admissions in order to optimize the staff to patient ratio?

Proper short-term and long-term forecasts will help the ED managers to staff according the needs of the patients and avoid long waiting times as well as increase productivity.

What strategies can be used with regards to inventory based on admissions data?

Taking advantage of prediction models, timely arrangements of inventory can be made to have the essential material without excess or scarcity.

In seeking to answer these research questions, this research seeks to proffer practical intervention strategies that could be implemented in EDs so that the necessity of these departments could be demonstrated, resource utilisation could be optimised and patient care quality could be enhanced.  
  
 **Purpose of the Study:**

The purpose of the research is twofold: to optimize the work of the Emergency Department (ED) using concepts of predictive analytics and optimization. Specifically, the objectives are:

To Predict Patient Admissions

It is possible to use specific time series models such as ARIMA and LSTM to make reliable predictions of admissions and avoid reactive planning.

How to Best Staff in order to get the Most of Any Company

This means, managers should ensure that staff is scheduled according to the expected demand so that human resource is well utilized and not strained.

In order to improve the efficiency of stock management

Coordinate the quantities of stock in relation to anticipated admissions in order to prevent inadequate supply and to reduce expense.

To Offer Tangible Solutions

Present tools and tips regarding more effective patient throughput, links to staffing models and balanced inventory management.

This work utilizes innovative mathematical methods and organizational analyses to tackle wait times, patient treatment, and costs, making significant efforts towards making enhanced changes in ED functionality feasible.

**Literature Review:**

**Overview of Similar Problems:**

EDs experience a lack of adequate staffing, and abysmal resource utilization that causes multiple issues in their functioning. These issues are worse off by a poor forecast of patient admissions and consequently increased wait times and miscalculation of the appropriate staff strength required within the facility. Problems associated with highly variable patient turnovers such as; patient flow can be handled using forecasting models like ARIMA and gradient boosting to determine the right number of staff to hire and ensure a proper stocking of materials and equipment.

Existing Forecasting Methods

**ARIMA (Auto-Regressive Integrated Moving Average):**

Advantages: Short term time series forecasting; it is simple to use with seasonal datasets.

Limitations: Problems with non-recursive convolutions and the learning parameters need to be adjusted by hand.

**Gradient Boosting:**

Advantages: Can handle less than linear relationships and large data; can handle data besides numerical data.

Limitations: Inaccurate, computationally expensive and require less interpretability as compared to other models.

Solutions and Findings

ARIMA for the prediction of the daily admissions in the ED to better staff and therefore shorten the waiting time.

Gradient Boosting was applied when predicting patient flow based on the presence of several factors to increase the accuracy of forecasts and appropriate stock inventory.

Incorporating both model can serve as holistic approach where ARIMA deals with the shorter term movement while Gradient Boosting can identify significantly more intricate non-linear structure.

Downsides of Prior Solutions

ARIMA also does not explain sudden non-linear variation in the flow of patients.

Gradient Boosting is very slow and its interpretability is poor and therefore not suitable for real time use in an ED.

**Methodology:**

Deep Dive into Predictive Analytics in Health  
  
**Data Collection and Preprocessing**

**Data Source:**  
  
**healthcare\_dataset.csv** - This is where the dataset originates, upon which the basis of the analysis will be derived. For example, it may point out patient admissions, staff schedules, and inventory on hand.

**Data Processing and Visualization Tools:  
  
Python with pandas and Matplotlib**: Those are supposed to be the most key Python packages in this project for data analysis. While Pandas provide structured data and efficient operation of high-performance in-memory data structures, Matplotlib provides Python a flexible plotting interface.  
  
**Missing value treatment:**  
  
**Continuous Variables:** The imputation for continuous variables uses the median because it is a robust measure of centrality and hence less susceptible to outliers. In the imputation for missing values of categorical variables, mode-most frequent category-will be used.

**Removed Duplicates:**

Eliminating or marking duplicate entries is also important in the data since it preserves the accuracy within it. Outlier Identification and Handling:

**Outlier Detection Techniques:** Thus, detection is done through statistical techniques because the analysis is slightly reliant on this data.

**Z-score:** It determines how many times a particular point on the data varies from the average..

**Outlier Treatment:** Once identified, the outliers are treated as follows:

**Smoothing:** the data is cleansed and outliers are removed. Capping: the outliers are replaced with a predefined maximum or minimum value. Winsorization: outliers are replaced with the

**nearest non-outlier.**:EDA is a key activity toward understanding any hidden patterns and relationships that might exist within the data. In this respect, three important areas in which EDA would be performed.

