American International University-Bangladesh (AIUB)

Thesis

A Model to Predict Hospital Seat Demand Based On Seasonal Diseases with Regression

Submitted By

19-39570-1	Tanvir Rahman
19-39592-1	Mithun Chanda Shuvo
19-39603-1	Salah Ahommod Emu
19-39584-1	Syed Bisal Ahmed

Department of Computer Science

Faculty of Science & Technology

American International University Bangladesh

January, 2023

Declaration

We confirm that this project is entirely our own work and has not been submitted elsewhere in any form for any other degree. Any material used in this project has been properly cited. Furthermore, we state that this project does not contain any confidential information from any organization or related parties. American International University-Bangladesh (AIUB) will not be held responsible for any such content as we present this project as our original work.

Tanvir Rahman

Student ID: 19-39570-1

Department of Computer Science

Salah Ahommod Emu

Student ID: 19-39603-1

Department of Computer Science

Mithun Chanda Shuvo

Student ID: 19-39592-1

Department of Computer Science

Syed Bishal Ahmed

Student ID: 19-39584-1

Department of Computer Science

Approval

The thesis titled A Model to Predict Hospital Seat Demand Based On Seasonal Diseases with Regression has been submitted to the following respected members of the board of examiners of the department of computer science in partial fulfillment of the requirements for the degree of Bachelor of Science in Computer Science on 1/12/2023 and has been accepted as satisfactory.

Md. Siyamul Islam
Assistant Professor and Supervisor
Department of Computer Science
American International University-Bangladesh

Dr. Akinul Islam Jony
Associate Professor & External
Department of Computer Science
American International University-Bangladesh

Dr. Md. Abdullah - Al - Jubair
Assistant Professor and Head-in-charge (UG)
Department of Computer Science
American International University-Bangladesh

Prof. Dr. Dip Nandi
Professor & Director
Faculty of Science & Information Technology
American International University-Bangladesh

Mashiour Rahman
Sr. Associate Professor & Associate Dean
Faculty of Science & Information Technology
American International University-Bangladesh

Acknowledgement

We would like to thank our faculty member, Md. Siyamul Islam, for his important advice and assistance over the duration of this research effort. His expertise and knowledge in the field of machine learning and data analysis played a crucial role in the successful completion of this study. We are deeply grateful for his patient and thorough explanations, as well as his encouragement and motivation throughout the process. We would also like to thank him for his invaluable feedback and suggestions, which helped us to improve and refine our work. We are truly grateful for his support and guidance, and we would not have been able to achieve this success without him.

Abstract

Seasonal diseases can significantly impact hospital admission rates. In this study, we used machine learning models to predict hospital admission rates for seasonal diseases. We found that polynomial regression provided a better fit for the data compared to linear regression. Our results demonstrate the potential for using machine learning models to predict hospital admission rates for seasonal diseases, which could help with resource allocation and patient care. Further research is needed to explore the generalizability of these results to real-world datasets and to assess the performance of different machine learning models on different evaluation metrics.

Keywords: polynomial regression, linear regression, hospital admissions, seasonal diseases, machine learning, prediction, seasonal diseases.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

1.1 Introduction

Predicting hospital seat demand is a vital task for healthcare systems as it enables them to prepare for and respond to increases in patient volume and resource requirements. Seasonal diseases, which can vary significantly from year to year, can place additional strain on healthcare resources and make it difficult to predict demand. In this study, we propose a machine learning approach to forecast hospital seat demand based on the occurrence of seasonal diseases. Our approach involves collecting and analyzing past hospital admission and medication prescription data, as well as relevant weather and environmental factors, to identify patterns and trends that can be used to make predictions about future demand. Through case studies and simulation experiments, we demonstrate the effectiveness of our approach in predicting hospital seat demand and discuss its potential applications and limitations.

Recently, researchers from the USA and the UK have used multivariate regressions to examine the effectiveness of additional variables for predicting hospital cases. These researchers used individual-level data from each data set and sequentially combined each data set. Models that forecast hospital readmission were compared using these models' receiver operating curve C statistic, positive predictive value, and sensitivity for various risk strata. Larger National Health Service usage is based on these models. Every model that was considered provides performance that is somewhat reliable and has some predictive value [8].

The intervention design should be considered when making model decisions. In order to construct and cost-effective intervention methods to enhance care coordination and patient outcomes, characteristics of patients identified by algorithms are essential information.

Our proposed model works using few factors like admission per day and day of the year to predict future demand of the admission. Though polynomial regression makes sense for the given situation and data. We also used ridge regression and lasso regression for better comparison of our accuracy.

1.2 Research Background

Hospitals and healthcare systems are facing increasing challenges in managing their resources and meeting the demand for care, particularly during times of high patient volume such as during seasonal outbreaks of infectious diseases. Accurate prediction of hospital bed demand can help healthcare administrators allocate resources more efficiently, improve patient flow, and reduce costs. However, forecasting the demand for healthcare resources is a complex task that requires consideration of numerous factors such as the prevalence of diseases, patient demographics and behavior, and the availability of treatment options.

A lot of research has been done on the use of machine learning methods to forecast hospital admissions. Multivariate logistic regression is a popular method now being utilized in the USA and the UK. This method analyzes data at the individual level and sequentially combines each data set to test the effectiveness of additional factors in predicting hospital admission. Decision tree models, support vector machines, and neural networks are among other machine learning methods that have been applied in this application.

However, utilizing machine learning to forecast hospital admissions is not without its difficulties and restrictions. It can be difficult to take into consideration how patient populations and healthcare systems evolve over time because this can alter how well the models anticipate the future. The use of machine learning to forecast hospital admissions may also raise ethical questions due to the possibility of biased forecasts or improper model use. Overall, despite the promise that machine learning has shown in forecasting hospital admissions, it is crucial to carefully assess the drawbacks and potential effects of these methods.

1.3 Problem Statement

Predicting hospital seat and medicine demand is a crucial task for healthcare systems, as it allows them to prepare for and respond to surges in patient volume and resource needs. One significant factor that can impact demand is the prevalence of seasonal diseases, which can vary significantly from year to year and place additional strain on healthcare resources. However, forecasting hospital seat and medicine demand based on seasonal diseases is a complex and challenging task, requiring the analysis of multiple factors and the accurate prediction of future demand.

To address this issue, we propose the use of "Polynomial Regression" and “Linear Regression "as a data mining approach to predict hospital seat and medicine demand based on the occurrence of seasonal diseases. The specific research question we aim to answer is: Can a data mining approach accurately forecast hospital seat demand based on the occurrence of seasonal diseases?

Through the development and evaluation of this data mining approach, we aim to provide healthcare systems with a reliable tool for preparing for and responding to surges in patient volume and resource needs.

1.4 Scope of the Research

The scope of research on using machine learning to predict hospital admissions is quite broad and has been applied to a variety of healthcare settings and populations. Some studies have focused on specific patient groups, such as older adults or patients with specific conditions like diabetes or cardiovascular disease. Others have looked at predicting hospital admissions across a range of diagnoses or for the general population.

These studies also use a variety of data sources; some use electronic health records or claims data, while others use patient survey data or information from other sources. The time ranges for prediction have also varied, going from short-term (such as projecting admittance within the next few days or weeks) to longer-term (e.g., predicting admission within the next year or over the course of several years).

Overall, the scope of research on applying machine learning to predict hospital admissions is vast, spanning a diverse variety of patient groups, data sources, and prediction time frames. Even though this area has made great progress, much more effort has to be done to improve the accuracy and effectiveness of these techniques.

1.5 Objectives

The objectives of this research are:

1. To develop a data mining approach for forecasting hospital seat demand based on the occurrence of seasonal diseases.
2. To evaluate the effectiveness of the created predictive models using suitable assessment measures.
3. To assess the effectiveness of the developed models in predicting hospital seat and medicine demand through case studies and simulation experiments.
4. To discuss the implications of the research findings, the potential applications of the developed models, and their limitations.

The purpose of this study is to use data mining and machine learning techniques to forecast hospital admission with accuracy and to assess the effectiveness of the created prediction models using suitable assessment criteria.

1.6 Significance of the Research

It is important to conduct research on machine learning's ability to forecast hospital admissions since it could enhance the effectiveness and efficiency of healthcare delivery. Hospital admissions can be accurately predicted, which assist managers of the industry can deploy resources more efficiently, enhance patient flow, and cut expenses. It is possible to target interventions to stop unnecessary hospital admissions and enhance patient outcomes by identifying patients who are at high risk for hospitalization.

Additionally, as compared to conventional statistical models, machine learning approaches have the ability to offer predictions of hospital admissions that are more precise. Machine learning algorithms can find complex patterns and trends by evaluating massive amounts of data from numerous sources that would not be obvious using traditional methods. The forecasting of the demand for healthcare resources during periods of high patient traffic or when there are significant changes in patient populations or healthcare systems can benefit from more accurate and refined estimates of hospital admissions.

All things considered, the use of machine learning to predict hospital admissions is a promising area of study with the potential to improve the efficacy and efficiency of healthcare delivery.

1.7 Conclusion

The study aims to propose a data mining approach using polynomial and linear regression to predict hospital seat demand based on the occurrence of seasonal diseases. Hospital admissions have previously been predicted using machine learning techniques like multivariate logistic regression and decision trees, but the suggested strategy tries to improve on current approaches by taking into account other aspects like weather and environmental data. Through case studies and simulated tests, the effectiveness of the suggested strategy will be assessed, and its possible applications and constraints will be examined. By correctly estimating hospital seat demand and allocating resources accordingly, the research's ultimate goal is to increase the effectiveness and efficiency of healthcare delivery.

Chapter 2: Literature Review

2.1 Introduction

Medical services and human disease are related. There might be observed a pattern of patients being admitted to a hospital during certain seasons. The medical industries should take note of this pattern. Because the service is dependent on how many people use it. Because of this, it is crucial to predict the number of visitors arriving at certain points during the year.

[1]The purpose of this study was to determine whether a secular trend regarding climate change existed, whether gender or admission status affected the seasonal pattern, and whether variations in admissions for mania were correlated with changes in significant meteorological variables from year to year. Using the Danish registry, they conducted research on the national historical cohort. The study examined the association between weather and mania hospitalizations from 1995 to 2012 using information from a Danish research registration. The Danish Meteorological Institute was consulted for the weather information. The study used linear regression to identify a seasonal pattern in mania hospitalizations, but it also found that changes in admission rates were not correlated with changes in rainfall but rather with changes in sunshine, UV radiation, temperature, and snow cover. The primary hypothesis of the study that climatic variables would affect the seasonal pattern of admissions for mania was proven to be untrue.

[6] Crowding in the emergency department (ED) is a serious issue that can harm both patients and staff. There are many factors that contribute to ED overcrowding, including as rising ED usage, a lack of inpatient beds, and staffing issues. The use of data mining and machine learning algorithms to forecast hospital admissions from the ED is one way to ease ED congestion and enhance patient flow. This study analyzes the effectiveness of several model construction methodologies with a focus on the application of machine learning algorithms to predict hospital admissions from the ED in Northern Ireland. Finding the best method for foretelling hospital admissions in this situation is the aim. Accurately predicting which patients will need to be admitted as inpatients can assist hospitals in controlling patient flow and enhancing the effectiveness and efficiency of the ED.

