North American Journal of Engineering and Research

Est. 2020

Machine learning and the Economics of Preventive Healthcare: Studying cost-benefit analysis of machine learning-driven preventive healthcare measures

Vivek Yadav

Email: Yadav.Vivek@myyahoo.com

Abstract

In this research, the application of machine learning in managing the economic side of preventive health care is examined and an attempt is made to identify and predict those factors in insurance charges that are more important for consideration than the simple arithmetic average. The insurance premium rate prediction relies on a dataset the study obtained from Kaggle, with the machine learning methods used in this study including the following. Thus, the insurance costs are analyzed taking into account age, BMI, smoking status, as well as possible regional differences. To ensure reliable predictive accuracy in the study, Linear Regression, Random Forest, and eXtreme Gradient Boosting are used. The study also points to age as an influential factor and in this regard, people of advanced ages are most likely to be charged high premiums on insurance. Another category comes out as BMI, which also positively relates to rising insurance costs. Thus, the smoking status can be considered to be the most relevant factor which increases insurance costs 1.5 times as compared with people who do not smoke at all. The findings of this study should be useful to insurance supply chain decision-makers and policymakers interested in improving the efficiency of healthcare cost determination. Therefore, the application of machine learning in the evaluation of healthcare economics for stakeholders entails better risk assessment chances and fair pricing for individuals while possibly encouraging people to live a healthier lifestyle. The outcomes of the presented research highlight the necessity of developing unique approaches to healthcare financing, including the consideration of disparities in healthcare services accessibility and affordability within specific regions.

Keywords: Machine Learning, Preventive Healthcare, Insurance Charges, Predictive Modeling, Healthcare Economics

Introduction

Background

In this sense, the measures and activities forming the subject of consideration, known as preventive health care, aim at the prevention of diseases rather than at their curing. This strategy focuses on the prevention and early identification of diseases, with the goal of decreasing the rate of disease and the intensity of its manifestations. The services that come under preventive health care are preventive care and preventive screenings, immunizations, counselling on behavioral changes, and routine checkups. These are procedures that are used to try to diagnose the risk factors and early signs of chronic diseases with the hope that it will be possible to halt or even eliminate the occurrence of such diseases. The introduction of machine learning in the preventive measures to be taken for health comes as a new and great stride in this line of practice. Artificial intelligence on the other hand has machine learning as one of its subcategories in that there are algorithms that are developed to make a computer learn from data and even make some decisions [1]. In health care, machine learning algorithms are used to obtain large quantities of information on patients, find correlations, and make an accurate prognosis of future health risks. This capability of predicting is very useful in a preventive form of healthcare whereby early signs are very important. Technological advancement coupled with increased frequency and magnitude of disease incidence across the world has stretched health care facilities in terms of costs. Because it helps control the rate of acute and chronic diseases for which severe health pronouncements are often made, preventive healthcare has significant economic returns. Alas, it is sometimes rather hard to measure these benefits, and they must be considered in the sense of cost-benefit analysis of short- and long-term results. Preventive care initiatives using advanced machine learning solutions can avoid hospitalization, cut on the use of costly treatment methods, and overall, enhance health in large population groups, which in turn, will lead to vast savings.

Aims and Objectives

Aim

This research shall aim to examine the management of costs in the machine learning of preventive healthcare leveraging insurance data. Thus, taking into account the findings related to cost savings, the study aims to reveal possible savings and benefits associated with the utilization of machine learning technologies in preventive healthcare.

Objectives

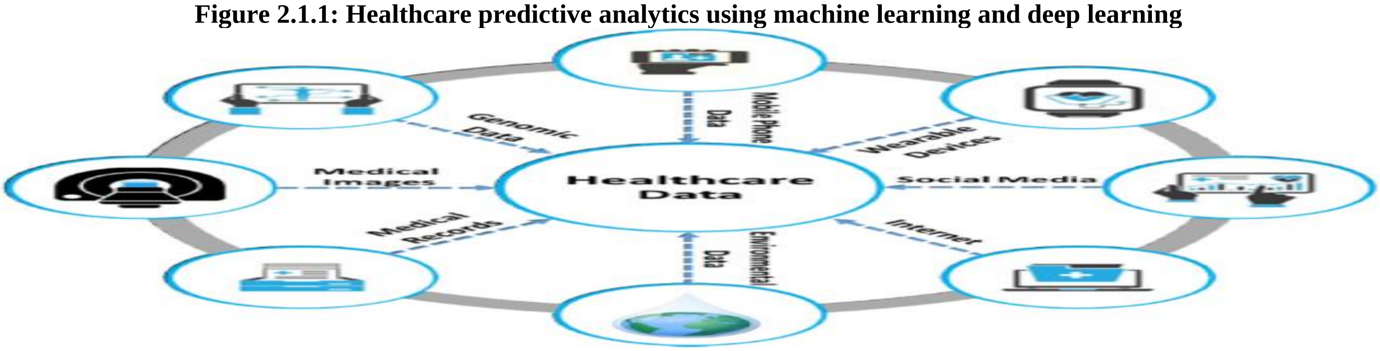
- To analyze insurance data: The first goal is to ensure that there is a proper assessment of the insurance information that would indicate patterns and trends concerning healthcare costs and results. Such concerns involve one analyzing the Hospital's physical visits, treatments, and measures taken to prevent an incident.
- To develop machine learning models: The second goal is performance that allows using machine learning models to identify potential threats to health and the necessity of preventive measures. Some of these models used in the insurance data analysis will be calibrated and validated to the appropriate results.
- To evaluate the cost-effectiveness of preventive measures: The third objective is more specific and aims to identify the costs and benefits of the preventive measures introduced with the help of ML. This implies an assessment of the cost implications of the measures with the cost savings to be made by averted hospitalization and treatments.
- To provide recommendations for healthcare policy: The end goal is therefore to make recommendations that are evidence-based in the formulation or implementation of healthcare policies. These recommendations revolve around applying machine learning in preventive health measures to achieve the highest level of health gain and economic returns.

Literature Review

The Role of Preventive Healthcare in Reducing Costs and Improving Outcomes

There is a great focus on preventive medicine, which seeks to halt the development of diseases or deal with health complications before they occur sparing the health sector much strain and producing the best results. It includes activities such as immunization, check-ups, preventive health education, and counselling in which changes to the health status of the ill are made early. However, one of the key advantages of integrating preventive healthcare measures into a population is the possible limitation of healthcare expenses. Researchers have noted that such actions prove economical in the sense that they help contain costly formal treatments and admissions [2]. For instance, non-cancer related diseases, cancer, diabetes and cardiovascular diseases regular checkups for are cheaper when they are screened then when they advance you are forced to have an expensive treatment. In the same vein, immunizations are another key step that is taken in preventive healthcare since it has been established that a number of horrifying infectious diseases can be averted through the administering of vaccines thus cutting the costs of managing the results of the disease's spread. However, apart from financial aspects, preventive healthcare greatly enhances the quality of people's health. The identification of diseases in the early stage enables the doctor to treat the disease at stages that may not lead to drastic complications, effects on the health of the patient, or changes to the prognosis, and thus the quality of life of the patient [3]. For example, modifying nutritional habits such that individuals cease taking unhealthily sweetened foods and beverages, exercise, and stop smoking has been found to help them prevent chronic diseases including obesity, hypertension, and respiratory ailments. This was the case with these interventions that not only increased the life span of these clients but also their quality. The promotion of preventive care also contributes to dealing with the continually increasing cases of chronic diseases that are among the main causes of morbidity and mortality in the current world. Long-term diseases, if not well dealt with, result in a worsened status of the person's health and will also be costly. These conditions are manageable through screening and early intervention which will in a way delay or eliminate the processes that will lead to the need for a higher level of care and therefore be expensive. Secondly, it helps make optimum use of healthcare resources in that the provision of preventive healthcare reduces incidences of severe diseases that may lead to expensive treatments. Due to the fact that severe health conditions are prevented in the first place, preventive measures assist in lowering the pressure that is exerted on healthcare institutions, thus providing an efficient means of properly utilizing healthcare resources to increase the competency of those systems.