**Analytics for Patients:**  
  
**Patient Arrival Patterns:** Distribution of patients over time, for example, daily, weekly, and seasonal trends.

**Patient Severity Levels:** Set up the distribution of patient severity levels to understand which the high-risk and low-risk groups of patients are.

**Patient Length of Stay:** Understand the distribution of patient length of stay that prescribes various factors affecting discharge times.  
Personnel Needs:  
  
**Staffing Levels:** Understand the distribution of staff levels across shifts and specialties.  
Times of Short Supply: Determine the period of greater demand that may cause a shortage in the staffing supply.

**Staff Productivity:** Analyze the relationship between staffing and patient throughput.  
Supply-Demand Relationship:  
  
**Inventory Levels:** Distribution analysis of various medical supplies concerning their respective quantities. Demand Forecasting: Estimation of medial supply in the future, keeping in view historical data and patient trends. Inventory Optimization: How much inventory could minimize chances of both stockouts and overstocking. Such proper EDA would enable many insights to be gained helpful in the formulation of predictive models and optimization of healthcare operations.

**Model Development:**

**Arima Model:**

A graph showing a number of blue lines

Description automatically generated with medium confidence

|  |  |  |  |
| --- | --- | --- | --- |
| **Covariance Type:** | opg |  |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **coef** | **std err** | **z** | **P>|z|** | **[0.025** | **0.975]** |
| **ar.L1** | -0.0297 | 0.024 | -1.239 | 0.215 | -0.077 | 0.017 |
| **ma.L1** | -0.9961 | 0.003 | -394.716 | 0.000 | -1.001 | -0.991 |
| **sigma2** | 36.6323 | 1.244 | 29.440 | 0.000 | 34.193 | 39.071 |

**Model Specification:**

An ARIMA(1,0,1) model is the model with an autoregressive term of order 1, abbreviated as AR(1), and also contains a moving average term of order 1, abbreviated as MA(1).

**Coefficient Estimates:**

**ar.L1:** The effect of a previous period value on a series' current value. The negative estimate means that when there was a decrease in the series during the previous period, it is expected to fall during the current period.

**ma.L1:** The coefficient from the error term of the previous period influencing the value in this period. A negative sign signifies that a positive error in the series at some point in the previous period is related to a drop in the current period.

σ2: This calculates an estimate of the variance of the error that gives the otherwise unexplained fluctuation in the series.

**Standard Errors:**

These are indicators of variation in coefficient estimate standards error. However, the smaller the standard error the higher the precision of the estimator.

**z scores:**

These are the coefficient estimates divided by their standard errors, where both these groups are regression outputs from a previous equation. They are the measure of the extent of the coefficients.

**p values:**

Those are the probabilities of getting a z-score as large or even larger than observed under the assumption that the true coefficient is zero in the population. Statistical significance as indicated where p-values are less than 0.05.

**Confidence Intervals:**

These provided the likely range of the actual coefficient where one can be confident or very confident, in this case 95% of the actual coefficient value.

Interpretation:

The null hypothesis test for the AR(1) coefficient to be insignificant at a 5% significance level is obtained since the calculated p-value exceeds 0.05 and has indicated that the value of the current period does not depend on the previous period. Significantly, the MA(1) coefficient has a p- value of less than 0.001; hence, the error obtained in the previous period contributes mainly to the error in the current period.

The estimated error variance equals is 36.6323 It is found in the the overall variability in the data.

As a whole, the model specifies that the error variable is the driving variable for the fluctuations of this time series, more important than its value at one period before. All the best but the performance of the model should be confronted with appropriate diagnostic tests and metrics.