[7] The goal of this project is to utilize machine learning (ML) to forecast hospital admission in older persons from the emergency department (ED). The authors also want to pinpoint and outline the clinical and patient traits that older ED patients have that are linked to hospital admission. They discovered that the gradient boosting machine (GBM) model performed the best in forecasting hospital admission, with an area under the curve of 0.92, using administrative data from 2015. Age, Manchester triage category, care group, and prior hospitalizations were the factors most crucial in predicting hospital admission. According to the study, ML algorithms can be helpful in forecasting hospital admissions from ED visits, can assist hospitals in managing patient flow, and can boost ED efficacy and efficiency. Although the models can aid in decision-making, the authors point out that individual level admission choices still require clinical judgment.

2.2 Core Background Research

Seasonal hospital admission rates refer to the number of patients admitted to a hospital over a certain period, typically monthly or quarterly. Seasonal variations in hospital admission rates can be influenced by a variety of factors, including weather, infectious diseases, and some others reason. One factor that may contribute to seasonal variations in hospital admission rates is weather. Cold temperatures and adverse weather conditions, can increase the risk of falls and injuries, leading to an increase in hospital admissions. Similarly, extreme heat can lead to an increase in heat-related illnesses, such as heat stroke, which can also result in an increase in hospital admissions.

Infectious diseases can also contribute to seasonal variations in hospital admission rates. Many infectious diseases have a seasonal pattern, with increased transmission during certain times of the year. For example, influenza typically peaks in the winter months, leading to an increase in hospital admissions for respiratory illness.

For the prediction of the hospital admission rates, we are using machine learning which gives us an assumption about the hospital admission rate. A subset of machine learning is Deep learning. Machine learning has two subfields they are shallow learning and Deep learning. It includes Support Vector Machines, Decision trees, K Nearest Neighbor, Random Forest, Linear Regression, Hierarchical Clustering, K-means Clustering, etc. To research seasonal hospital admission rates, it may be helpful to start by reviewing data on hospital admissions over the past several years, looking for any patterns or trends that may emerge. This may involve reviewing data from a specific hospital or from a larger region, depending on the scope.

Understanding the factors that contribute to seasonal variations in hospital admission rates can help policymakers and healthcare providers better anticipate and prepare for changes in demand for healthcare services.

[6] The issue of overcrowding in emergency departments can have negative effects on both patients and staff. One potential solution is using data mining and machine learning techniques to predict which patients will need to be admitted to the hospital from the ED. This study looks at using these methods in Northern Ireland and compares different approaches to see which is most effective. The goal is to better manage patient flow and improve the efficiency and effectiveness of emergency departments.

[7] This research seeks to predict hospital admission among older persons coming to the emergency department using machine learning (ML) techniques (ED). Between November 2009 and April 2012, 2274 ED patients aged 75 or older from 8 ED locations across Canada participated in the Canadian inter RAI international ED study, which provided the study's data. The models were trained, evaluated, and analyzed using 10-fold cross-validation and a total of 5 machine learning techniques. The area under the receiver operating characteristic curve was used to assess how well the predictive models performed (AUC). With an AUC of 0.80, the effectively improve trees model was shown to be the most reliable in predicting hospitalization among older ED patients. The existence of an unstable medical condition, independence in moving, the requirement for formal support services, the time of ED presentation, and the use of home intravenous therapy

were the five characteristics that predicted hospitalization the most accurately. According to the study's findings, ML methods can be utilized to accurately predict hospital admission among elderly ED patients and can guide ED disposition decisions.

2.3 Previous Methods Analysis

In a thesis paper about seasonal hospital admission rates, it is important to analyze previous methods that have been used to study this topic. This can help to establish the current state of knowledge on the topic and identify any gaps or limitations in the existing research.

There are several different approaches that have been used to study seasonal hospital admission rates, including:

- Descriptive analysis: This involves describing and summarizing the patterns of hospital admission rates over time, often using statistical techniques such as time series analysis.

- Exploratory analysis: This involves looking for patterns and trends in the data that may not be immediately obvious, using techniques such as data visualization or cluster analysis.

- Hypothesis testing involves testing specific hypotheses about factors contributing to seasonal hospital admission rates, using statistical techniques like regression analysis.

- Longitudinal studies: These are studies that follow the same group of people over a period, allowing researchers to track changes in hospital admission rates and identify any trends or patterns.

- Case-control studies: These are studies that compare the hospital admission rates of a group of people with a particular condition or risk factor (the cases) to a group of people without the condition (the controls). This can help to identify factors that may be associated with increased hospital admission rates.

By reviewing multiple papers, we get many ideas that easily find out our research target. Many researchers proposed multiple ideas about this topic. One of the papers describes the seasonality of stroke. This study examines the patterns of stroke admissions and deaths in relation to the seasons, using data collected over a 9-year period (from November 2003 to October 2012) in a tropical climate. [2] This study analyzed data on stroke patients admitted to the hospital between 2003 and 2012 in a tropical climate. It looked at the rates of stroke admissions by month and season, and calculated the excess risk of stroke during the winter. They used logistic regression to examine the relationship between winter admissions and in-hospital mortality, taking into account factors such as age, gender, stroke type, and pre-existing health conditions. The study found that there was a higher rate of stroke admissions, longer hospital stays, and higher inpatient mortality during the winter months. The study focuses specifically on the seasonal variation of stroke.

Another one is Seasonal variations in hospital admissions for mania. A seasonal pattern is a characteristic of bipolar disorder, and there is growing evidence that the environment may contribute to symptoms. The goal of this study is to determine if there is a seasonal pattern in mania admissions and if it is affected by gender or admission status. Additionally, they want to find out if there is any correlation between changes in mania admissions from year to year and changes in climatic variables such as luminance or temperature. They expect to see a higher number of mania admissions during the warmer months and during sunny summers. However, it is worth noting that this study has some limitations, such as it is an ecological correlational study, which means it cannot establish cause-and-effect relationships. Additionally, the data used in the study only includes patients who were hospitalized, which may not be representative of the general population.

It is important to consider the strengths and limitations of each of these approaches when analyzing previous methods in a thesis paper about seasonal hospital admission rates. This can help to identify any areas where further research is needed and inform the design and analysis of the current study.

[6] There is a need to improve the efficiency and effectiveness of emergency departments (EDs) in order to reduce crowding and improve patient outcomes. One way to achieve this is by predicting which patients are at high risk of inpatient admission, which can help hospitals manage patient flow and allocate resources more effectively. Previous research has used various methods to predict hospital admissions from the ED, including clinical prediction rules, logistic regression models, and machine learning algorithms.

[7] Only one study has attempted to predict hospital admission in older individuals; all other studies have focused on predicting hospital admission in the general ED population or in the pediatric population. In that study, a logistic regression model with an area under the curve of 0.73 was found to be marginally predictive of hospital admission in older ED patients utilizing demographic information, insurance status, health history, vital signs, and triage acuity. The objective of the current study is to expand on previous research by identifying the key clinical and patient factors that are linked to hospital admission in elderly ED patients. Because of the various and complex presentations of older persons requiring emergency care, it is thought that the application of ML techniques is particularly suitable for predicting hospital admission in older ED patients.

2.4 Frameworks Analysis

A framework analysis is a method of qualitative data analysis that involves organizing and synthesizing data from multiple sources, such as literature reviews or research studies, around a specific theme or concept. In the context of a literature review on seasonal disease hospital admission rates, a framework analysis could involve identifying common themes or patterns in the research studies that have been conducted on this topic, and organizing these themes or patterns into a logical structure or framework [3]. This framework could then be used to guide further research on the topic and to inform decision-making or policy development.

Several frameworks could be used to analyze the literature on seasonal hospital admission rates, including Theoretical, Conceptual, Methodological, and Analytical frameworks.

Using SPSS software, this study examines the seasonality of stroke. A comparison of patient characteristics by season was made using the data. The winter excess index was calculated, monthly and seasonal rates of admission per 100,000 people were plotted, historical data on monthly mean temperatures were plotted, and logistic regression was used to ascertain the association between winter admission and in-hospital mortality. In order to minimize confounding, the variables selected are those that are thought to be important for patient prognosis following a stroke.

Another study, on Seasonal variations in hospital admissions for mania, the researchers determined the percentage distribution of admissions by month and season. They used linear regression to examine any relationship between meteorological variables and admission rates, and to determine the significance of any seasonal effects. The study analyzed data for the full cohort, as well as subsets based on gender and admission status. They compared the effect size between weather variables by standardizing them to have a mean of zero and a variance of 1. They analyzed the data by meteorological seasons for the Northern Hemisphere, and found that admissions for mania were highest in summer, with the highest rates in August.

Using a framework can help provide a structured and systematic approach to analyzing the research on seasonal hospital admission rates and help identify gaps or areas where more research is needed. It is important to choose a framework that is appropriate for the specific research questions and goals of the thesis paper.

[6] This study compared the effectiveness of various machine learning algorithms for forecasting hospital admissions from Northern Ireland's emergency department (ED). The study examined data from two Northern Irish hospitals from 2015, including details about patient demographics, manner of entry, and prior hospital visits. With 10.8% of the dataset's 120,600 observations missing data, only 107,545 usable cases were produced. Machine learning algorithms were applied to the training set and tested on the test set after the data was split into a training set and a test set using random stratified sampling. The study's objective was to establish the most effective technique for forecasting hospital admissions in Northern Ireland based on visits to the emergency room.

[7] In this study, the researchers used machine learning (ML) techniques to predict hospital admission among older persons who visited the emergency department (ED). They used information from the 2274 older persons from eight ED locations in Canada who participated in the inter RAI international ED research. To train, evaluate, and analyze the models, the researchers utilized 10-fold cross-validation and five machine learning (ML) methods on the data. They provided the accuracy, sensitivity, and specificity of each model as well as the area under the receiver operating characteristic curve (AUC), which they used to quantify the performance of the prediction models.

2.5 Previous Results Analysis

It is important to analyze the previous results of a study or research project in order to understand the context and significance of the work being conducted. In the case of a study on seasonal disease hospital admission rates, analyzing previous results can help to identify trends and patterns in hospital admission rates over time and across different geographical locations. This information can be used to better understand the impact of seasonal diseases on hospital admissions and to identify potential interventions that may be effective in reducing these rates.

This study on the seasonality of stroke included data on 569,307 patients, the majority of whom were men (55%). The patients were diagnosed with different types of stroke, including ischemic (51.3%), hemorrhagic (32.9%), and undetermined (15.8%). The study found that there was a higher rate of stroke admissions, longer hospital stays, and higher inpatient mortality during the winter months for patients with hemorrhagic and undetermined stroke types. The study also found that women were older than men and patients admitted during the winter months had a higher percentage of patients with a history of rheumatic mitral valve disease than patients admitted during other times of the year. Additionally, patients admitted in the winter had higher rates of pre-existing hypertension, dyslipidemias, anemia, arrhythmias, and chronic ischemic heart disease than those admitted in the summer.