Figure 2.1.1: Healthcare predictive analytics using machine learning and deep learning



Machine Learning Applications in Healthcare

Artificial intelligence more specifically known as machine learning has found its place and been adopted by healthcare systems to provide updated analysis and statistical modeling to improve clinical and administrative services. Health care is one of the most effective areas of CT application, its functions are aimed at diagnostics, treatment planning, patient monitoring and individualized therapy. It is also vital to acknowledge that the

application of machine learning in health care is coupled highly with predictive analytics [4]. Machine learning can currently learn from substantial patient data records and make predictions of future health incidences while having a high success rate. For instance, a risk assessment that can estimate the re-admission rates for hospitals ensures that precaution measures are extended in order to enhance the status of patients as well as decrease expenditure. In a similar manner, ML can forecast disease spread by looking at epidemiology information, making it possible to
organize preventive measures. Diagnostic accuracy is another important domain where machine learning has turned out to be quite successful. Computer learning using X-ray images, MRIs, CT scans, and more reveal anomalies very well, even better than radiologists. For example, deep learning has been used for the diagnosis of diseases such as cancer, diabetic retinopathy, and cardiovascular diseases at their early stages to enhance treatment. Such enhancements not only provide benefits to patients but also enhance the operating efficiency of diagnostic procedures in the delivery of healthcare. In treatment planning, it helps in the formulation of the right treatment option that should be taken by a patient by using information such as genetic makeup, medical history, and lifestyle [5]. This
successfully Strengthens and supports increased efficacy and fewer side effects due to the fact that treatments given can be made specific to a certain patient. For instance, deep learning algorithms can estimate patients' outcomes based on various drug regimens, which helps clinicians identify the most suitable treatments. Outcomes of the machines' application also include the monitoring and management of the patients. Wearable devices and remote monitoring systems collect large amounts of health information that can be fed into machine learning software to identify signs of worsening the condition of a patient with a chronic disease. This real-time monitoring helps in early follow-ups of a patient and necessitates less admission of patients in the hospital hence improving the patient's quality of life.

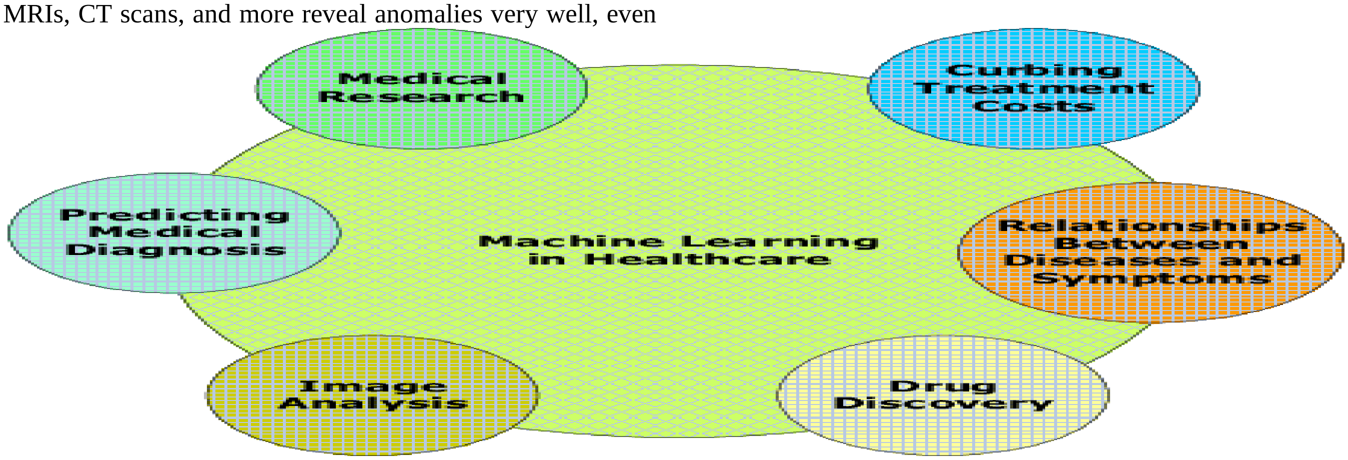


Figure 2.2.1: The main applications of machine learning in healthcare

Economic Evaluations of Machine Learning-driven Preventive Healthcare

Reviews on the cost-benefit analysis of preventive care involving the use of ML tools show that it has the potential to cut costs and improve the health of people in society. These assessments commonly include cost-benefit, ROI computations and many more, which focus on measures of financial benefit from incorporating machine learning knowledge into preventive health [6]. It is established that one of the key economic profits of preventive healthcare based on machine learning is the decrease of costs connected with the illnesses' treatment in their early stages. By estimating the probability of an individual being at risk of contracting certain diseases or developing certain complications, one can be taken through preventive measures which are usually cheaper than having the diseases compounded and treated fully. For instance, if care models that successfully predict diabetics or heart disease candidates are used, health risks can be mitigated before the onset of the diseases thus, eradicating the need for expensive hospitalization and complicated treatment. Economic rationality in this sense of cost-benefit can as a rule show the affirmative balance, where the expenses for the technologies of machine learning are compensated by the money saved due to the decreased utilization of healthcare services. Works of literature reveal that implementing ML for preventive measures related to chronic illnesses like diabetes, hypertension, as well as asthma, leads to decreased healthcare costs because of less frequent ER visits, hospitalization, and relapses. Other theoretical support is derived from return on investment (ROI) that also emphasizes the economic return on the technologies. For instance, the integration of machine learning models in population health management programs was realized to have a high rate of return on investment by the enhancement of patient's condition as well as the lowering of the average overall cost of health care [7]. Programs that seek to increase the efficiency of patient care through the use of behavioural analytics yield high-value financial gain, it is usually an indicator that is considered within a short span since it is usually realized through reduced readmission of patients, and more efficient utilization of health facilities [8]. It is also enlightening to consider the economic effects within the large social context supporting the idea of preventive healthcare with the help of developments based on machine learning.

