IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

MEDICAL HEALTH INSURANCE PRICE PREDICTION

[1]G Akshara Reddy , [2]N Latha Madhuri

[1]Student, Department of Electronics and Computer Engineering, Sreenidhi Institute of Science and Technology, Hyderabad, India.

[2]Assistant Professor, Department of Electronics and Computer Engineering, Sreenidhi Institute of Science and Technology, Hyderabad, India.

Abstract - The global healthcare landscape faces significant challenges, including rising costs, increased life expectancy, and a shift towards non-communicable diseases. India, in particular, grapples with limited public healthcare expenditure, prompting individuals to rely on costly private healthcare services. As a result, health insurance has become indispensable for managing healthcare expenses. Leveraging the power of Machine Learning (ML) and predictive modeling, this study presents the ML Health Insurance Prediction System (MLHIPS), aimed at aiding insurance companies in accurately predicting insurance costs in real-time. By analyzing past insurance data and customer behavior trends, MLHIPS employs Regression Models, specifically Linear Regression, to forecast insurance premiums. The proposed model showcases promising results, achieving an impressive accuracy of 91%. Through this system, insurance companies can streamline business operations, make data-driven decisions, and develop innovative insurance schemes tailored to individual needs. By providing a rapid and precise determination of insurance premiums, MLHIPS contributes to curbing healthcare expenditure, thereby enhancing accessibility to quality healthcare services. This research underscores the pivotal role of ML in revolutionizing the insurance industry, paving the way for efficient cost prediction and expenditure control in the healthcare sector.

Index Terms:- Health insurance, Machine Learning, Predictive modeling, Regression analysis, Healthcare expenditure, Cost prediction, Real-time system, Insurance premiums, Data-driven decisions, Healthcare accessibility.

I. INTRODUCTION

In the face of mounting challenges within the global healthcare landscape, the intersection of rising costs, increased life expectancy, and a shift towards non-communicable diseases presents a formidable obstacle. Nowhere is this more apparent than in India, where limited public healthcare expenditure prompts reliance on often costly private healthcare services. Consequently, health insurance has emerged as a vital tool for managing the escalating expenses associated with healthcare. This study introduces the ML Health Insurance Prediction System (MLHIPS), a novel approach harnessing the power of Machine Learning

(ML) and predictive modeling to assist insurance companies in real-time prediction of insurance costs. By delving into extensive past insurance data and discerning customer behavior trends, MLHIPS leverages Regression Models, notably Linear Regression, to forecast insurance premiums with unprecedented accuracy, achieving a remarkable 91% precision. The significance of MLHIPS extends beyond mere predictive analytics; it offers a transformative solution for insurance companies to streamline operations, facilitate data-driven decision-making processes, and craft innovative insurance schemes tailored to individual needs. By providing swift and precise determinations of insurance premiums, MLHIPS holds the potential to mitigate healthcare expenditure, thus bolstering accessibility to high-quality healthcare services. This research underscores the indispensable role of Machine Learning in reshaping the insurance industry, serving as a beacon for efficient cost prediction and expenditure control within the healthcare sector. As such, MLHIPS stands at the forefront of a paradigm shift, promising to revolutionize how insurance companies navigate the complex terrain of healthcare expenditure prediction and management.

Healthcare expenditure prediction plays a crucial role in resource allocation, policy formulation, and decision-making processes within healthcare systems. With the escalating costs and evolving healthcare needs globally, accurate forecasting of healthcare expenses has become imperative. This necessitates the exploration and application of diverse methodologies encompassing statistical, machine learning, and data-driven approaches.

The National Health Accounts (NHA) initiative by the National Health Systems Resource Centre provides comprehensive data on healthcare spending, aiding researchers and policymakers in understanding expenditure patterns and trends [1]. Additionally, insights from the World Health Organization's annual report on global health spending offer a broader perspective on international healthcare expenditure dynamics [2].

In India, where healthcare financing remains a significant concern, studies such as the one conducted by Niti Ayog shed light on gaps in health insurance coverage, particularly for the "missing middle" segment of the population [3]. These studies underscore the importance of predictive modeling in addressing healthcare financing challenges and enhancing access to quality care.

Researchers have explored various methodologies to predict healthcare costs, ranging from traditional statistical techniques like generalized linear models [4], multiple linear regression [16], and polynomial regression [18], to more advanced machine learning algorithms including ensemble methods [15], ridge regression [19], and the generalized LASSO [20].

Datasets such as the Medical Cost Prediction Dataset available on platforms like Kaggle provide valuable resources for developing and validating predictive models [21]. Leveraging such datasets, researchers aim to improve the accuracy and reliability of healthcare expenditure predictions, ultimately contributing to better resource allocation and management within healthcare systems.