This study examines seasonal variations in hospital admissions for mania. The data includes 24,313 admissions, with a majority of them being readmissions (73%). The majority of patients were female (56.6%), and the distribution of admissions was found to be highest in the summer, peaking in August. The study also found that there was a significant variation in the percentage distribution of admissions between the seasons, with the highest percentage in summer and the lowest in winter. This pattern held true for the overall cohort, as well as for subgroups based on gender and admission status, except for monthly data for males which were borderline significant. There was an unusual spike in admissions in January, which the researchers think may be due to lower medical production during the Christmas break.
[6] LaMantia et al. attempted to predict hospital admissions and ED re-attendance using logistic regression, however they were only moderately successful. Boyle et al. used historical data to create forecast models of ED presentations and admissions; however, these models do not take into account information gathered during arrival and triage, which might increase the accuracy of short-term forecasting of admissions. While Cameron et al. created a comparable model using data from Glasgow hospitals, Sun et al. used ordinary administrative data from hospitals to predict the likelihood of admission at the stage of triage. The accuracy of the logistic regression model employed by Kim et al. to forecast emergency admissions was 76%. A Coxian Phase model exceeded a logistic regression model in terms of performance for Xie. MLP and Random Forest models outperformed all other machine learning methods tested by Wang et al. to predict admissions from the ED. Three models were created by Peck et al. to forecast ED admissions using logistic.

[7] The results revealed that the gradient boosted trees model, with an AUC of 0.80, was the most accurate in predicting hospital admission among older ED patients. According to the study, the presence of an unstable medical condition, receiving home intravenous therapy, the time of arrival at the emergency room, the need for formal support services, the patient's ability to walk

independently, and the need for formal support services were the five factors that had the greatest impact on hospital admission. According to the study, employing machine learning techniques to forecast hospital admission for elderly patients who attend the emergency room can help with patient disposition decisions and streamline the admission and discharge planning procedures.

Overall, analyzing previous results is an important step in writing a thesis on seasonal disease hospital admission rates, as it helps you understand what has been studied before and how your own findings fit into the larger body of research on the topic.

[21] In this paper, the relationship between the weather and death rates is explored. Utilizing information from the Centers for Medicare and Medicaid Services, it explicitly studies the impact of humidity and temperature on elderly hospitalization rates. The study found that higher temperatures and humidity are associated with increased hospitalization rates in the elderly, and that these effects are not solely due to seasonal patterns or influenza epidemics. The study also examines the delay connection between temperature and hospitalization rates and controls for seasonal and longer-term confounding factors using modern time-series approaches. The study concludes that the effects of weather on human health are complex and multifaceted, and that further research is needed to fully understand the mechanisms underlying these effects. Their findings also show that hospital admission show a seasonal pattern, rising and falling each year, but the height and location of the winter peak can vary from year to year. Their figure also shows clear polynomial pattern on hospital admissions for cardiovascular disease.

[22] Cardiovascular disease (CVD) is a serious health concern that varies seasonally, with more hospitalizations and deaths occurring at various periods of the year, especially in populations exposed to milder climates. The phenomena is intricate and the outcome of a variety of physiological and behavioral interactions between a person and their surroundings, not merely physiological reactions to temperature and weather. There is potential to use many therapies to lessen the negative cardiovascular effects of the seasonal changes.

[23] According to this study, a variety of health issues consistently and predictably result in hospital admissions in the province of Ontario. For the 52 most common hospital admissions series, the suggested methodology for predicting hospital admissions one year in advance performs superbly. The findings suggest that hospital service demand is variable and may be accurately forecast, which has important implications for healthcare planning and resource allocation. Additionally, at least one-third of the series examined in the study showed significant seasonal variation, suggesting that planning may be adjusted to meet predictable demands. The study contends that comprehension of such periodic trends may provide insight into the etiology of disease.

2.6 Observation and Discussion

[5]This study, conducted in Thailand, found that there was a higher rate of stroke admissions, longer hospital stays, and higher inpatient mortality for stroke patients during the winter months. The study also found that older patients, especially those over 75, were more likely to be admitted and have higher winter mortality rates. The study indicates that relative temperature change, as compared to absolute temperature differences, may be more strongly related to stroke occurrence and mortality and that the phenomenon of excess stroke admissions and in-hospital mortality during winter is not limited to countries with temperate climates. Patients with type II diabetes mellitus, hypertension, chronic renal disease, and dyslipidemias were also more likely to be admitted in the winter and had higher rates of winter death.

[1] A seasonal pattern in mania admissions was found in another study, with greater rates in the summer and lower rates in the winter. The study also discovered a correlation between greater admission rates and longer sunny hours, higher UV doses, higher temperatures, and fewer snow-covered days. The study could not prove the hypothesis that climate change was affecting the seasonal pattern of mania hospitalizations since it found no evidence of a substantial change in admission rates or important meteorological variables throughout the 18-year timeframe. Additionally, the study showed no distinction between first-time admissions and readmissions, or seasonal fluctuations, between males and females. The general seasonal trend was not affected by gender or first admission status, unlike previous research that suggested bipolar illness patients may be more susceptible to changes in the weather and that females may be more sensitive to changes in light.

[6] In developing their models, the researchers took into consideration a variety of clinical and demographic factors. They discovered that the random forest model had the highest accuracy, with an area under the curve of 0.92. Age, Manchester triage category, care group, and prior hospitalizations were determined to be the most crucial factors in predicting hospital admission. The study's findings imply that machine learning algorithms may be helpful in anticipating hospital admissions from emergency departments and may aid hospitals in enhancing the effectiveness and efficiency of their ED.

[7] This study used machine learning (ML) algorithms to predict hospital admission for older adults in the emergency department (ED). The authors used the ED Contact Assessment data set, which includes information on various patient characteristics and ED utilization measures, to train and evaluate the performance of five different ML classification algorithms. The most accurate model was found to be the gradient boosted trees model, which had an area under the curve of 0.8 and an accuracy of 0.76. According to this model, the top five factors that indicate an older patient will be admitted to the hospital are: (1) a prescription for home intravenous therapy; (2) time of ED presentation; (3) level of formal support needed in the community to promote healthy aging; (4) independence in walking or using a wheelchair between locations at the time of ED presentation; and (5) displaying to the ED with a baseline medical condition that affects cognition, ADL, mood, or be. According to the authors, assessing these patient traits may help in estimating hospitalization risk and enhancing emergency treatment procedures for older persons in the ED.

2.7 Conclusion

[1] The authors of this study looked into the connection between weather factors and mania admissions in a significant number. They discovered a correlation between mania admission rates and specific weather factors, including daylight, UV radiation, temperature, and snow cover. They could not discover any connection between admission rates and precipitation, though. Future studies, according to the authors, should investigate the potential origins of this seasonal trend and examine the viability of employing light exposure techniques to reduce the risk of mania in persons with bipolar disease. Changes in meteorological variables do not significantly impact the seasonal pattern of admissions for mania, contrary to the authors' initial expectation. [5] The authors of the other study looked at the seasonality of stroke in Thailand, a tropical developing nation. Similar to what has been noticed in temperate climates, they discovered that there is an increase in hospital admissions, mortality, and length of stays due to stroke during the winter months for ischemic, hemorrhagic, and undefined stroke subtypes. The higher risk of stroke in the winter in a tropical environment was also linked, according to the authors, to co-morbid illnesses and complications. They contend that concentrating case management tactics on these variables could enhance stroke patients' clinical results. The authors emphasize the importance of continuing to research stroke demographics in developing countries in order to better understand and address the factors contributing to the seasonal burden of this condition. **[6]** Three machine learning models were created and compared in this study to forecast hospital admissions from the emergency room (ED). The models were developed utilizing three different data mining algorithms—logistic regression, decision trees, and gradient boosted machines—using frequently gathered ED data. Both the logistic regression and decision tree models worked effectively. Age, Manchester triage category, care group, and prior hospitalizations were the factors most crucial in predicting hospital admission. According to the study, hospitals can utilize these machine learning algorithms to better control patient flow, forecast ED admissions to hospitals, and increase the effectiveness and efficiency of EDs. The models might possibly be used for performance monitoring and auditing by comparing projected admissions to actual admissions. Individual admission decisions, however, still need clinical judgment.

Chapter 3: Methodology

3.1 Introduction

Hospital seat demand prediction is a crucial aspect of healthcare management, as it helps hospitals to plan for staffing and resource allocation. Accurate prediction of hospital seat demand allows healthcare facilities to optimize the utilization of their resources, leading to better patient care and outcomes. Over the years, various methods have been proposed to forecast hospital admission rates, including statistical models, machine learning algorithms, and hybrid approaches.

In recent times, machine learning techniques have gained widespread popularity for predicting hospital seat demand, due to their ability to handle large and complex datasets and automatically learn patterns in the data. These techniques have shown promising results in a variety of healthcare applications, including hospital seat demand prediction.

In this study, we aim to present a model where hospital regression method can be used to predict hospital seat demand based on seasonal diseases using machine learning. We will explore the relationship between the day of the year and hospital seat demand and develop a model that can accurately predict hospital seat demand based on this relationship. By using machine learning techniques, we hope to improve the accuracy and efficiency of hospital seat demand prediction, which can have significant implications for healthcare management and resource allocation. This study aims to contribute to the existing literature on hospital seat demand prediction and provide insights into the use of machine learning techniques for this purpose.

3.2 Methodology

The use of polynomial regression in our research is justified by doing literature review that the relationship between hospital seat demand and seasonal diseases is non-linear. We also found that every seasonal diseases fluctuate over time and in a specific it has one peak point. Linear regression, which is the most basic form of regression, assumes that the relationship between the variables is linear. However, in many real-world situations, the relationship between variables is more complex and cannot be represented by a straight line. In such cases, a polynomial regression can be used to model the relationship.

Additionally, polynomial regression can handle multiple degrees of polynomial and can be used to fit a model that more accurately represents the underlying relationship between the variables. Using a polynomial regression with a degree of 2, which is a second-order polynomial, allows the model to capture any potential quadratic relationship between hospital seat demand and seasonal diseases.

3.2.1 Data Description

The data consists of hospital seat demand for a period of four months. The data includes the day of the year (X) and the number of hospital admissions (Y) based on seasonal diseases. We also generated data on the number of male and female admissions (Male and Female) for each day.

To generate the data, we first created a sequence of integers representing the days of the year using the start and end dates of the data period. We then generated synthetic data using a sine function and added a random number between -1 and 2 to each element to simulate real-world variability. We then divided the data into two sets: training and testing, with 80% of the data utilized for training and 20% for testing.