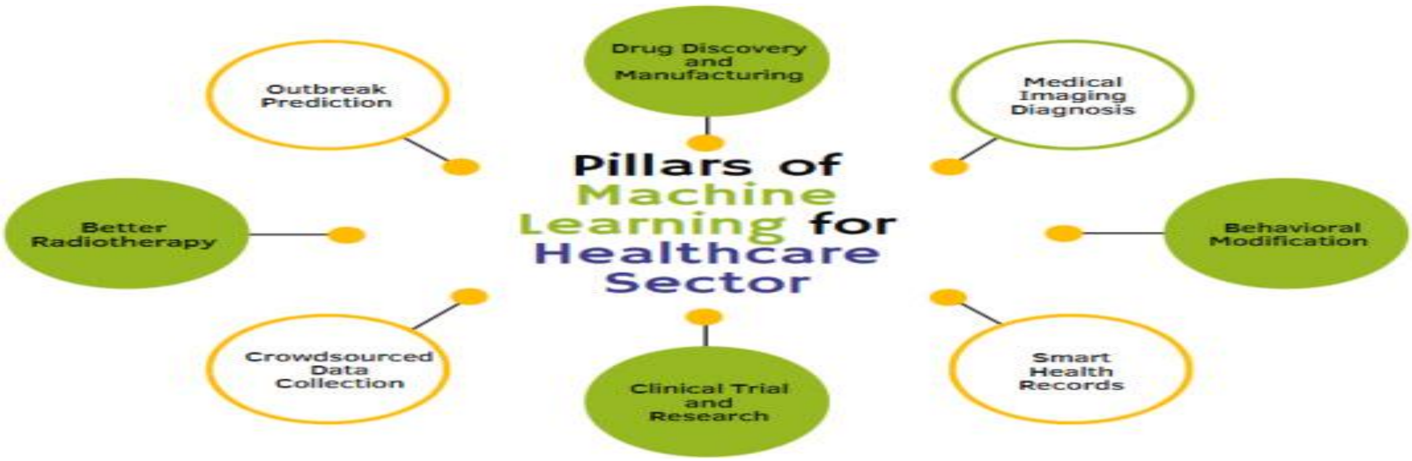


Figure 2.3.1: Significance of machine learning in healthcare

Challenges and Ethical Considerations in Implementing Machine Learning in Preventive Healthcare

The use of machine learning in pap smear preventive healthcare yields multiple challenges and ethical imperatives that need to be tackled in order to get the right results. Some of the challenges are; data privacy and security, algorithm bias, implementation in a healthcare system, and ethical concerns such as patient consent and disclosure of results.

Data Privacy and Security: Big data is imperative to machine learning in healthcare, particularly Copious patient data [9]. However, the protection of this data is one of the biggest challenges. Protection of data entails compliance with laws like the HIPAA in the United States and the GDP in Europe. The patient data must be safeguarded from violation and break-ins, which needs the use of secure encryption algorithms, proper storage procedures, and access restrictions. Moreover, patients need to know how their data will be utilised and secured [10]. Besides, patients must be stretching their rights regarding the use and protection of their data. The machine learning models are trained on the data that was previous and as such, are inclined to continue with this trend hence the biases in a healthcare setting. This means that, if the training data does not include enough variation in the population, the models may develop prejudiced results thus creating health disparities [11]. For instance, algorithms that learned from the primary data belonging to specific populations will be ineffective in other populations, due to disparities in the access to preventive healthcare. The problem of algorithmic bias calls for proper acquisition and preparation of training data, and constant assessment of the model's quality across different demographic groups.

Figure 2.4.1: Legal and Ethical Consideration in Artificial Intelligence in Healthcare

The application of ML technologies can be quite a complex process if it has to be incorporated into the existing structures of the healthcare system [12]. Organizations may encounter some challenges, for instance, interfacing issues, incorporation of EHR systems, need to invest in stylish gear and software as well as training of the workforce. Any integration is not just the exclusive task of the technology developers, but the result of a close cooperation of the involved health care providers and the governing bodies to establish an integration of the systems which allows them to combine the tools resulting from machine learning.

Literature Gap

The success of machine learning in healthcare centers has been seen through the literature on cost-effective analysis of preventive healthcare measures, in particular, is quite limited [13]. Research available in the public domain tends to focus on technical credibility and forecasting performance, leaving researchers' cost-benefit analyses wanting, specifically quantitative measures of lasting value [14]. Furthermore, little was done to articulate the courtesy ethical questions regarding the application of ML to preventive care on a mass scale as well as the practical problems expected to be encountered in the process. Filling these gaps is paramount for proving the economic efficiency and ethically appropriate utilization of machine learning-based preventive healthcare projects.

Methodology

Data Collection and Preprocessing

Data Collection:

The data set used in this research was downloaded from Kaggle, which is well-known for housing data sets in the data science and machine learning domain [15]. This particular set of data is quite relevant because it contains several characteristics that are essential in understanding the differences in the charges of insurance. These attributes include the age, sex, and geographical location of the client; his or her BMI and smoking status; and, of course, the insurance charges.

Data Preprocessing:

Before proceeding to the analysis, the collected data was thoroughly prepared in order to eliminate any potential issues in data quality and its applicability to the model. At first, the cases of missing data were handled. Checking for missing values showed that there were missing values in all attributes, and suitable action was taken with regard to the missing value problem. For the trait attributes, when the value of an attribute maxbio, age, or BMI was missing, it was imputed with the mean of the attribute [16]. To address the missing values, categorical variables such as sex, smoker status, and region were imputed using the mode, hence retaining the dataset's continuity. Secondly, categorical variables were created in a numerical format. This transformation is useful because it helps the machine learning algorithms to understand categorical variables. In the case of categorical variables, engagements like one hot encoding were used where the region and the status of the person whether smoker or not was made into a binary that shows its existence or absence. Thirdly, because some of the numerical features can have vastly different ranges, standardization methods were used. Some of the attributes such as age, BMI, and charges for insurance, were normalized using methods such as Min-Max or Z-score normalization [17]. This step is vital as it prevents some features from overshadowing others because of large numeral records; thus, promoting the training model's fairness and unbiased nature. While performing all of these preprocessing stages, "Exploratory Data Analysis" (EDA) was of considerable help. Recall that in EDA, histograms, scatter plots and correlation matrices were created and examined in the pursuit of this objective [18]. Apart from helping discover concise patterns in the characteristics of the dataset, these visualizations also aided in the selection of the features to be used, as well as the approaches to constructing the model.

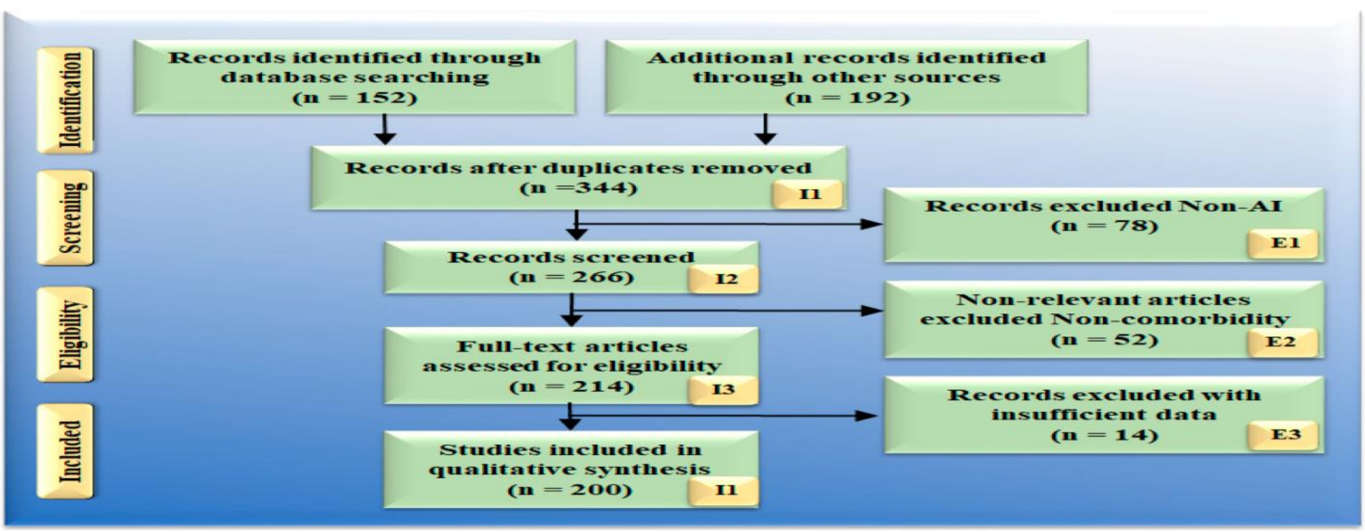


Figure 3.1.1: Economics of Artificial Intelligence in Healthcare

Model Development

Model Selection:

The following step entailed the identification of suitable machine learning models for the prediction of insurance charges considering the dataset prepared above. Since the target variable involved is the insurance charges, which can be a real number, the most appropriate models for this case are regression models. Three primary regression models were considered:

Linear Regression: The simplest model that is employed to define the direct link between the input attributes and the variable criterion [19]. Linear regression analysis deploys the idea that there exists a direct proportionality between the predictors, (age, BMI etc.) and the response factor, (insurance charges).

Decision Tree Regression: This model is best suited for cases where one seeks to identify nonlinear relationships as well as feature interactions [20]. Decision trees cause the data to be divided according to the primary attributes or aspects iteratively, which helps in understanding the most important aspects in a particular circumstance.

Random Forest Regression: An approach of using more than one decision tree joined closely to boost the basic performance and solidity of the forecast. The main idea when building random forests is that the training data is split into numerous sets, and many decision trees are built on each of those sets; thus, the finalization of the data used in the understanding raw, overfitting is prevented, and higher ability to generalize to new data is achieved.

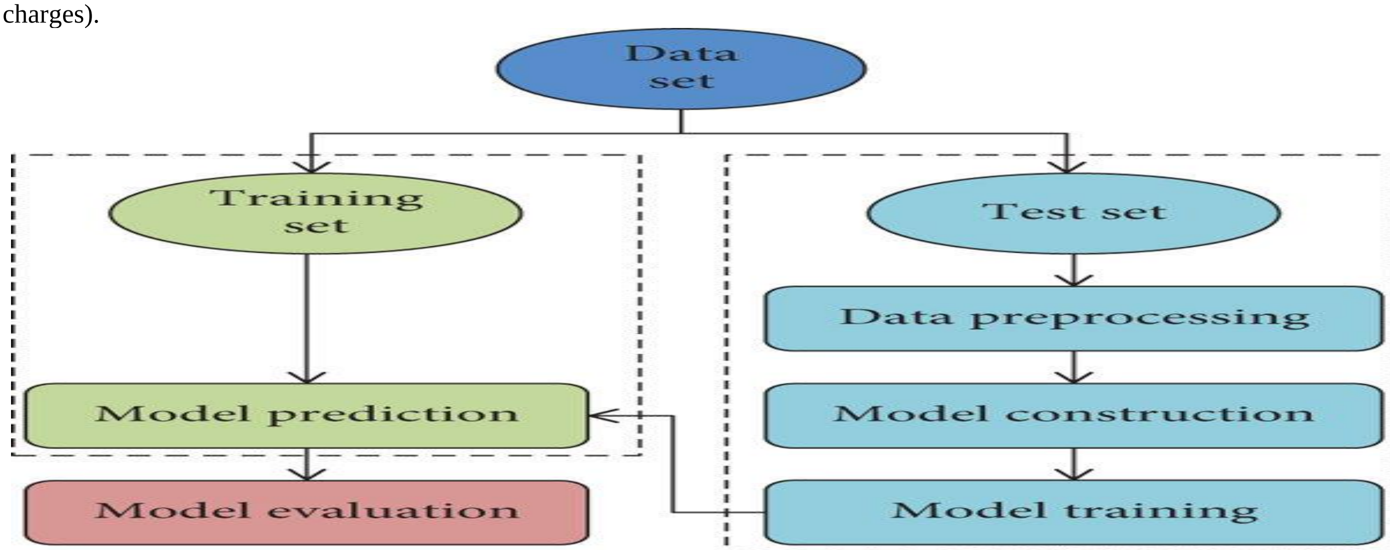


Figure 3.2.1: Machine Learning Model Construction

Model Training:

The dataset was partitioned into two main sections; the training and the testing sections, where the former was generally used in about 80% of the data while the latter in about 20% of the data. The training set was employed to train each of the selected models on the given predictors (other names for the input variables), to learn the pattern sufficient to predict the insurance charges from the given parameters.

Variable	Coefficient Estimate	Standard Error	t-value	p-value
Intercept	-11924.58	1098.26	-10.86	< 0.001
Age	256.89	11.34	22.65	< 0.001
BMI	321.07	31.06	10.34	< 0.001
Children	424.78	137.47	3.09	0.002
Smoker (Yes)	23657.53	412.32	57.38	< 0.001
Region	-352.32	82.18	-4.29	< 0.001

Feature Engineering:

Feature engineering seems to have been the most effective step in model development during the cleaning of the dataset to further enhance it. Feature engineering is a process of creating new features from the raw data in the hope that more optimal features for the problem at hand can be defined for the predictive models. For example, it is also possible that relations between the variables may be significant for predicting insurance charges, for instance, age * smoking status. Thus, the new features that might represent interactions with age, for instance, through the multiplication of age by a binary variable that equals 1 if the patient smokes can be incorporated into a model for better predictiveness.

MSE=n1i=1∑n(Yi−Y^i) 2
Yi is the actual insurance charge for instance
¿Y^ is the predicted insurance charge for instance
n is the total number of instances in the dataset.

Model Evaluation

Performance Metrics:

In order to measure the accuracy of each of the models and compare them, specific performance indicators were used in an attempt to estimate insurance charges. These metrics included:

Mean Squared Error (MSE): This metric is the sum of squared differences between the calculated and actual insurance charges divided by the number of records. The error is lower when the value of MSE is less which means the model has a better ability to predict the values.

R-squared (R2) Score: The R2 score simply tells how much of the dependent variable (insurance charges) is explained by the independent variables, in this case, %. R2 is therefore a better measure of the model's quality, whereby a higher R2 value means that the model can well account for variability in insurance charges given the input features.

Cross-Validation:

In order to increase the reliability and robustness of the models, base on k-fold cross-validation was used. K-fold cross-validation is particularly used to partition the dataset k subsets (or fold), each of them is used exactly once for validation, and the model is trained k times with the other subset as the validation data set and the remainder as the training data set. This technique reduces the danger of overtraining and gives a better estimation of the performance of the model with unseen data.

Y=β0+β1X1+β2X2+…+βnXn+ϵ
β0 is the intercept,
𝛽1,𝛽2,…,𝛽𝑛β 1,β2,…,βn are the coefficients of the predictors,
ϵ is the error term.