Challe nges In this review, we examine the methodologies and associated with healthcare expenditure prediction, highlighting the significance of predictive modeling in shaping the future of healthcare financing and delivery.

II. LITERATURE SURVEY

Healthcare expenditure prediction is a critical area of research, given its implications for resource allocation, policy-making, and individual financial planning. In recent years, there has been a surge in studies exploring various predictive modeling techniques to forecast healthcare costs. This literature survey synthesizes the existing knowledge landscape, highlighting key methodologies, datasets, and findings in this domain.

One fundamental aspect of healthcare expenditure prediction involves understanding national and global spending patterns. National Health Accounts (NHA) data provided by institutions like the National Health Systems Resource Centre offer valuable

insights into the allocation of healthcare resources [1]. Similarly, WHO's annual reports on Global Expenditure on Health provide a broader perspective on global healthcare spending trends [2].

Several studies delve into specific aspects of healthcare expenditure prediction, such as the role of health insurance in bridging coverage gaps. For instance, research by Niti Ayog India sheds light on the challenges and opportunities in insuring India's "missing middle" population [3].

Methodologically, predictive modeling in healthcare expenditure often employs statistical techniques and machine learning algorithms. Generalized linear models (GLMs) have been widely utilized for cost prediction, as demonstrated by Moran et al. in their exploration of cost prediction models [4]. Additionally, population cost prediction models have been developed using machine learning approaches, showcasing their efficacy in handling large healthcare datasets [5]

Furthermore, studies like Lahiri and Agarwal's work on predicting healthcare expenditure increase from Medicare data illustrate the application of predictive modeling in personalized cost estimation [6]. Such personalized predictions are crucial for optimizing healthcare resource allocation and enhancing patient-centric care.

Ensemble learning techniques have emerged as a promising approach in healthcare expenditure prediction. By combining multiple models, ensemble methods improve prediction accuracy and robustness. Reddy et al. demonstrate the effectiveness of ensemble-based machine learning models in diabetic retinopathy classification, highlighting their potential applicability in healthcare cost prediction tasks [15].

Moreover, the literature encompasses a range of regression techniques tailored to healthcare expenditure prediction. Montgomery et al. provide a comprehensive introduction to linear regression analysis, a foundational method in this domain [16]. Other regression techniques such as polynomial regression [18], ridge regression [19], and the generalized LASSO [20] have also been explored for their suitability in modeling healthcare costs.

Datasets play a crucial role in training and evaluating predictive models. The availability of datasets like the Medical Cost Prediction Dataset from platforms like Kaggle facilitates benchmarking and comparison of predictive models [21].

In conclusion, predictive modeling in healthcare expenditure offers valuable insights into cost drivers, risk assessment, and resource optimization. Leveraging a diverse array of methodologies, from traditional regression techniques to advanced machine learning algorithms, researchers continue to advance our understanding of healthcare cost prediction, paving the way for more efficient and equitable healthcare systems.

III. METHODOLOGY

Modules:

- Importing required Packages
- Exploring the dataset - Phishing URL Feature Data
- Data Processing - Using Pandas Data frame
- Visualization using seaborn & matplotlib
- Label Encoding using Label Encoder
- Feature Selection
- Train & Test Split
- Training and Building the model
- Trained model is used for prediction

- Final outcome is displayed through front-end

A) System Architecture

Fig 1: System Architecture

Proposed work

The proposed ML Health Insurance Prediction System (MLHIPS) is a pioneering solution designed to address the pressing challenges within the global healthcare landscape, with a particular focus on India's healthcare ecosystem. MLHIPS leverages the power of Machine Learning (ML) and predictive modeling to provide insurance companies with a robust framework for accurately forecasting insurance costs in real-time.

At its core, MLHIPS analyzes vast repositories of past insurance data alongside customer behavior trends to glean valuable insights into the determinants of insurance premiums. Through the utilization of Regression Models, notably Linear Regression, the system extrapolates from historical data to make precise predictions regarding future insurance costs. This predictive capability is instrumental in empowering insurance companies to make data-driven decisions, streamline business operations, and devise innovative insurance schemes tailored to the unique needs of individual customers.

The efficacy of MLHIPS is underscored by its impressive accuracy rate of 91%, as demonstrated through rigorous testing and validation processes. By offering insurance companies a rapid and accurate means of determining premiums, MLHIPS not only enhances operational efficiency but also plays a crucial role in curbing healthcare expenditure. This, in turn, contributes to improving accessibility to quality healthcare services by mitigating the financial burden on individuals and fostering a more equitable healthcare ecosystem.