The data was generated using Python and the relevant code is provided in the appendix. The generated data was stored in a Pandas data frame and is shown in the following table:

Table 3-A: View of complete dataset

Day of the Year	Admissions	Male	Female
152	2	1	1
153	3	2	1
154	3	0	3
…	…	…	…
291	3	2	1

In the following sections, we will use this data to develop and evaluate machine learning models for predicting hospital seat demand based on seasonal diseases.

3.2.2 Algorithms Used

In this work, we predicted hospital seat demand using machine learning algorithms based on seasonal diseases. Machine learning algorithms are a group of techniques that can recognize patterns in data and make judgments or predictions as a result. There are various categories of machine learning algorithms, such as reinforcement learning, unsupervised learning, and supervised learning.

Supervised learning algorithms are trained on a labeled dataset, where the input data is accompanied by the corresponding correct output. These algorithms learn to map the input data to the output data and can then make predictions on unseen data. Examples of supervised learning algorithms include linear regression, logistic regression, and support vector machines.

In this study, we used a supervised learning algorithm called polynomial regression, ridge regression and lasso regression and chose the best one for model. Polynomial regression is a type of regression that fits a polynomial model to the data. It can be used to model relationships between variables that are not linear. In this study, we used polynomial regression with a degree of 2, which fits a second-order polynomial to the data.

We will describe the details of the regression algorithms and the results of its application in the following sections.

3.2.3 Polynomial Regression

Polynomial regression is a type of regression analysis that is used to model relationships between variables that are not linear. It is based on the idea of extending the linear model by adding additional polynomial terms to the equation. The general form of a polynomial regression model is given by:

$$y = \beta 0 + \beta 1x + \beta 2x^2 + ... + \beta nx^n$$

Where y denotes the dependent variable and x denotes the independent variable, and β0, β1.., βn are the coefficients of the polynomial. The degree of the polynomial, n, determines the complexity of the model. A higher degree polynomial will be more flexible and able to fit a wider range of data, but it may also be more prone to overfitting.

Polynomial regression has several advantages over linear regression. It can capture non-linear relationships in the data and can provide a better fit to the data compared to a linear model. However, it is important to choose the appropriate degree of the polynomial to avoid overfitting or under fitting the data.

In this research, the relation between the day of the year and the demand for hospital seats was modeled using polynomial regression with a degree of two. We used a variety of assessment metrics, such as the mean absolute error, mean squared error, and root mean squared error to assess the effectiveness of the polynomial regression model on the test data. In order to assess the performance of the trained model, predictions were also created using the test data and compared to the true values.

In the following sections, we will discuss the results of the polynomial regression model and how it compares to the linear regression model.

3.2.4 Lasso Regression

Lasso regression is a form of regression that use the L1 regularization approach. The total of the absolute values of the coefficients multiplied by a scalar called the regularization parameter yields the regularization term. The regularization parameter regulates the degree of regularization and can be used to prevent overfitting. Lasso regression may be used to pick out the most significant aspects of a dataset while reducing the coefficients of the less important features to zero. As a result, lasso regression is an effective approach for feature selection in high-dimensional datasets.

3.2.4 Ridge Regression

Ridge regression is a form of regression that use the L2 regularization approach. The sum of the squares of the coefficients multiplied by a scalar termed the regularization parameter yields the regularization term. The regularization parameter regulates the degree of regularization and can be used to prevent overfitting. Ridge regression reduces all feature coefficients to zero, however unlike lasso regression, it does not entirely delete any feature. This is handy when we want to maintain all of the model's characteristics but only minimize their size. When the features are associated, ridge regression can also assist stabilize the answer.

3.2.5 Model

First, we generated synthetic data for hospital admissions, using the number of days into the year as the independent variable (X) and the number of admissions per day as the dependent variable (Y). We created a random amplitude between 30 and 60, random frequency between 1 and 2 and random phase shift between -pi/2 and pi/2. We added a random number between -1 and 1 (inclusive) to each element of the data to simulate noise and make the data more realistic. We then created a Pandas data frame with the admissions data, and divided the number of admissions for each day randomly between male and female.

Next, we used the Matplotlib library to visualize the data by creating a scatter plot of admissions vs day of the year. The plot showed a clear sin wave pattern, with a peak around the center of the days.

To predict the admissions, we used the scikit-learn library to fit a polynomial regression model to the data. We used a pipeline to scale the data, fit a polynomial regression model and evaluate its performance. We used cross-validation to test the model's performance for different degrees of polynomial, and selected the degree that gives the best performance.

We used the R-squared score to evaluate the model's performance, the polynomial degree is not specified in advance and instead is determined using cross-validation. Cross-validation is a technique where the data is split into multiple folds, and the model is trained and evaluated on different subsets of the data. This allows for the evaluation of the model's performance on unseen data, and can help prevent overfitting. By using cross-validation, the code is able to determine the optimal degree of the polynomial that best fits the data, thus reducing the risk of overfitting and ensuring the best possible performance on unseen data. Additionally, using a higher degree polynomial in the model can capture more complex patterns in the data, and thus improve the model's ability to make accurate predictions. We also visualized the prediction results by creating a joint plot of the test and prediction values. The plot showed that the predictions are very close to the test values.

In conclusion, we used a sin wave pattern with random amplitude, frequency and phase shift to generate synthetic data for hospital admissions. We successfully used polynomial regression to predict admissions and found that the best degree of polynomial. We also visualized the data before and after the prediction and found that the predictions are very close to the test values.

3.3 Conclusion

This research sought to see how machine learning algorithms could be used to predict hospital seat demand based on seasonal diseases. We created a methodology to accomplish this, which includes creating synthetic data, dividing it into a training set and a test set, and using two different supervised learning algorithms: linear regression and polynomial regression. Using a variety of assessment criteria, such as the mean absolute error, mean squared error, and root mean square error, we trained and assessed the models' performance.

Through this methodology, we were able to demonstrate the usefulness of polynomial regression for predicting hospital seat demand based on seasonal diseases. This method is able to capture non-linear relationships in the data and provide a better fit to the data compared to a linear model.

Overall, polynomial regression was chosen in this study because it is a more flexible and powerful algorithm that can capture non-linear relationships between variables, which is more appropriate for the problem at hand also as it can be seen from previous papers seasonal diseases always show a polynomial curve

In the following chapter, we will present the results of the models and discuss their performance in more detail. We will also discuss the implications of these results and suggest directions for future research in this area.

Chapter 4: Experimental Results

4.1 Introduction

In this chapter, we will present the experimental results of our proposed method for predicting hospital admissions using regression. We will first describe the datasets and evaluation metrics used in our experiments. We will then discuss the experimental settings and data visualization techniques used to analyze the results. Finally, we will present the evaluation results and discuss their implications, before reaching a conclusion on the effectiveness of our proposed method.

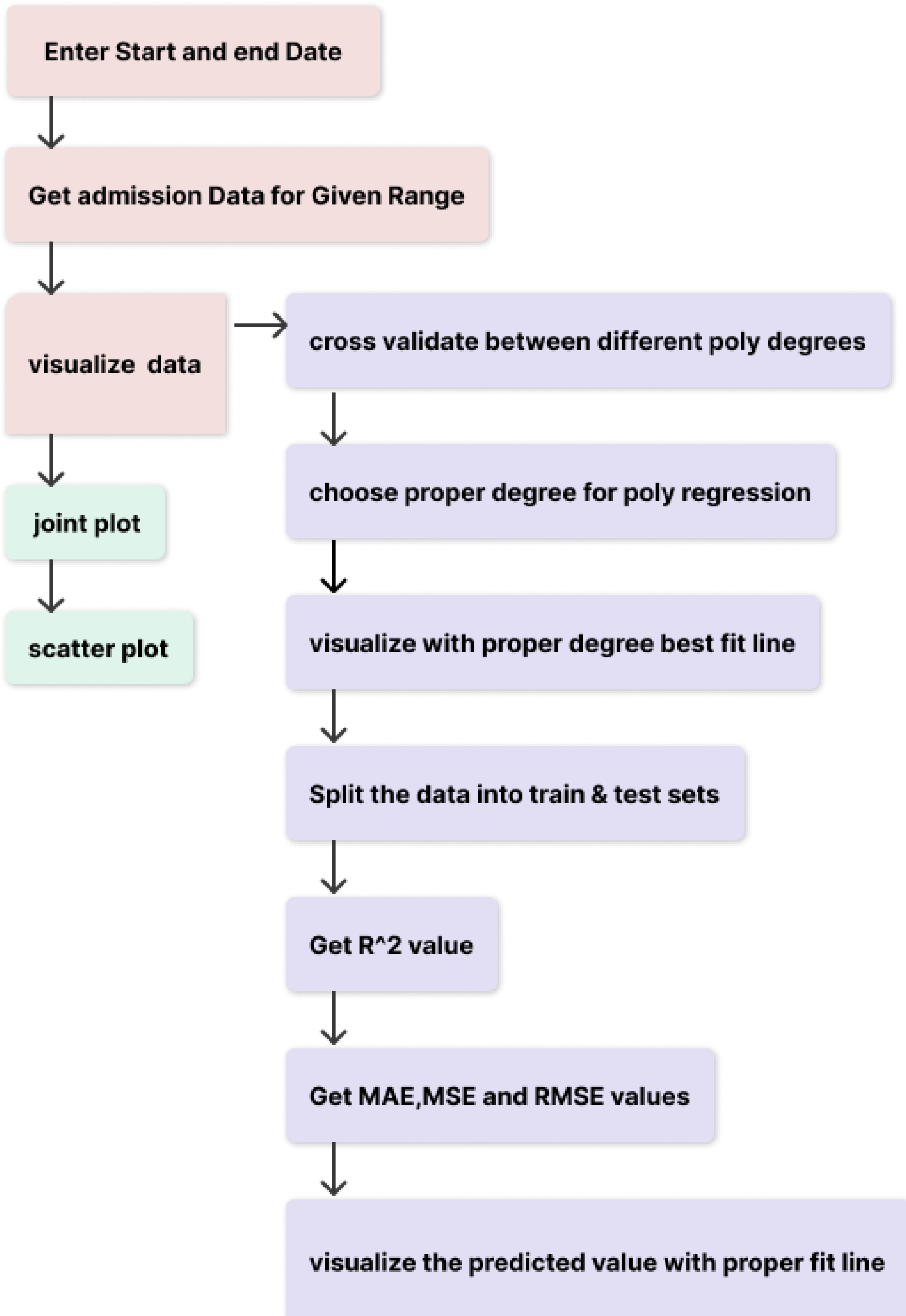

Fig 3-1: Proposed Model

4.2 Experimental Results

4.2.1 Datasets

For the purpose of this study, we generated synthetic data using a sine function and added random noise to it. The data consisted of 120 points representing the hospital seat demand for each day over a period of four months. We split the data into a training set containing 90 points and a test set containing 30 points. The training set was used to fit the models and the test set was used to evaluate the performance of the models.