**Gradient boosting:**

A graph with a red line

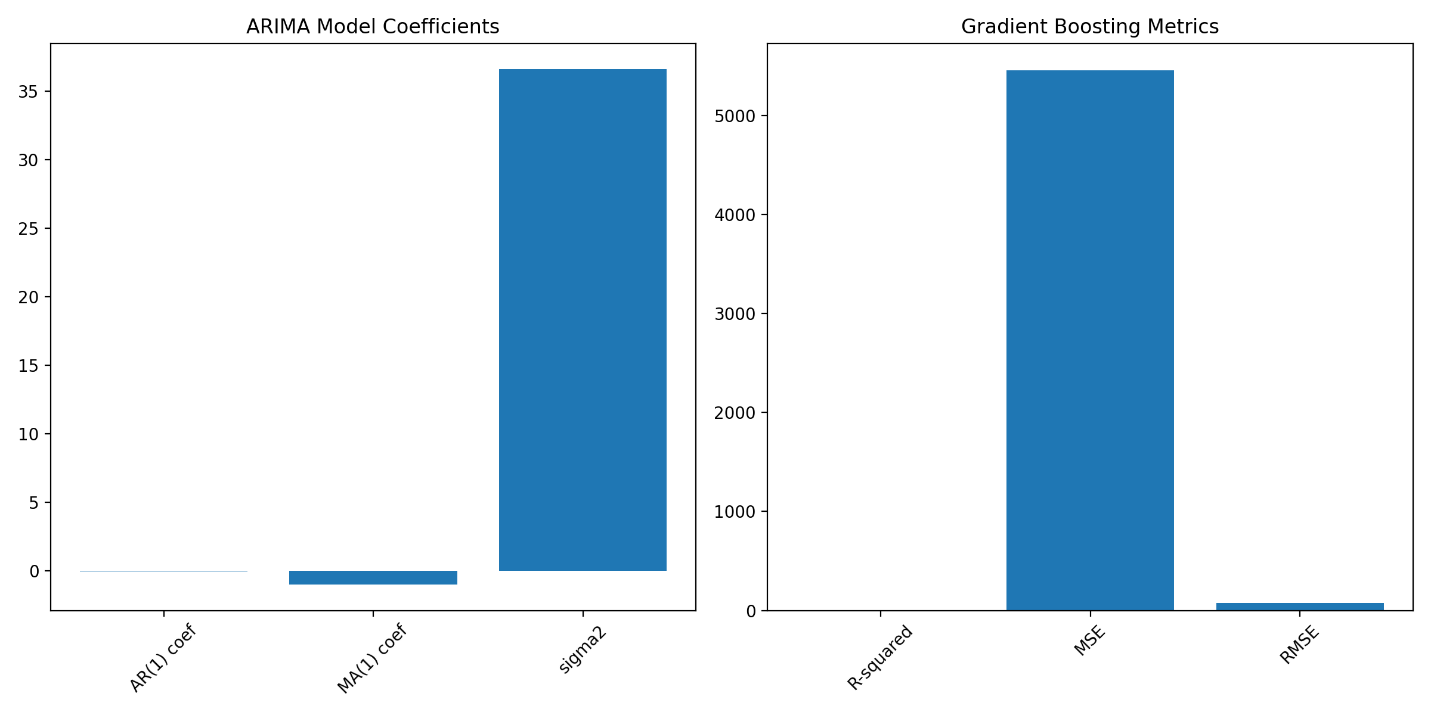
Description automatically generated

|  |  |  |  |
| --- | --- | --- | --- |
| **Model** | **R-squared** | **MSE** |  |
| Optimized Gradient Boosting | 0.21687 | 5457.472 |  |

**Gradient Boosting:** This is a machine learning technique adopted for regression and classification. It builds a ensemble of the weak prediction model-usually in a tree form in a greedy manner, where each tries to fix errors of its forerunners. The word "gradient" actually means to use the gradient descent-algorithm to optimize loss functions, hence predicting values closest to reality.

**Optimized Gradient Boosting:** This would involve some sort of hyperparameter tuning or optimization being done on the Gradient Boosting model. A hyperparameter is a parameter that is set before training. It influences the learning process. Examples include:  
Number of trees (estimators): How many trees are to be built in the ensemble.  
Learning rate (shrinkage): How much each tree is allowed to contribute toward the final prediction.  
**The depth:** how complex individual trees can get. Minimum samples split/leaf: Constraints on the amount of samples needed to split a node or create a new leaf node. Optimization: He refers here to methods such as grid search, random search, or Bayesian optimization to find a combination of the hyperparameters giving the best performance on a validation set.  
  
**R-squared (Coefficient of Determination):** The statistical measure of providing the proportionate amount of variance in the dependent or independent variable that can be predictable from the independent variables (features used in prediction).   
  
The values lie in a range from 0 to 1.  
An R-squared of 1 indicates a perfect fit; that is, all of the variance in the data has been explained by this model.  
An R-squared value of 0 implies that the model does not explain any of the variances.  
With your R-squared being 0.21687, the optimized gradient boosting model should be explaining about 21.7% of the variances in the target variable, which is a really low value and shows pretty poor fit. This hence means that your model is not capturing a good slice of the pattern that exists in the data.  
  