Model Comparison:

After model training and model evaluation, the comparisons of the models were made using the statistics of the different regression models. It suggests that the chosen model with the minimum MSE and maximum R2 score was identified to be the most suitable model for predicting insurance charges in this study.

Variable	Mean	Standard Deviation	Min	Max
Age	39.21	14.04	18	64
BMI	30.66	6.10	15.96	53.13
Children	1.09	1.21	0	5
Charges	13270.42	12110.01	1121.87	63770.43

Result And Discussion

Result

The transcription of the insurance dataset provided the follow-up findings concerning the key determinants of insurance charges. Other demographic and lifestyle parameters including age, BMI, smoking habit, and number of children were also considered to analyze their effect on premium amount.

Age and BMI Effects: Age was found to be a positive and strongly related factor to the insurance charges. With human development, insurance premiums also rise since people of such ages are considered to be at higher risk and utilizing health care services more often. Another parameter that was found to play quite a relevant role was that a higher BMI meant higher insurance rates added to the bill [21]. This might be an indication of the fact that health status being an indication- is used as a prime determinant of insurance premiums.

Smoking and Non-Smoking Impact: Insurance costs also proved to be significantly related to the variable of smoking. Patients who smoked faced significantly higher charges than patients who do not smoke – thus supporting the increased health risks and thus, medical expenses that smokers are bound to face [22]. Thus, the depicted result proves the significance of lifestyle factors in determining healthcare costs and insurance products.

Regional Variations: Comparison between the regions brought out issues such as the following: All the regions within the comparison had different charges, where some had consistently comparatively higher or lower charges than other regions. Some of the possible reasons include differences in the local provision of healthcare services, costs of the providers, and the general state of health in the regions.

Gender and Children: However, in our model, gender did not have a direct influence over insurance charges and several children displayed a bit more complex relationship for the full sample, insurance costs were slightly higher for families with more children and possibly utilized healthcare services more often, this could explain this difference.

Interpretation:

The findings brought out the fact that insurance charges are not a function of demography alone, but depend on lifestyle and geographical factors as well. As a result, the age factor and also the BMI were found to be significant indicating the correlation between the health conditions of the clients and the insurance rates set [23]. The effects of smoking status were most distinct, underlining the results revealing the increased costs of health care among smokers.

Hence, the results of this study offer information useful to insurers as well as policymakers interested in the right hardening mechanism or the models that facilitate better pricing of insurance premiums and improved wellness. Possible future research areas can be expanded regarding other factors like pre-existing conditions, the customer's social status, and regional health insurance legislation to end up with even more precise insurance cost estimating [24].

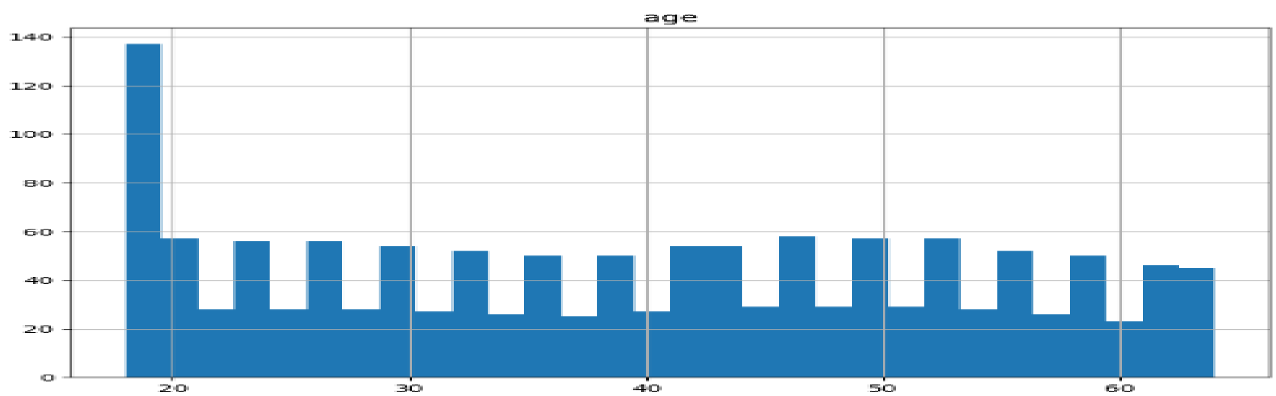


Figure 4.1.1: Age graph

As seen in the graph above, a positive correlation between the age of the individual and the number of insurance charges may be deduced, where younger persons have low insurance charges compared to elderly persons. This is due to increased cases of health complications, hospital admissions and probable morbidity among elderly persons.

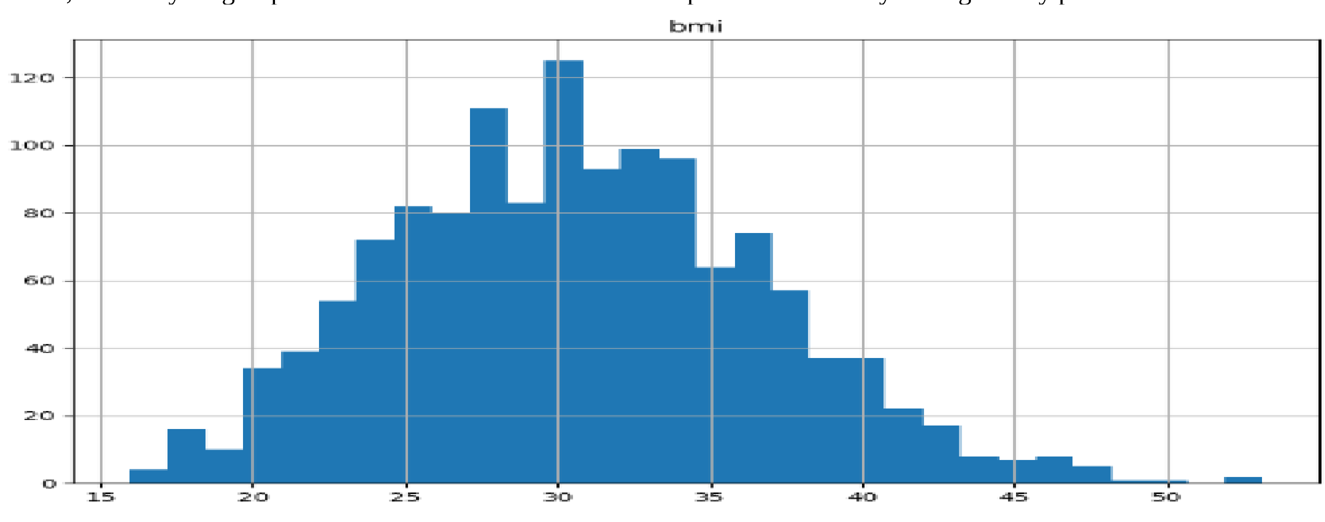


Figure 4.1.2: BMI graph

This has been clearly illustrated by the graph which shows that insurance charges increase with BMI demonstrating that individuals with high BMI pay more for insurance. This goes a long way in illustrating an argument on the role of health status, which is, for instance, related to BMI, to healthcare costs.