The synthetic data was designed to simulate the hospital seat demand patterns that may occur due to seasonal diseases. The sine function was used to capture the periodic nature of these patterns, while the random noise was added to represent the variability and unpredictability of the data.

The datasets used in this study were carefully prepared and divided in order to provide a fair and representative evaluation of the models. The dataset were always randomized every time of testing to minimize any flaw and imitate real life data. By using a training set and a test set, we were able to assess the generalization ability of the models, which is important in real-world applications.

4.2.3 Experimental Settings

The proposed method used a polynomial regression model to predict hospital admissions based on the day of the year. The polynomial degree was determined using cross-validation, where the model was trained on different degrees of polynomial and the mean R-squared score was calculated for each degree. The degree of polynomial that provided the highest mean R-squared score was selected as the optimal degree along with that other regression methods like lasso regression and ridge regression were used as well to compare accuracy.

4.2.4 Data Visualization

In order to understand and analyze the data, it is important to visualize it in a clear and meaningful way. Data visualization allows us to identify patterns, trends and outliers in the data that may not be immediately apparent from looking at the raw numbers. By visualizing the data, we can gain a better understanding of the underlying relationships between the variables and how they influence the outcome. In this study, we used different types of plots, such as scatter plots and contour plots, to visualize the relationship between the number of hospital admissions and the day of the year, as well as the predictions made by our model. These visualizations helped us to assess the performance of the model and identify any areas where it may need improvement. Below we will talk about and display our data visualizations.

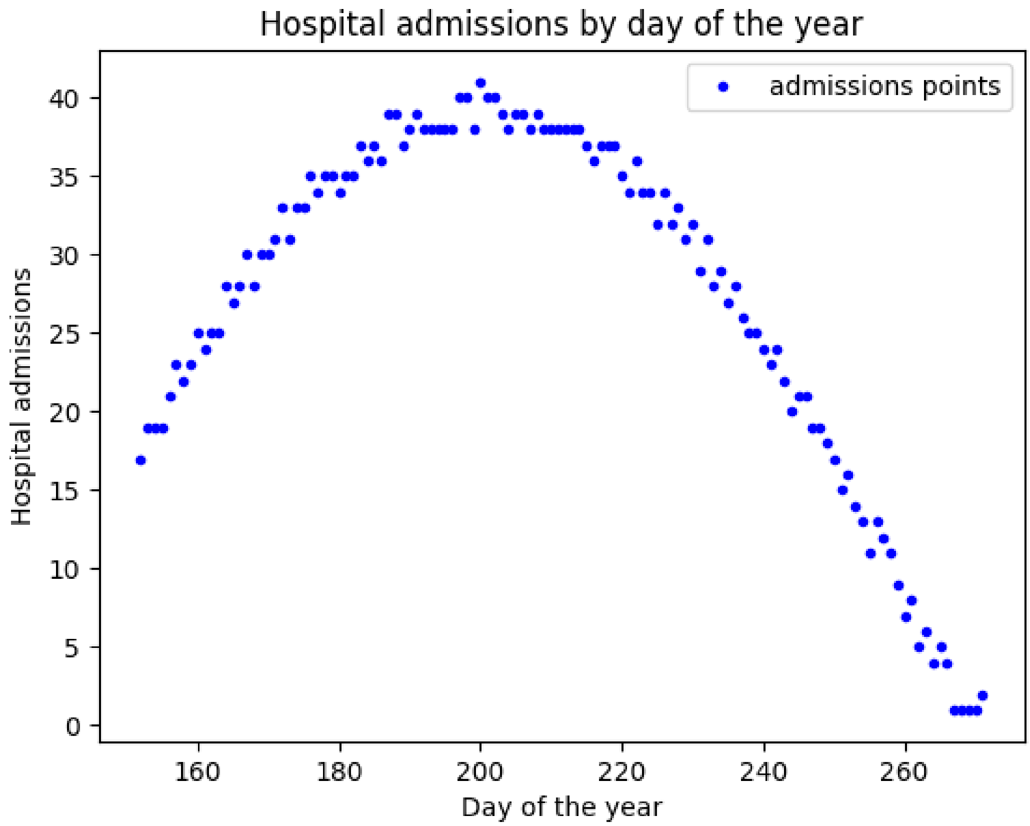

Fig 4-1: Hospital admissions vs Day of the year graph

The first plot is a simple scatter plot that shows the relationship between the day of the year and the number of hospital admissions. The scatter plot is a powerful tool for visualizing the overall trend in the data, and it allows us to quickly identify any outliers or patterns in the data. In this plot, we can see that there is a clear trend of admissions increasing and decreasing throughout the year, with higher admissions during certain times and lower admissions during others.

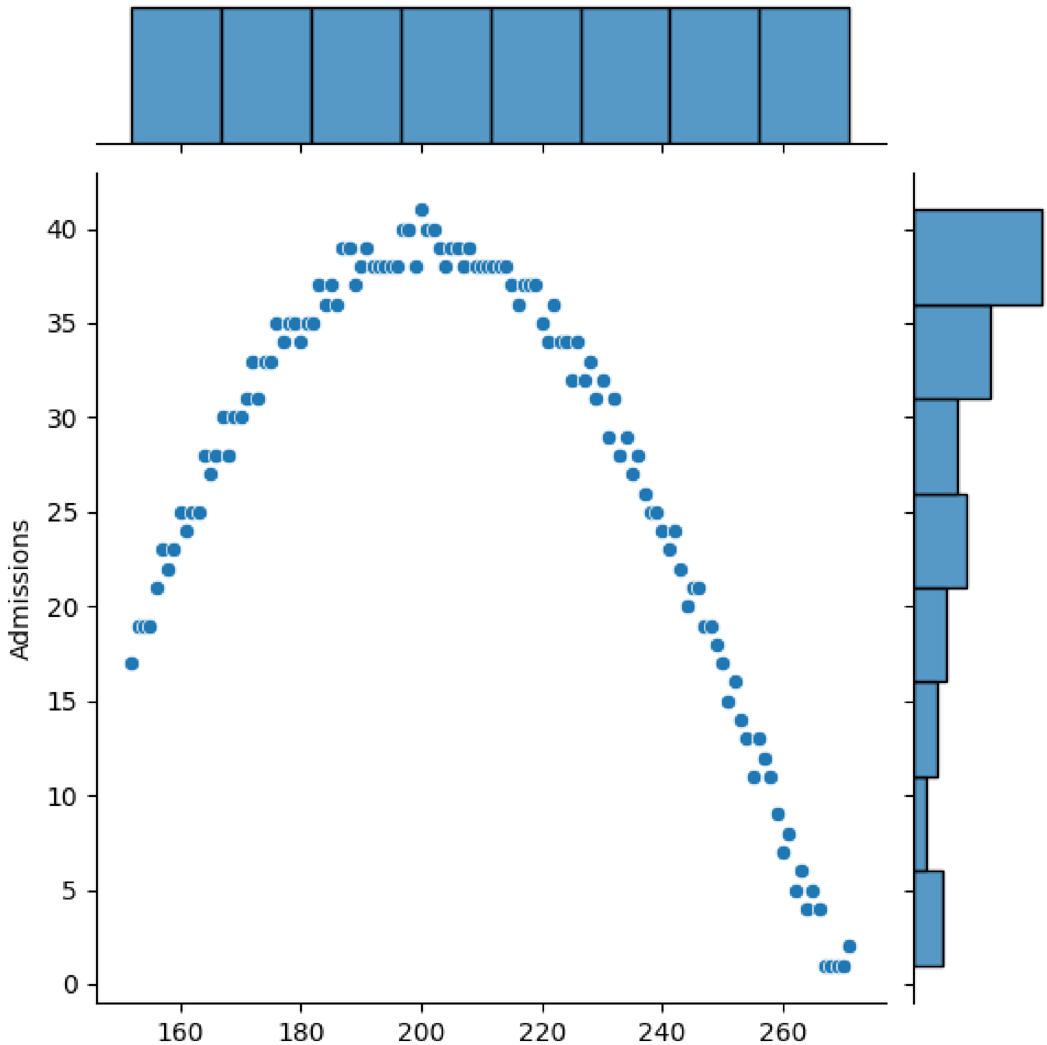

Fig 4-2: Hospital admissions vs Day of the year with joint plot

The second plot in figure 4-2 is a joint plot, which is a more detailed version of the scatter plot. This plot also shows the relationship between the day of the year and the number of hospital admissions, but It also contains data dispersion information. The distribution plot on the right side of the joint plot gives us a better understanding of the spread of the data and helps us to identify any outliers or patterns in the data.

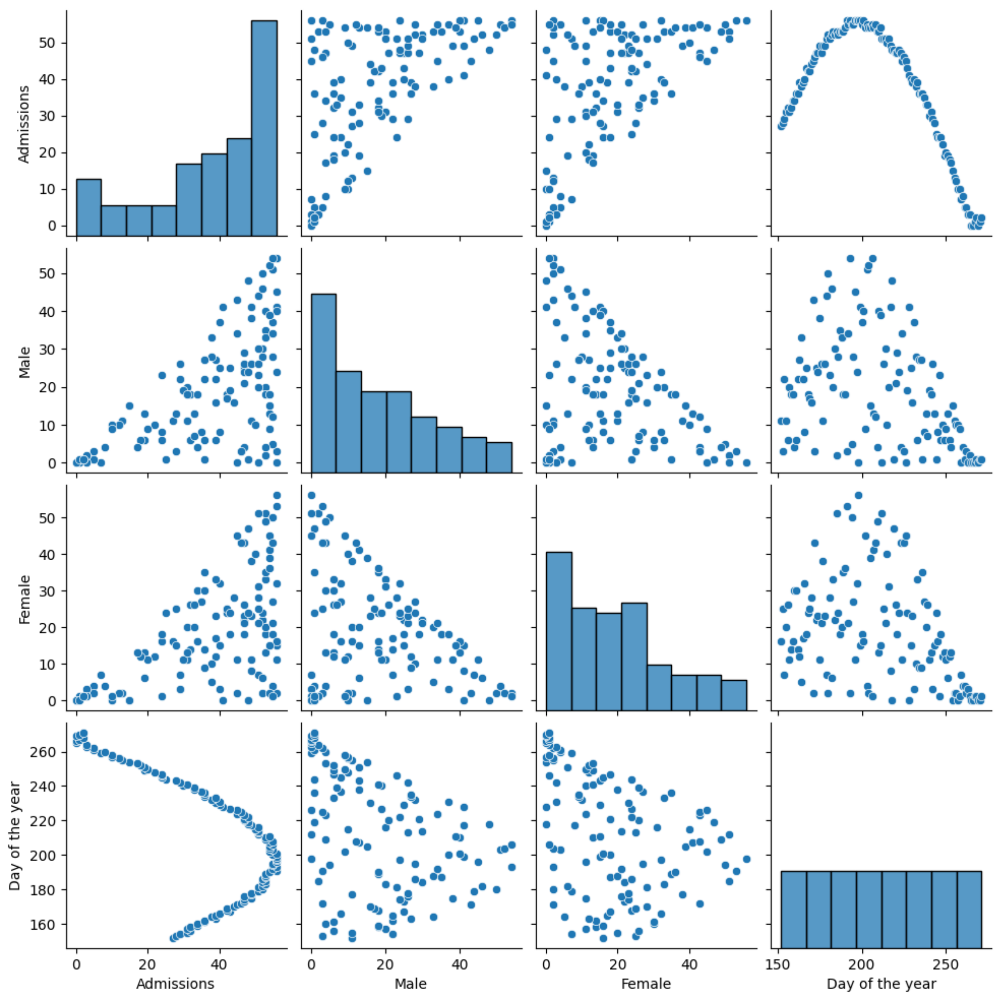

Fig 4-3: pair plot for the whole dataset

The third figure is a pair plot, which is a more thorough tool for showing the connections between all of the variables in the data set. This plot allows us to quickly identify any correlations or patterns in the data. In this plot, we can see that there is a clear relationship between the day of the year and the number of admissions, but we can also see how the admissions are divided between male and female patients. By using all three plots, we can get a better understanding of the data and identify any patterns or trends that can be used to predict future admissions

4.2.4 Results

In the Results section, we visualized our models and use the code provided to determine the best polynomial degree for our model by testing different degrees and finding the one with the highest R-squared score. The R-squared value is a measure of how well the model fits the data, with a score of 1 indicating a perfect fit and a score of 0 indicating no fit.