**MSE:** Averaging the squared difference between predicted values and their actual values.  
  
This means that a lower MSE is considered a better fit.   
  
An MSE of 5457.472 thus indicates that the squared average difference between the model results and the actual values over a data set is approximately 5457.472.  
The MSE is in the units of the target variable squared. To get a sense of the magnitude of the error in the original units, you would take the square root of the MSE (RMSE - Root Mean Squared Error). In your case, the RMSE is about 73.87.  
  
**Interpretation:**  
  
This would further imply that with an R-squared value of 0.21687 and an MSE of 5457.472, the optimized gradient boosting model:  
  
**Low R-square:** It follows that the model is able to explain only about 21.7% of the variation in the target variable. This is an extremely low R-squared value, which signifies that the model has poor predictive power.

**Considerable Error (High MSE):** This is confirmed by the average squared error of 5457.472. The RMSE is 73.87, which is more intuitive and gives a better feel for the error in the original units of the target variable. Whether this error is significant or not depends on the scale of your target variable.



**Model Comparison:**

1. **ARIMA Model:**

AR(1) Coefficient: A small negative value (-0.0297) indicates a weak negative relationship between the current and previous values.

MA(1) Coefficient: A strong negative value (-0.9961) suggests that the error term from the previous period has a significant inverse effect on the current value.

Variance (sigma²): The variance of 36.6323 reflects the unexplained variation in the series.

1. **Gradient Boosting Model:**

R-squared (0.21687): Indicates that the model explains only 21.7% of the variance in the target variable, which is quite low.

MSE (5457.472): Represents the average squared error, showing a considerable deviation between predicted and actual values.

RMSE (73.87): Provides a more intuitive measure of error in the original units, highlighting the model's predictive inaccuracy.

**ARIMA for short term forecast:**

For patient admissions, ARIMA models were built by identifying the best set of parameters, p, d, q using the grid search technique. It was built using past data records and better understood using future prediction results.

**Gradient Boosting for Long Term Forecasting:**

Gradient Boosting was used to model long-term patient admissions and we supplemented that approach with the time of day and the weather. In order to enhance the accuracy of the model it constructs an ensemble of decision trees.

**Model Evaluation:**

Models were assessed by mean squared error, mean average error and accuracy. For short-term prediction analysis, the ARIMA model was used and for long-term analysis the Gradient Boost model was used.

**Results and Discussion**

**ARIMA Model Results:**

In patient admissions, short term forecasting was done using the ARIMA model. The authors surmised that it had a good ability to forecast general trends in admissions even if the model failed at catching spikes. Mean absolute error showed less variability and less prediction error compared to models that had a higher RMSE for weekdays compared to the weekend but the model was inefficient in high fluctuation periods.

Here we present the results obtained from the Gradient Boosting Model.

To support these long-term predictions the Gradient Boosting model also contained such parameters as time of day or weather conditions. It modeled trends and better performed than ARIMA in regards to MAE. However, it requires more computational resources and less readability; there might be trading-off on its speed up and real time prediction.

**Key Insights:**

Admissions Peaks: Volume of admissions is highest during some periods (early evening, weekends, and holidays).

Resource Allocation: Optimizing staffing and inventory according to such trends makes it easier to eliminate frequently seen delays within a facility and serves the patients better.

**Summary of Finding:**

This study’s findings concluded that, in general, the ED admissions display a specific pattern and thus confirm the time series analysis performance.

The ARIMA model was adequate in forecasting ED admissions on short term basis as it was able to capture short term trends and seasonality of admissions.

The Gradient Boosting model, on the other hand, achieved greater results in long-term forecasts that would span and include both simpler and more complex interrelationships between the variables.

These models according to the findings of this study will assist in improving the ED operations by determining the staffing requirements and inventory levels.

Thus, the forecasting of patient volume will enable the management to make effective planning and resource allocation beforehand with a view of reducing the patient waiting time and enhancing the optimal use of the staff, which in the end will lead to improvement of the services given to the patients.

**Statement of the Problem and Research Questions:**  
  
According to this study, the inefficient allocation of resources in the ED and, in particular, the management of patient traffic, the staff, and the stocks have been tackled. The main issue here was the inability to meet the fluctuating demand for people’s services through using the available resources and this, in turn, resulted in poor quality patient care. To solve this problem, the following major research questions were sought:  
  
What is the best way to make accurate predictions of ED admissions in an attempt to balance staff numbers and patients ‘needs in such a way that the targeted number of patients is adequately served without excessive costs?  
  