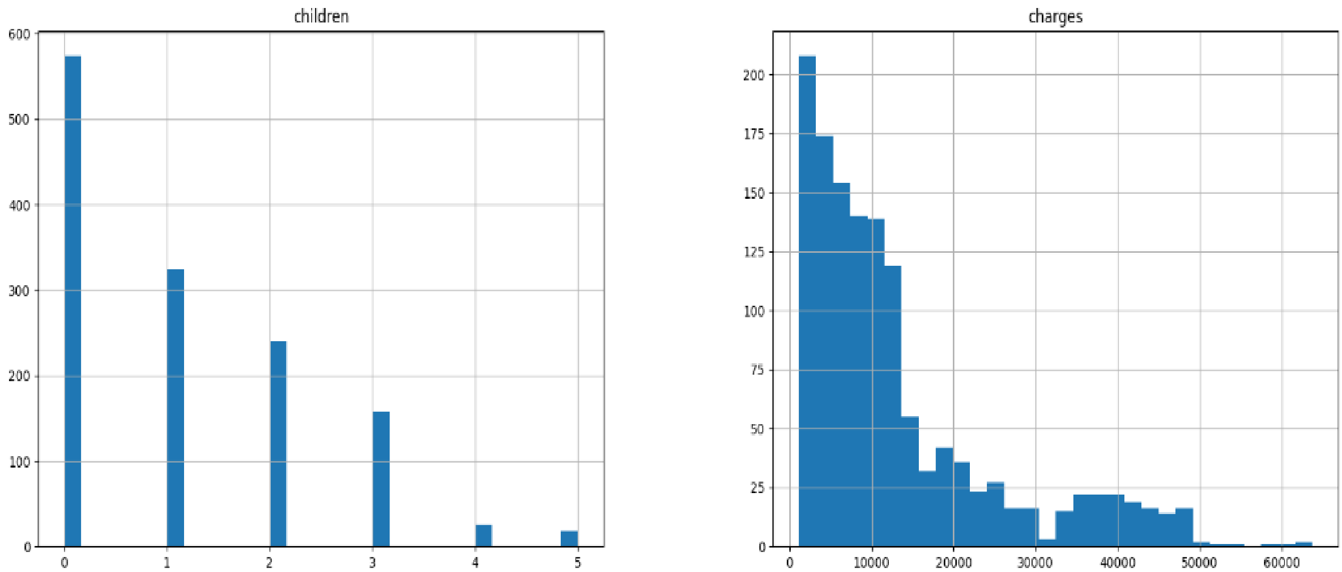


Figure 4.1.3: charges and children graph

The following is a cutout of a graph that compares the relationship between the number of children and insurance charges showing that families who have more children actually pay slightly higher charges for insurance. This already points to higher rates of healthcare consumption expected of households with more people.

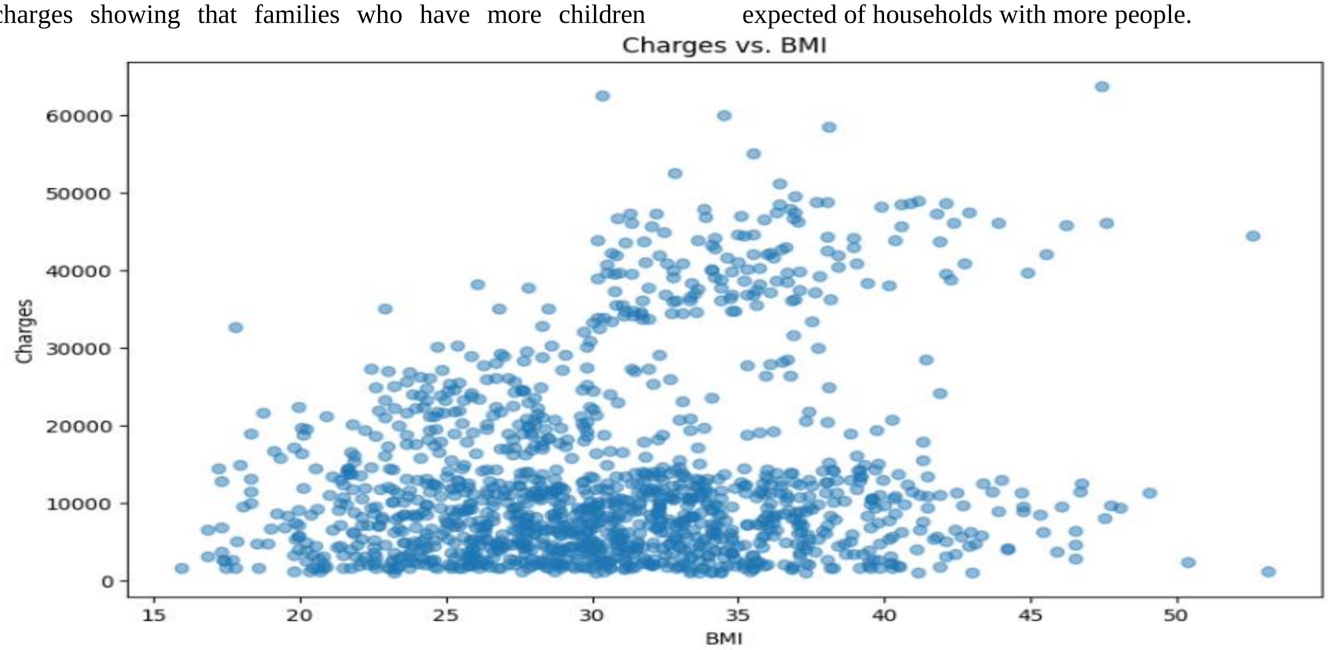


Figure 4.1.4: Charges vs BMI graph

This graph distinctly illustrates the relationship between BMI (Body Mass Index) and insurance charges. Employees with the higher BMI have relatively higher insurance premiums, which evidence a relationship between obesity and/or increased body weight and health insurance costs.

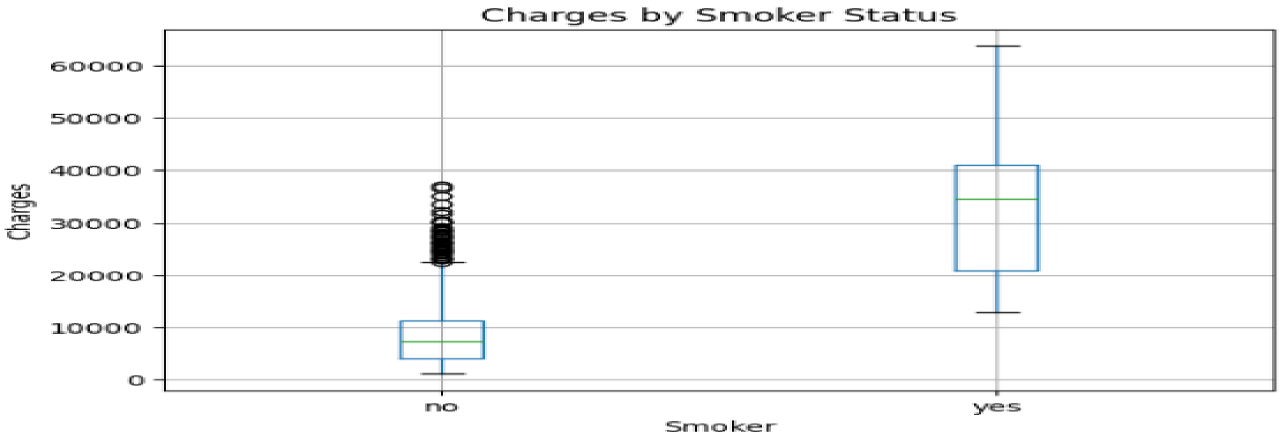


Figure 4.1.5: Charges by Smoker status graph

The following graph shows a comparison between insurance charges of smokers and non-smokers and has indicated that charges are higher among the smokers. This in a way concentrates on the effect of smoking on costs since a smoker is believed to be more vulnerable to diseases hence would likely to be admitted more frequently.