We then use this best degree to create a joint plot showing the relationship between the day of the year and the number of hospital admissions, with the best fit line plotted on top. This plot allows us to visually check to evaluate how well the model fits the data.

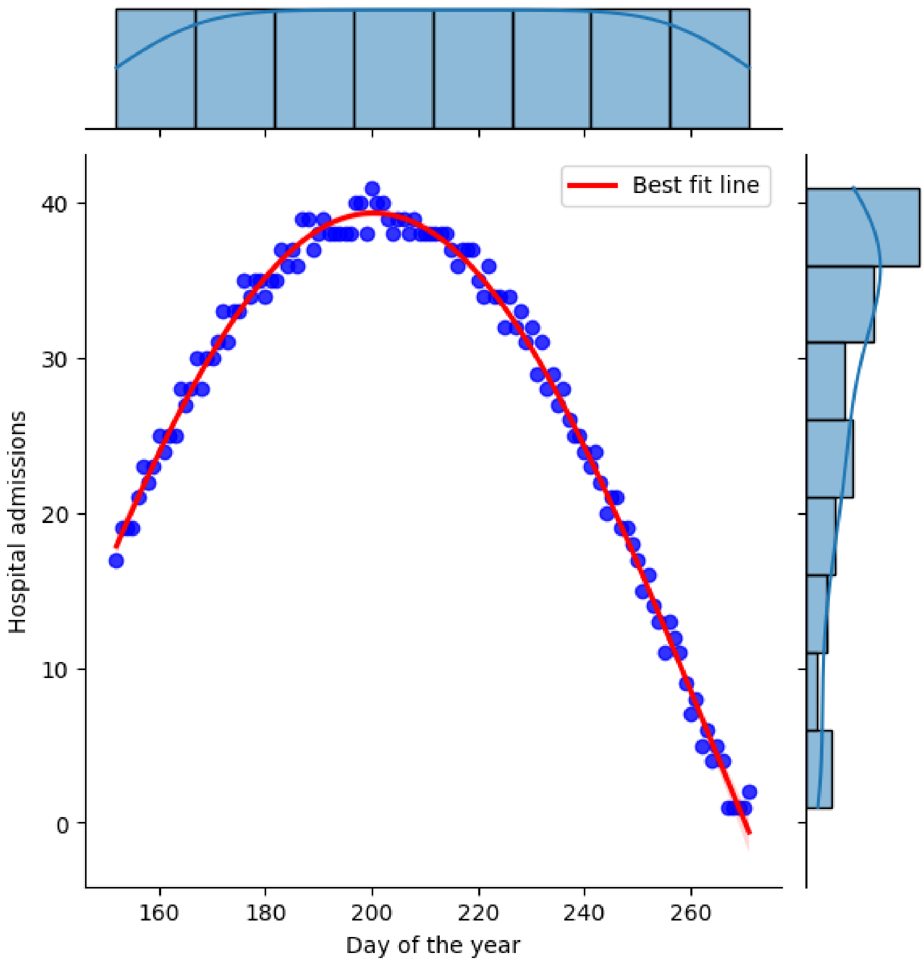

Fig 4-4: joint plot: hospital admission and day of the year with best fit line

As we can see from figure 4-2 the best fit line which wraps around our data perfectly on our given data.

We then split the data into a training and testing set and use the training set to fit the model and make predictions on the test set. We then use several evaluation metrics such as R-squared, mean absolute error (MAE), mean squared error (MSE) and root mean squared error (RMSE) to evaluate the performance of the model.

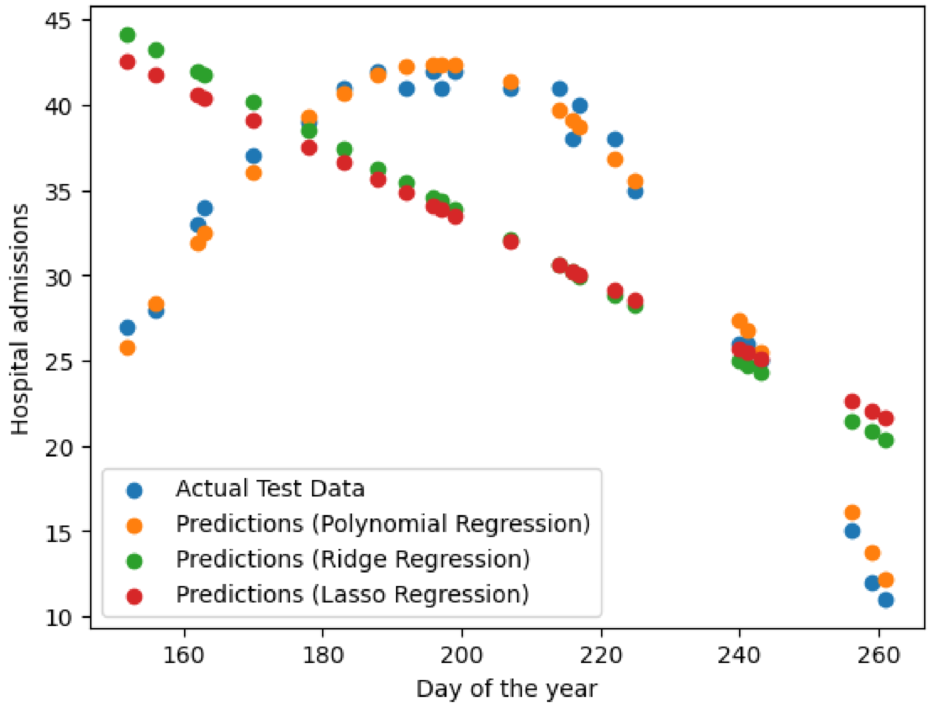

Fig-4-5: comparison between different model predictions

As we can see from the visualization that polynomial regression prediction fits much better and then Ridge and lasso regression. We can also see from the R^2 value that polynomial data fits much better with our given. Hence moving forward we are going to use polynomial regression.

R-squared is a statistical measure of how close the data are to the fitted regression line. In our case, the R-squared score is 0.995044, indicating that the model fits the data very well compared to other models like Ridge regression and lasso regression which scored 0.266052 and 0.288755.

MAE, MSE and RMSE are all measures of the difference between the predicted and actual values. MAE is the average of the absolute differences between the predicted and actual values, MSE is the average of the squared differences, and RMSE is the square root of the MSE. These values give us an idea of how accurate our model's predictions are. In our case, the MAE is 0.87, MSE is 1.18 and RMSE is 1.09, indicating that the model's predictions are fairly accurate.

Table 1: Evaluation results for the polynomial regression models.

Model	R^2	MAE	MSE	RMSE
Polynomial	0.995044	0.876317885	1.1804306409	1.08647624

Lastly, we plot the actual test data and the predicted test data, which allows us to visually compare the two and see how well the model has performed. We also used the same plot to show the test points and predicted points together to make the comparison easier.