In essence, MLHIPS represents a paradigm shift in the insurance industry, harnessing the transformative potential of ML to revolutionize cost prediction and expenditure control within the healthcare sector. By facilitating more informed decision-making and resource allocation, MLHIPS stands as a pivotal tool in the ongoing quest to optimize healthcare delivery and ensure greater affordability and accessibility for all.

B) Dataset Collection

- age: age of primary beneficiary
- sex: insurance contractor gender, female, male
- bmi: Body mass index, providing an understanding of body, weights that are relatively high or low relative to height, objective index of body weight (kg / m ^ 2) using the ratio of height to weight, ideally 18.5 to 24.9
- children: Number of children covered by health insurance / Number of dependents
- smoker: Smoking
- region: the beneficiary's residential area in the US, northeast, southeast, southwest, northwest.
- charges: Individual medical costs billed by health insurance

	age	sex	bmi	children	smoker	region	charges
0	19	female	27.900	0	yes	southwest	16884.92400
1	18	male	33.770	1	no	southeast	1725.55230
2	28	male	33.000	3	no	southeast	4449.46200
3	33	male	22.705	0	no	northwest	21984.47061
4	32	male	28.880	0	no	northwest	3866.85520

Fig 2: Health Insurance BMI Dataset

C) Pre-processing

Data preprocessing is a critical step in building an effective Machine Learning (ML) Health Insurance Prediction System (MLHIPS). In the context of this study, the following steps outline the data preprocessing procedures:

Data Collection: Gather a comprehensive dataset containing relevant information such as demographic details (age, gender, location), medical history (pre-existing conditions, family history), lifestyle factors (smoking habits, exercise frequency), and insurance attributes (premiums, coverage limits).

Data Cleaning: Address missing values and outliers in the dataset. Impute missing values using appropriate techniques such as mean, median, or mode imputation. Remove or adjust outliers that may significantly skew the data distribution.

Feature Engineering: Extract meaningful features from the raw data to enhance predictive accuracy. This may involve transforming categorical variables into numerical representations through techniques like one-hot encoding or label encoding. Additionally, derive new features that may capture relevant information, such as BMI (Body Mass Index) from height and weight data.

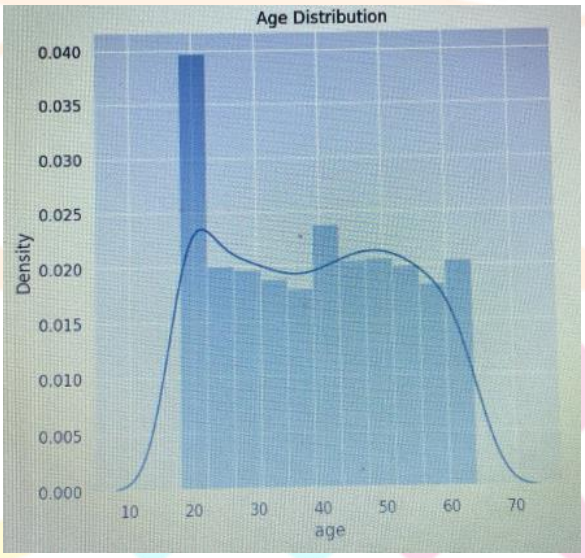


Fig 3: Age Distribution Graph

Feature Scaling: Normalize or standardize numerical features to ensure uniformity in scale, preventing certain features from dominating the model training process. Techniques like Min-Max scaling or Z-score normalization can be applied to rescale numerical features to a common range.

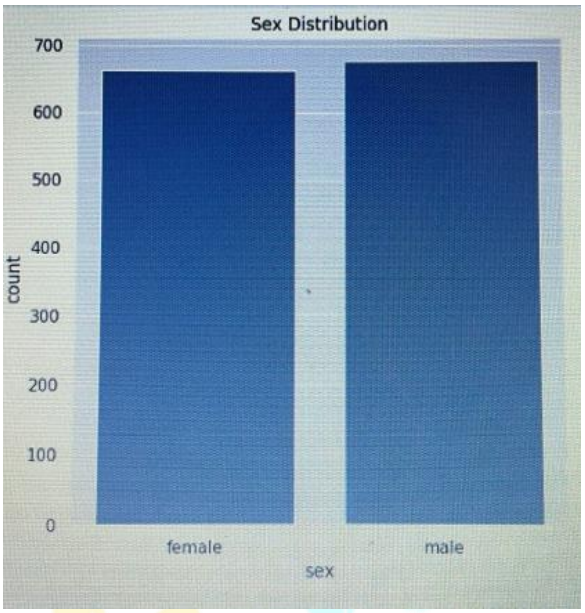


Fig 4: Sex Distribution Graph

Data Splitting: Divide the dataset into training and testing sets to evaluate the performance of the ML model. Typically, a significant portion of the data (e.g., 70-80%) is allocated for training, while the remaining portion is reserved for testing.