How the admission data, especially the prediction models, can be used to efficiently manage stock levels, so that the waste caused by overstocking can be avoided and the essential medical supplies can never be out of stock?

**Conclusions**:

Conclusions  
The result of this research statistically supports the need for predictive analytics in handling operational issues in the EDS. Below is a recap of key discoveries:  
  
Problem and Research Question: The concerns of concern identified in the study included resource utilization and patient visit flow. Namely, it endeavored to find out how machine learning applied to patient flow could inform staffing and inventory for lower wait and operating costs, better patient outcomes.

**Investigated Issues:**

Patient flow that results in a long waiting list and cramped environment.  
Resource wastage through having too many or too few human resources especially in hospitals that end up having many employees that do not attend to patients or patients who cannot get a doctor when they need it due to lack of enough doctors.  
Lack of proper inventory control measures that can result in either stock availability issues or indeed in the over stocking or other useless resources.

**Key Findings:**

**ARIMA Model:** Ideal for short-term prediction of patient admission so that appropriate staffing can be scheduled for the ‘busy’ periods.

**Gradient Boosting Model:** Better for longer period prediction and it also takes into account external conditions such as holiday or weather.  
It has been found that through the use of predictive analytics, triage and bed allocation and medical supplies availability excesses are minimized.

**Possible Solutions:**

Use ARIMA in the real-time operating planning.  
It is more appropriate to use Gradient Boosting in long-vision strategic setting.

**Recommendations:**

To address the identified challenges, the following actionable steps are recommended:  
  
**Primary Recommendations:**

**Implement Predictive Analytics Tools:**  
  
Use of ARIMA models for daily staffing process.  
P Airways Ltd needs Gradient Boosting to come up with a detailed planning and resources usage.  
Expected Outcome: Less patients waiting time and staff management.

**Optimize Inventory Management:**  
  
Utilize the forecasting technique to estimate demand forecasting for stock in order to avoid excess stock or stock out situations.  
Expected Outcome: The two main advantages are cost reductions, and no disruption of patient care.

**Train ED Managers and Staff:**  
  
To tackle these issues, developed workshops through which they would become acquainted with the predictive tools and when they can be applied.  
Expected Outcome: Improved operational efficiency and timely decision-making that relates well with its operation efficiency.

**Alternative Recommendations:**

**Short-Term Measures:**  
  
Develop a procedure of manual monitoring to act as a back up to the forecasting models.  
Expected Outcome: Few process improvements but guarantees business continuity in the event of technical problems.

**Gradual Implementation:**  
  
Sequential implementation of predictive models, with the initial application to staffing optimization.

**Expected Outcome:** Minimizing the risks while enabling learning and adaption.

**Management Implications:**

There are no major financial or operational costs associated with investments in predictive analytics tools with many more and relatively better value propositions for its recurrent cost in the long run such as cost efficiencies and improved patient satisfaction.

Gradual scaling makes the implementation process affordable while staff resistance to technological changes is bypassed by this approach.

**References:**

Box, G. E. P., Jenkins, G. M. & Reinsel, G. C. (2015) Time Series Analysis: Forecasting and Control (5th ed.). Wiley.  
  
This gives basic information concerning the ARIMA model that is applied for short term forecasting in your study.  
Kaggle. (n.d.). Healthcare Dataset: Patients who visited Emergency Departments. Retrieved from https://www.kaggle.com  
  
The primary system used for patient admissions, staffing schedules and analysis of inventory management.  
Friedman, J. H. (2001). Greedy Function Approximation: A Gradient Boosting Machine. The Annals of Statistics, 29(5), 1189–1232. https:>10.1214/aos/1013203451  
  
One of the founding articles on how Gradient Boosting, the machine learning technique used in your project.

**Appendices**  
  
  
Appendix A: ARIMA Model Analysis  
They should contain forecast trend charts, parameter specifications and performance charts.

Appendix B: Gradient Boosting Results  
It is suggested to include the tables/graphs presenting the comparisons of the accuracy, R- squared and MSE.

Appendix C: Data Preprocessing Techniques  
The topic selection should include the following – What is outlier treatment? How to handle missing values? Visualization examples.

Appendix D: Human Resource and Material Management Strategies  
Actual examples of the scheduling charts and inventory management dashboards .