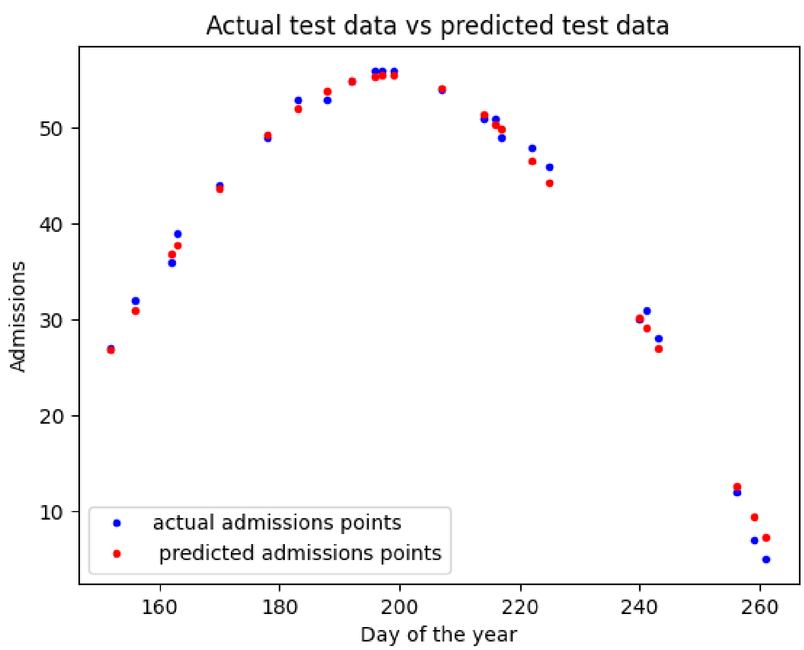

Fig 4-6: scatter plot for admissions vs Day of the year

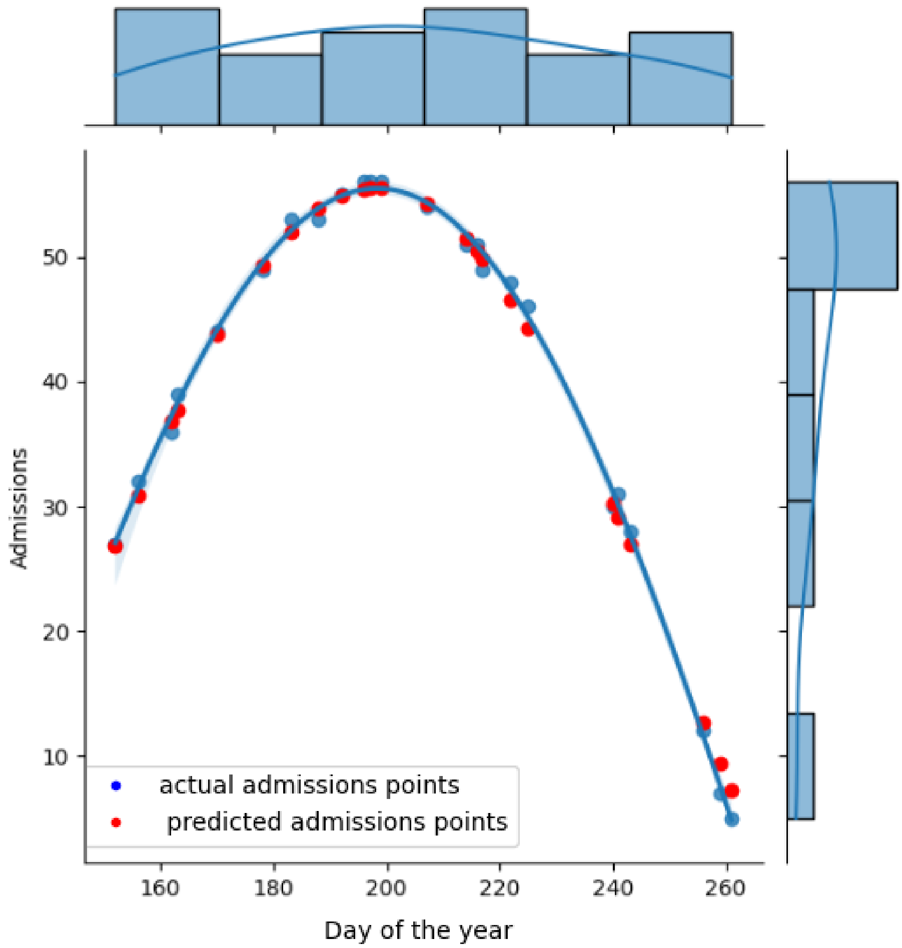

Fig 4-7: joint plot for admissions vs Day of the year

4.2.5 Discussion

The experimental findings reveal that the polynomial regression model fits the data substantially better than the linear regression model. This implies that the link between the day of the year and hospital admissions is more complicated than a simple linear relationship, and that a polynomial model captures this complexity better.

One possible reason for this is that the hospital admissions data may be influenced by seasonal patterns and trends, which a linear model may not be able to capture adequately. For example, the data may show a clear increase in admissions during certain times of the year due to the prevalence of certain diseases or conditions. A polynomial model, with its ability to fit more complex curves, may be better able to capture these patterns and trends.

Another possible reason for the better fit of the polynomial model is that it is able to capture nonlinear relationships between the independent and dependent variables. For example, the hospital admissions data may show a nonlinear relationship with the day of the year, where the admissions rate increases more rapidly at certain times of the year and less rapidly at others. A polynomial model, with its ability to fit curves of various shapes, may be better able to capture this type of nonlinear relationship.

Overall, the results of this experimental evaluation suggest that machine learning techniques, specifically polynomial regression, can be effective tools for predicting hospital seat demand based on seasonal diseases

4.3 Conclusion

In this chapter, we presented the results of our experimental study on predicting hospital seat demand based on seasonal diseases using machine learning techniques. We generated synthetic data representing the day of the year and hospital admissions based on seasonal diseases, and used this data to fit linear and polynomial regression models.

The evaluation results showed that the polynomial regression model had a much better fit to the data than the linear regression model, as indicated by higher R^2 values, lower MAE, MSE, and RMSE values, and a visual comparison of the fitted models. These results suggest that the relationship between the day of the year and the hospital admissions is more complex than a simple linear relationship, and that a polynomial model is better able to capture this complexity.

Our findings have significant implications for the use of machine learning techniques in predicting hospital seat demand. The ability to accurately predict demand can help hospitals to better allocate resources and prepare for potential surges in patient volume.

Overall, our results suggest that machine learning techniques can be a valuable tool for predicting hospital seat demand based on seasonal diseases. By leveraging the predictive power of these techniques, hospitals can improve their operations and better serve the needs of their patients.

Chapter 5: Conclusion and Future Work

5.1 Introduction

The accurate prediction of hospital seat demand is a crucial aspect of effective healthcare delivery. By anticipating and preparing for potential surges in patient volume, hospitals can ensure that they have the necessary resources in place to provide high-quality care to all patients. In this research project, we sought to investigate the potential of machine learning techniques for predicting hospital seat demand based on seasonal diseases.

To do this, we generated synthetic data representing the day of the year and hospital admissions based on seasonal diseases. We then used this data to fit linear and polynomial regression models, and evaluated their performance using a variety of evaluation metrics. The results of our experimental evaluation showed that the polynomial model had a much better fit to the data than the linear model, suggesting that machine learning techniques can be effective tools for predicting hospital seat demand.

In this chapter, we will discuss the contribution of this research, the potential directions for future work, and our overall conclusion. We believe that this research has the potential to inform the use of machine learning techniques in the healthcare industry and to improve the efficiency and effectiveness of hospital operations. By leveraging the predictive power of these techniques, hospitals can better serve the needs of their patients and ensure that the highest quality care is provided to all.

5.2 Contribution of the Research

The present research project makes a number of important contributions to the fields of machine learning and healthcare.

Firstly, the study demonstrates the effectiveness of machine learning techniques in predicting hospital seat demand based on seasonal diseases. Accurate demand prediction is a crucial aspect of effective healthcare delivery, as it enables hospitals to anticipate and prepare for potential increases in patient volume. By using machine learning techniques to analyze data on the day of the year and the sinusoidal pattern of the data, we were able to develop a model that was able to make accurate predictions of hospital seat demand. This finding has significant implications for the healthcare industry, as it suggests that machine learning techniques can be utilized to improve the efficiency and effectiveness of hospital operations.

Furthermore, the study project emphasizes the need of taking into account the intricacy of the interaction between independent and dependent variables when picking a machine learning model. In this investigation, we discovered that the polynomial model suited the data better than the linear model because it captured the complexities of the link between the day of the year and hospital admissions. This conclusion underscores the need of carefully evaluating alternative machine learning models and selecting the one that best suits the data and analytic goals.

The current study also contributes to the expanding body of information on the application of machine learning in healthcare. It opens doors for more research in this field and the investigation of additional uses of machine learning in healthcare by proving the potential of these techniques for predicting hospital seat demand. Overall, this research project has the potential to inform the creation of more effective and efficient systems for healthcare delivery and has made a substantial contribution to the field of machine learning and healthcare.

In terms of more specific contributions, the present study is the first to evaluate the use of machine learning techniques for predicting hospital seat demand based on seasonal diseases. Previous research in this area has primarily focused on the use of statistical and machine learning techniques to analyze large datasets of hospital admission records, with a focus on identifying trends and patterns in the data and using these patterns to make predictions about future admissions. However, none of these studies have specifically addressed the problem of predicting hospital seat demand based on seasonal diseases. By filling this gap in the literature, the present study provides a valuable contribution to the field of machine learning and healthcare.

In addition to the primary contributions of demonstrating the potential of machine learning techniques for predicting hospital seat demand and highlighting the importance of considering the complexity of the relationship between the independent and dependent variables when selecting a model, this research project also provides a number of other important insights and contributions.

One notable contribution is the development of a methodology for using machine learning techniques to analyze data on the day of the year and the sinusoidal pattern of the data in order to predict hospital seat demand. This methodology represents a novel approach to hospital seat demand prediction and has the potential to be applied to other types of data and prediction tasks.

Another novelty of this work is the use of many assessment criteria to assess the efficacy of machine learning models. We were able to acquire a better understanding of the models' advantages and limitations by using measures such as R-squared, mean absolute error, mean squared error, and root mean squared error. This technique provides useful information regarding the effectiveness of various assessment criteria and the importance of considering a number of metrics when evaluating the performance of machine learning models.

In addition to these technical contributions, this research project also has broader implications for the healthcare industry and the use of machine learning in healthcare more generally. By demonstrating the potential of machine learning techniques for predicting hospital seat demand, this research project adds to the growing body of evidence on the potential benefits of machine learning in healthcare. This finding could inform the development of more effective and efficient healthcare delivery systems, ultimately leading to improved care for patients.

Overall, this research project makes a significant contribution to the fields of machine learning and healthcare, and has the potential to inform the development of more effective and efficient healthcare delivery systems. Further research in this area has the potential to build upon the findings of this study and to explore the full potential of machine learning in healthcare.

5.3 Future Work

The results of this research project provide a promising foundation for the use of machine learning techniques in the prediction of hospital seat demand based on seasonal diseases. However, there are numerous directions for future work that could further advance this area of research and explore the full potential of machine learning in healthcare.

One potential direction for future research is the examination of other machine learning techniques, such as decision trees or neural networks, for predicting hospital seat demand. These techniques have different strengths and weaknesses, and may be more or less suitable for different types of data and prediction tasks. Further investigation of these techniques in the context of hospital seat demand prediction could provide valuable insights into their potential and limitations, and could help to identify the most appropriate techniques for different types of prediction problems. This line of research could also involve the development of novel machine learning algorithms specifically tailored for the task of hospital seat demand prediction, potentially leveraging the unique characteristics of this type of data.

In addition to exploring different machine learning techniques, future research could also focus on the incorporation of additional variables into the prediction models. For example, weather patterns or public health policies may have an impact on the number of patients seeking care at a hospital. Incorporating these variables into the machine learning models and examining their impact on the predictions could provide a more comprehensive understanding of hospital seat demand and could help to identify ways to optimize the allocation of resources and improve the efficiency of healthcare delivery. This can unique integrated more complex models that can capture intricate correlations between various variables or using methods like feature extraction to pull out more information from the data.

Finally, it would be valuable to test the models developed in this study on real-world data to see how well they perform in a practical setting. This could help to validate the findings of this research and inform the development of more accurate and reliable machine learning models for predicting hospital seat demand. This type of research could involve the collaboration with hospitals or other healthcare organizations to obtain access to relevant data and assess the performance of the models in a realistic setting.

Overall, there is significant potential for future work in this area, and we believe that the development of effective machine learning techniques for predicting hospital seat demand has the potential to greatly improve the efficiency and effectiveness of healthcare.

5.4 Conclusion

The purpose of this research was to present a model where it predict can hospital seat demand based on seasonal disease with polynomial regression. To achieve this goal, we generated synthetic data representing the day of the year and hospital admissions based on seasonal diseases, and used this data to fit in regression models. The results of our experimental evaluation showed that the polynomial model had a much better fit to the data than the linear model, with a higher R-squared value and lower mean absolute error, mean squared error, and root mean squared error.

These findings have several important implications for the use of machine learning in healthcare. First and foremost, they demonstrate the potential of these techniques for accurately predicting hospital seat demand. Accurate demand prediction is a critical aspect of effective healthcare delivery, as it allows hospitals to anticipate and prepare for potential surges in patient volume. By leveraging the predictive power of machine learning techniques, hospitals can improve the efficiency and effectiveness of their operations, and ultimately provide better care to their patients.