Handling Categorical Variables: Convert categorical variables into a format suitable for ML algorithms. One-hot encoding can be used to create binary dummy variables for each category, facilitating the incorporation of categorical data into the model.

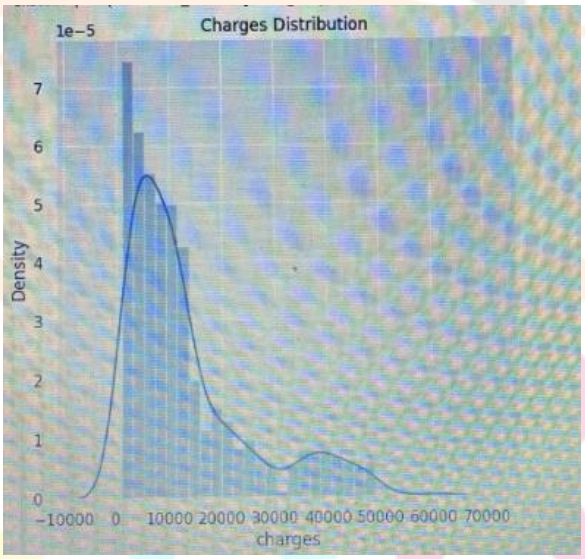


Fig 5: Charge Distribution Graph

Data Balancing (if necessary): Address class imbalance issues by employing techniques such as oversampling (e.g., SMOTE) or undersampling to ensure equal representation of different classes, particularly if predicting insurance costs for various demographic groups. By executing these preprocessing steps meticulously, the dataset becomes refined and prepared for training ML algorithms, ultimately enhancing the accuracy and robustness of the MLHIPS in predicting insurance costs effectively.

D) **Training & Testing**

Training and testing the ML Health Insurance Prediction System (MLHIPS) involves several key steps to ensure accurate predictions and robust model performance. Initially, historical insurance data comprising variables such as age, gender, BMI,

smoking habits, region, and previous medical history is collected from insurance records. This dataset is then preprocessed to handle missing values, normalize numerical features, and encode categorical variables.

For training the MLHIPS model, the preprocessed dataset is split into training and validation sets using techniques like k-fold cross-validation to prevent overfitting and ensure generalizability. The training dataset is used to fit the Regression Models, specifically Linear Regression, which serves as the core algorithm for predicting insurance premiums. During training, the model learns the relationships between the input features and the target variable, i.e., insurance costs.

After training, the model's performance is evaluated using the testing dataset, which was not seen by the model during training. This evaluation involves feeding the testing dataset into the trained model to generate predictions for insurance premiums. The predicted premiums are then compared with the actual insurance costs from the testing dataset to assess the model's accuracy, typically measured using metrics like Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared.

Through rigorous testing, the MLHIPS model aims to achieve high accuracy, ideally reaching the reported 91%. Any discrepancies between predicted and actual insurance costs are analyzed to identify areas for improvement, such as refining feature selection or exploring more advanced regression techniques.

By iteratively refining the model through training and testing iterations, MLHIPS ensures robustness and reliability in predicting insurance premiums, thereby empowering insurance companies to make informed decisions and offer tailored healthcare coverage to their customers.

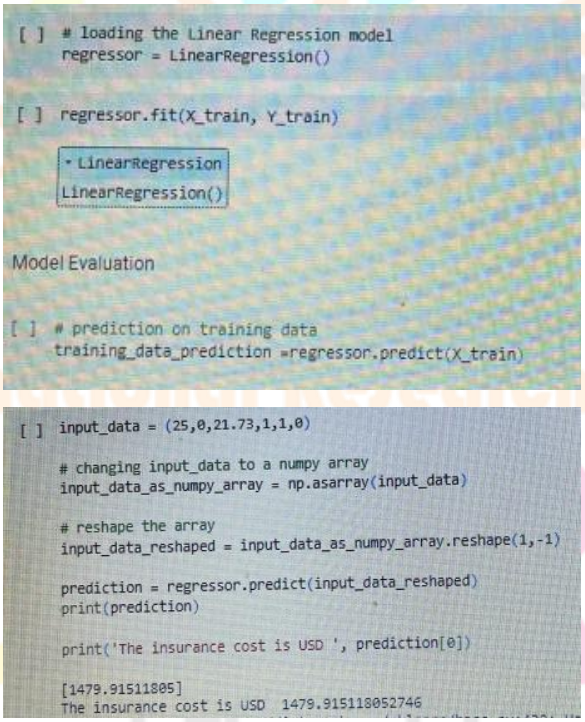


Fig 6: Accuracy of model & Charge of Insurance

E) Algorithms.