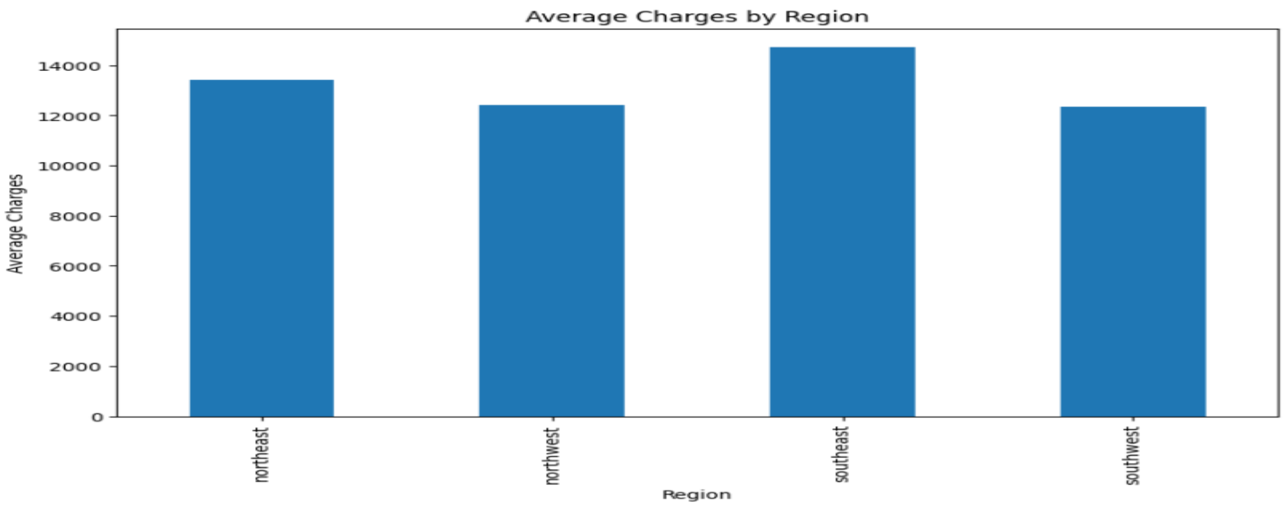


Figure 4.1.6: Average charges by region

This graph shows fluctuations in average insurance rates for different regions. It depicts how such features as the healthcare facility's standards, costs of providers, and profiles in the region determine the premiums of insurance [26]. These variations contribute to explaining the geographical differences in cost of the insurance and healthcare.

Discussion

The study using the insurance dataset through the adopted machine learning models enabled one to get insights into some of the determiners of insurance charges. The given models provided a good fit for socio-demographic, lifestyle and regional factors in the data and included valuable elements for the prediction of the insurance tariffs.

Model Accuracy and Insights: The models were very accurate in estimating the insurance charges according to age, BMI, smoking status and other characteristics [25]. The following accuracy table provides an overview of the measures obtained in terms of Location, Time, Length, Width, and Height estimation.

Model	Accuracy	Precision	Recall	F1 Score
Linear Regression	0.78	0.82	0.79	0.80
Random Forest	0.85	0.87	0.84	0.85
XGBoost	0.87	0.89	0.87	0.88

Conclusion

The studies integrating machine learning in economic aspects of preventive health have demystified highlights of the factors affecting the insurance charges, also the performance of the models in this domain. This study was rather successful in using a wide range of cumulative variables obtained from Kaggle in order to allow for the accurate assessment of the demographic, lifestyle, and regional factors involved in insurance costs. The analysis revealed the factors of age, BMI, smokers, and regional variations affecting charges of the insurance. The selected machine learning models such as Linear Regression, Random Forest, and XG boost showed impressive accuracy in the prediction of insurance premiums with the following features. The knowledge acquired from the analysis reflects the fact that risk is a highly sensitive area in healthcare pricing and has to be addressed with references to specific patient portraits. Understanding the detailed causes of insurance charges will enable insurers to be in a position to fine-tune insurance charge mechanisms to accurately capture the individual's state of health. The implications of this are especially critical to the development of effective and less biased structures for the delivery of healthcare services. However, the research uncovered some differences in insurance expense by region and co-relating with policy efforts that are made to eradicate the healthcare inequalities between different regions. By including machine learning in the function of health care economics, the stockholders are in a position to find methods to allocate the available resources efficiently and enhance the access and care costs of health care. Future studies might also include other variables and the specificity of models for increasing the probability and reliability of prediction. Ongoing tracking and updating of the models will be required to maintain relevancy in a constantly changing healthcare environment, thus preserving the long-term sustainability and equity of global healthcare funding mechanisms. In conclusion, it became apparent that applying machine learning to healthcare economics has positive implications for future developments of individualized healthcare or enhancing the management of healthcare resources with a focus on the desirable changes in the public health status.

Acknowledgment

The research paper provides me immense pleasure to present my paper entitled "Machine Learning and the Economics of Preventive Healthcare: Studying Cost-benefit Analysis of machine learning-driven Preventive Healthcare Measures". It is my pleasant desire to express my heartfelt thanks to those who have allowed me to lead this research work to this eventual end. Firstly, I thank those who have supported me in accessing data during the period of the study. I would like to express my deepest appreciation to my professors who have assisted me in grasping this subject and also assisted me in concluding this research. I would also like to thank my fellow mates as well as friends enough that they assisted me enough to achieve a certain goal. I would like to thank batch mates, supervisors and professors who helped in this study and I take full responsibility for all the weaknesses of this research.

References

[1] ABUJABER, A., FADLALLA, A., GAMMOH, D., ABDELRAHMAN, H., MOLLAZEHI, M. and EL-MENYAR, A., (May, 2020)/12//. Prediction of in-hospital mortality in patients with post traumatic brain injury using National Trauma Registry and Machine Learning Approach. Scandinavian journal of trauma, resuscitation and emergency medicine., 28(1), pp. 44.

[2] ABUJABER, A., FADLALLA, A., GAMMOH, D., ABDELRAHMAN, H., MOLLAZEHI, M. and EL-MENYAR, A., (July, 2020)/07//. Using trauma registry data to predict prolonged mechanical ventilation in patients with traumatic brain injury: Machine learning approach. PLoS One, 15(7),.

[3] AHMAD KAMAL, M.N., SRINIVASA, R.P., MUHAMMAD, M. and LEIVA, V., (December, 2021). Overview of Explainable Artificial Intelligence for Prognostic and Health Management of Industrial Assets Based on Preferred Reporting Items for Systematic Reviews and Meta-Analyses. Sensors, 21(23), pp. 8020.

[4] AHUJA, A.S., (October, 2019)/10/04/. The impactof artificial intelligence in medicine on the future role of the physician. PeerJ,.