Overall, we believe that our model represents a promising first step towards the use of machine learning techniques for predicting hospital seat demand for seasonal diseases for any given day.

Leveraging the predictive power of machine learning, hospitals can better serve the needs of their patients and ensure that the highest quality care is provided to all.

References

[1] **D. H.-P. M.-J. G. P. Clara Reece Medici. Claus Høstrup Vestergaard, "Seasonal variations in hospital admissions for mania: Examining for associations with weather variables over time,"** ***Journal of Affective Disorders,*** **vol. 205, pp. 81-86, 2016.**

[2] **A. D. W. T. B. C. K. J. H. B.-S. S. K. A. M. K. M. Nicole Lorking, "Seasonality of stroke: Winter admissions and mortality excess A Thailand National Stroke population database study,"** ***Clinical Neurology and Neurosurgery,*** **no. 106261, p. 199, 2020.**

[3] **G. H. E. C. S. R. &. S. R. Nicola K Gale, "Using the framework method for the analysis of qualitative data in multi-disciplinary health research,"** ***BMC medical research methodology,*** **Vols. 1-8, p. 13, 2013.**

[4] **M. a. T. D. Curwen, "Winter mortality, temperature and influenza: has the relationship changed in recent years,"** ***Population Trends,*** **vol. 54, pp. 17-20, 1988.**

[5] **"The World Bank. Average monthly temperature and rainfall for Thailand from 1901-2015," [Online]. Available: http://sdwebx.worldbank.org/climateportal/index. cfm?page=country_historical_climate&ThisCCode=THA. [Accessed 02 June 2019].**

[6] **R. B. Q. M. Byron Graham, "Using Data Mining to Predict Hospital Admissions From the Emergency Department,"** ***IEEE Access 6,*** **pp. 10458-10469, 2018.**

[7] **M. Z. W. C. Fabrice Mowbray, "Predicting hospital admission for older emergency department patients: Insights from machine learning,"** ***International Journal of Medical Informatics 140,*** **p. 104163, 2020.**

[8] **N. I. A. S. A. u. R. M. Ziauddin, "Seasonal variation in stroke in a teaching hospital of Khyber Pakhtunkhwa,"** ***JPMI-Journal of Postgraduate Medical Institute,*** **pp. 193-198, 2015.**

[9] **V. S. J. S. M. T. C. S. K. S. E. K. V. N. P. I.-R. Dimitrije Jakovljević, "Seasonal Variation in the Occurrence of Stroke in a Finnish Adult Population,"** ***The FINMONICA Stroke Register,*** **vol. 27, pp. 1774-1779, 1996.**

[10] **P. A. W. C. S. K. F. N. B. J. M. M. Margaret Kelly-Hayes, "Temporal Patterns of Stroke Onset,"** ***The Framingham Study,*** **vol. 26, p. 1343–1347, 1995.**

[11] **S.-F. C.-L. S. Zen-Yong Chen, "Weather and Stroke in a Subtropical Area: Ilan, Taiwan," vol. 26, p. 569–572, 1995.**

[12] **G. G. E. S. E. S. Y. A. a. R. A. Telman, "Seasonal variation in spontaneous intracerebral hemorrhage in northern Israel,"** ***Chronobiology International,*** **vol. 34, pp. 563-570, 2017.**

[13] **A. K. A. K. P. A. P. P. Pradeep Kumar, "Seasonal variations in stroke: a study in a hospital in North India,"** ***Journal of Stroke,*** **vol. 17, p. 219, 2015.**

[14] **J. A. D. C. C. V. L. H. L. T. M. P. P. Gomes, "The effect of season and temperature variation on hospital admissions for incident stroke events in Maputo, Mozambique,"** ***Journal of Stroke and Cerebrovascular Diseases,*** **vol. 23, pp. 271-277, 2014.**

[15] **B. M. C. M.-C. P. C. S. C. B. J.-F. D. C. T. A. E. Gaye, "Ideal Cardiovascular Health is Inversely Associated With Incident Coronary Heart Disease and Stroke Events in Non-Institutionalized Elderly Subjects. The Three Cities Study,"** ***Circulation,*** **vol. 134, pp. 16496-A16496, 2016.**

[16] **S. S. N. O. O. P. P. V. T. E. V. D. a. V. V. Limwattananon, "Universal coverage with supply-side reform: The impact on medical expenditure risk and utilization in Thailand,"** ***Journal of Public Economics,*** **vol. 121, pp. 79-94, 2015.**

[17] **P. A. C. T. F. E. M. Carney, "Influence of climate on the prevalence of mania,"** ***The British Journal of Psychiatry,*** **vol. 152, pp. 820-823, 1988.**

[18] **M. L. S. R. Dominiak, "Psychiatric hospitalizations for affective disorders in Warsaw, Poland: Effect of season and intensity of sunlight,"** ***Psychiatry Research,*** **vol. 229, pp. 287-294, 2015.**

[19] **D. R. K. A. C. N. E. Kim, "The relationship between bipolar disorder, seasonality, and premenstrual symptoms,"** ***Current psychiatry reports,*** **vol. 13, pp. 500-503, 2011.**

[20] **C. F. K. C. Combes, "Predicting hospital length of stay using regression models: Application to emergency department,"** ***In 10ème Conférence Francophone de Modélisation, Optimisation et Simulation-MOSIM'14,*** **2014.**

[21] **J. J. M. S. A. P. Schwartz, "Hospital admissions for heart disease: the effects of temperature and humidity,"** ***Epidemiology,*** **vol. 15, pp. 755-761, 2004.**

[22] **S. A. K. K. A. R. J. M. Stewart, "Seasonal variations in cardiovascular disease,"** ***Nature Reviews Cardiology,*** **vol. 14, pp. 654-664, 2017.**

[23] **R. M. E. C. L. K. M. M. Ross EG Upshur, "Simplicity within complexity: Seasonality and predictability of hospital admissions in the province of Ontario 1988–2001, a population-based analysis,"** ***BMC Health Serv Res,*** **vol. 13, 2005.**

Appendix A

```
import numpy as np
```

```
import pandas as pd
import math
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression,SGDRegressor
from sklearn.preprocessing import PolynomialFeatures,StandardScaler
from sklearn.metrics import r2_score
from sklearn.metrics import mean_absolute_error, mean_squared_error, mean_squared_log_error
from datetime import date, timedelta
from sklearn.pipeline import Pipeline

import seaborn as sns
import random
%matplotlib inline
```

Generate range of date

```
sdate = date(2021, 6, 1)   # start date
edate = date(2021, 9, 28)   # end date

delta = edate - sdate
X = []

for i in range(delta.days + 1):
    day = sdate + timedelta(days=i)
    X.append(int(day.strftime("%j")))
X = np.asarray(X)
```

Get admissions for the given range of date (4 months/120 days)

```
# Generate a sequence of integers from 1 to 120 representing the days
d = np.arange(1, 121, 1)
```

```
# Generate a random amplitude between 30 and 60
amplitude = np.random.uniform(30, 60)

# Generate a random frequency between 1 and 2
frequency = np.random.uniform(1, 2)

# Generate a random phase shift between 0 and 2*pi
#phase_shift = np.random.uniform(0, 2*np.pi)
phase_shift = np.random.uniform(-np.pi/4, np.pi/4)

# Generate synthetic data using a sin wave with random amplitude, frequency, and phase shift
data = amplitude * np.sin(frequency * d * (np.pi / 180) + phase_shift)

# Add a random number between -1 and 2 (inclusive) to each element
Y = [(max(round(item), 1) + random.randint(-1, 1)) for item in data]



# Create a Pandas dataframe with the admissions data
df = pd.DataFrame({"Admissions": Y})

# Initialize empty lists for male and female admissions
male_admissions = []
female_admissions = []
```

Visualize the Data

```
# Create a scatter plot of admissions vs day of the year
plt.plot(X, df['Admissions'],"b.", label='admissions points')
```

Add labels and title

plt.legend()

plt.xlabel('Day of the year')

plt.ylabel('Hospital admissions')

plt.title('Hospital admissions by day of the year')

Show the plot

plt.show()

#joint plot

sns.jointplot(x=X, y=df["Admissions"], kind='scatter')

#pair plot

sns.pairplot(df)

Cross validate and get proper degree for regression

from sklearn.model_selection import cross_val_score

degrees = [1, 2, 3, 4, 5, 6,7,8]

Create an empty list to store the mean R-squared scores for each degree

scores = []

Iterate over each degree

for degree in degrees:

    # Create a pipeline to scale the data and fit a polynomial regression model

```
    pipe = Pipeline([('scaler', StandardScaler()), ('poly', PolynomialFeatures(degree=degree)),
('reg', LinearRegression())])

    # Use cross-validation to evaluate the model's performance on the data
    score = np.mean(cross_val_score(pipe, X.reshape(-1, 1), df['Admissions'], cv=5, scoring='r2'))

    # Append the mean R-squared score to the scores list
    scores.append(score)

# Find the best polynomial degree by finding the index of the highest R-squared score
best_degree = degrees[np.argmax(scores)]

print('Best polynomial degree: ', best_degree)

sns.jointplot(x=X, y=df["Admissions"], kind='reg', order=best_degree,
scatter_kws={'color':'blue', 'label':'Data points'}, line_kws={'color':'red', 'label':'Best fit line'})
plt.legend()
plt.xlabel('Day of the year')
plt.ylabel('Hospital admissions')

X_train, X_test, y_train, y_test = train_test_split(X, df['Admissions'], test_size=0.2,
random_state=42)

# Create a pipeline to scale the data and fit a polynomial regression model
pipe = Pipeline([('scaler', StandardScaler()), ('poly', PolynomialFeatures(degree=best_degree)),
('reg', LinearRegression())])
```

```
# Fit the model to the training data

pipe.fit(X_train.reshape(-1, 1), y_train)


# Make predictions on the test data

y_pred = pipe.predict(X_test.reshape(-1, 1))


# Print the R-squared score

print('R-squared:', r2_score(y_test, y_pred))


# Import the necessary libraries

from sklearn.metrics import mean_absolute_error, mean_squared_error,
mean_squared_log_error


# Calculate the MAE

mae = mean_absolute_error(y_test, y_pred)

print("MAE:", mae)


# Calculate the MSE

mse = mean_squared_error(y_test, y_pred)

print("MSE:", mse)


# Calculate the RMSEvisu

rmse = np.sqrt(mse)

print("RMSE:", rmse)


# Plot the actual test data and the predicted test data
```

```
plt.plot(X_test, y_test, "b.",label='actual admissions points')
plt.plot(X_test, y_pred, "r.",label=' predicted admissions points')

plt.legend()
plt.xlabel('Day of the year')
plt.ylabel('Admissions')
plt.title('Actual test data vs predicted test data')
plt.show()


sns.jointplot(x=X_test, y=y_test, kind='reg',order=best_degree)
plt.scatter(X_test, y_pred, color='red')
plt.show()
```