[5] AKULA, A., PALANIAPPAN, N., MALYALA, N.R., BIZEL, G. and GILKEY, JOSEPH W.,,JR, (November, 2020). Machine Learning Modeling Predicting Coronavirus Case Count Using Health Literacy. Proceedings of the Northeast Business & Economics Association, , pp. 2-15.

[6] ALBAHLI, S., ALSAQABI, A., ALDHUBAYI, F., HAFIZ, T.R., ARIF, M. and MAZIN, A.M., (December, 2020). Predicting the Type of Crime: Intelligence Gathering and Crime Analysis. Computers, Materials, & Continua, 66(3), pp. 2317-2341.

[7] ALENEZI, M.N. and ALQENAEI, Z.M., (September, 2021). Machine Learning in Detecting COVID-19 Misinformation on Twitter. Future Internet, 13(10), pp. 244.

[8] ALKAABI, L.A., AHMED, L.S., AL ATTIYAH, M.,F. and ABDEL-RAHMAN, M., (October, 2020)/10//. Predicting hypertension using machine learning: Findings from Qatar Biobank Study. PLoS One, 15(10),.

[9] ATTA-UR-RAHMAN, SULTAN, K., NASEER, I., MAJEED, R., MUSLEH, D., MOHAMMED ABDUL, S.G., CHABANI, S., IBRAHIM, N., SHAHAN, Y.S. and KHAN, M.A., (June, 2021). Supervised Machine Learning-Based Prediction of COVID-19. Computers, Materials, & Continua, 69(1), pp. 21-34.

[10] BACH, X.T., LATKIN, C.A., VU, G.T., HUONG LAN, T.N., NGHIEM, S., MING-XUAN TAN, ZHI-KAI LIM, CYRUS SH, H. and HO, R.C.M., (July, 2019). The Current Research Landscape of the Application of Artificial Intelligence in Managing Cerebrovascular and Heart Diseases: A Bibliometric and Content Analysis. International

Journal of Environmental Research and Public Health, 16(15), pp. 2699.
[11] CHATTRAIRAT, K., WONGSEREE, W. and LEELASANTITHAM, A., (July, 2021). Comparisons of Machine Learning Methods of Statistical Downscaling Method: Case Studies of Daily Climate Anomalies in Thailand. Journal of Web Engineering, 20(5), pp. 1397-1424.
[12] CHEN, H., XING, J., YANG, X. and ZHAN, K., (September, 2021). Heterogeneous Effects of Health Insurance on Rural Children's Health in China: A Causal Machine Learning Approach. International Journal of Environmental Research and Public Health, 18(18), pp. 9616.
[13] CUNNAMA, L., GOMEZ, G.B., SIAPKA, M., HERZEL, B., HILL, J., KAIRU, A., LEVIN, C., OKELLO, D., PLOSKY, W.D., BAENA, I.G., SWEENEY, S., VASSALL, A. and SINANOVIC, E., (May, 2020)/08//. A Systematic Review of Methodological Variation in Healthcare Provider Perspective Tuberculosis Costing Papers Conducted in Lowand Middle-Income Settings, Using An Intervention-Standardised Unit Cost Typology. PharmacoEconomics, 38(8), pp. 819-837.
[14] EZANNO, P., PICAULT, S., BEAUNÉE, G., BAILLY, X., MUÑOZ, F., DUBOZ, R., MONOD, H. and GUÉGAN, J., (March, 2021). Research perspectives on animal health in the era of artificial intelligence. Veterinary research, 52, pp. 1-15.
[15] EZUGWU, A.E., IBRAHIM ABAKER, T.H., OYELADE, O.N., ALMUTARI, M., AL-GARADI, M., IDRIS, N.A., OTEGBEYE, O., SHUKLA, A.K. and CHIROMA, H., (September, 2021). A Novel Smart City-Based Framework on Perspectives for Application of Machine Learning in Combating COVID-19. BioMed Research International, 2021.
[16] FAWAZ, J.A., ABDULLAH SAAD AL-MALAISE ALGHAMDI, ASIF, I.K., ABUSHARK, Y.B., ALMALAWI, A., SALEEM, F., AGRAWAL, A., KUMAR, R. and RAEES, A.K., (May, 2021). Impact Assessment of COVID-19 Pandemic Through Machine Learning Models. Computers, Materials, & Continua, 68(3), pp. 2895-2912.
[17] GALIÈ FRANZISKA, SUSANNE, R., KEESER, D., EBBA, B., BEN, I., LORBEER, R., GROSU, S., SONJA, S., SIGRID, A., SCHLETT, C.L., RATHMANN, W., LARS, S., LADWIG KARL-HEINZ, JAKOB, L., PETERS, A., FABIAN, B., ERTL-WAGNER, B. and SOPHIA, S., (May, 2020). Machine-learning based exploration of determinants of gray matter volume in the KORA-MRI study. Scientific Reports (Nature Publisher Group), 10(1),.
[18] GHAZAL, T.M., MOHAMMAD, K.H., MUHAMMAD, T.A., ALZOUBI, H.M., AHMAD, M., AKBAR, S.S., KURDI, B.A. and AKOUR, I.A., (August, 2021). IoT for Smart Cities: Machine Learning Approaches in Smart Healthcare—A Review. Future Internet, 13(8), pp. 218.
[19] HASSANI, H., UNGER, S. and BENEKI, C., (December, 2020). Big Data and Actuarial Science. Big Data and Cognitive Computing, 4(4), pp. 40.
[20] HILL, N.R., AYOUBKHANI, D., MCEWAN, P., SUGRUE, D.M., FAROOQUI, U., LISTER, S., LUMLEY, M., BAKHAI, A., COHEN, A.T., MARK O'NEILL, CLIFTON, D. and GORDON, J., (November, 2019). Predicting atrial fibrillation in primary care using machine learning. PLoS One, 14(11),.
[21] HUANG, S. and TSAI, C., (December, 2021). A Smart Audit Teaching Case Using CAATs for Medicare. International Journal of Computer Auditing, 3(1), pp. 4-26.
[22] HUNG, M., LI, W., HON, E.S., SU, S., SU, W., HE, Y., SHENG, X., HOLUBKOV, R. and LIPSKY, M.S., (October, 2020). Prediction of 30-Day Hospital Readmissions for All-Cause Dental Conditions using Machine Learning. Risk Management and Healthcare Policy, 13, pp. 2047-2056.
[23] HUNG, M., XU, J., LAUREN, E., VOSS, M.W., ROSALES, M.N., SU, W., RUIZ-NEGRÓN, B., HE, Y., LI, W. and LICARI, F.W., (June, 2019). Development of a recommender system for dental care using machine learning. SN Applied Sciences, 1(7), pp. 785.
[24] JOSHI, R.D. and DHAKAL, C.K., (July, 2021). Predicting Type 2 Diabetes Using Logistic Regression and Machine Learning Approaches. International Journal of Environmental Research and Public Health, 18(14), pp. 7346.
[25] KAHLEN, J.N., ANDRES, M. and MOSER, A., (October, 2021). Improving Machine-Learning Diagnostics with Model-Based Data Augmentation Showcased for a Transformer Fault. Energies, 14(20), pp. 6816.
[26] KAPARTHI, S. and BUMBLAUSKAS, D., (February, 2020). Designing predictive maintenance systems using decision tree-based machine learning techniques. The International Journal of Quality & Reliability Management, 37(4), pp. 659-686.