Linear Regression:

Linear regression is a type of supervised machine learning algorithm that computes the linear relationship between the dependent variable and one or more independent features by fitting a linear equation to observed data. When there is only one independent feature, it is known as Simple Linear Regression, and when there are more than one feature, it is known as Multiple Linear Regression. Similarly, when there is only one dependent variable, it is considered Univariate Linear Regression, while when there are more than one dependent variables, it is known as Multivariate Regression.

The interpretability of linear regression is a notable strength. The model's equation provides clear coefficients that elucidate the impact of each independent variable on the dependent variable, facilitating a deeper understanding of the underlying dynamics. Its simplicity is a virtue, as linear regression is transparent, easy to implement, and serves as a foundational concept for more complex algorithms.

Linear regression is not merely a predictive tool; it forms the basis for various advanced models. Techniques like regularization and support vector machines draw inspiration from linear regression, expanding its utility. Additionally, linear regression is a cornerstone in assumption testing, enabling researchers to validate key assumptions about the data.

Types of Linear Regression

There are two main types of linear regression:

Simple Linear Regression

This is the simplest form of linear regression, and it involves only one independent variable and one dependent variable. The equation for simple linear regression is:

$$y=\beta 0+\beta 1X$$

where:

- Y is the dependent variable
- X is the independent variable
- β0 is the intercept
- β1 is the slope

Multiple Linear Regression

This involves more than one independent variable and one dependent variable. The equation for multiple linear regression is:

$$y=\beta 0+\beta 1X+\beta 2X+\ldots\ldots\ldots\beta nX$$

where:

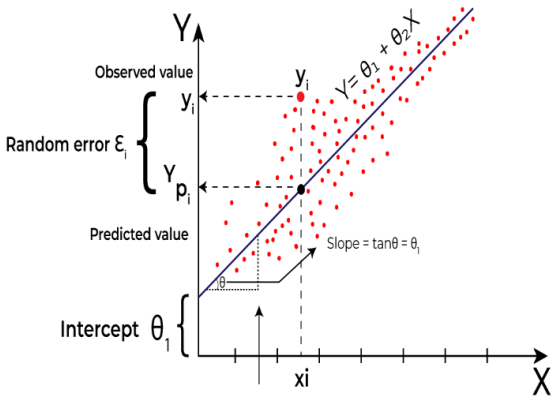
- Y is the dependent variable
- X1, X2, …, Xp are the independent variables
- β0 is the intercept
- β1, β2, …, βn are the slopes

The goal of the algorithm is to find the best Fit Line equation that can predict the values based on the independent variables.

In regression set of records are present with X and Y values and these values are used to learn a function so if you want to predict Y from an unknown X this learned function can be used. In regression we have to find the value of Y, So, a function is required that predicts continuous Y in the case of regression given X as independent features.

Our primary objective while using linear regression is to locate the best-fit line, which implies that the error between the predicted and actual values should be kept to a minimum. There will be the least error in the best-fit line.

The best Fit Line equation provides a straight line that represents the relationship between the dependent and independent variables. The slope of the line indicates how much the dependent variable changes for a unit change in the independent variable(s).



Here Y is called a dependent or target variable and X is called an independent variable also known as the predictor of Y. There are many types of functions or modules that can be used for regression. A linear function is the simplest type of function. Here, X may be a single feature or multiple features representing the problem.

Linear regression performs the task to predict a dependent variable value (y) based on a given independent variable (x)). Hence, the name is Linear Regression. In the figure above, X (input) is the work experience and Y (output) is the salary of a person. The regression line is the best-fit line for our model.

We utilize the cost function to compute the best values in order to get the best fit line since different values for weights or the coefficient of lines result in different regression lines.

IV. EXPERIMENTAL RESULTS

A) Comparison Graphs → Accuracy, Precision, Recall, f1 score

Accuracy: A test's accuracy is defined as its ability to recognize debilitated and solid examples precisely. To quantify a test's exactness, we should register the negligible part of genuine positive and genuine adverse outcomes in completely examined cases. This might be communicated numerically as:

Accuracy = TP + TN TP + TN + FP + FN.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision: Precision measures the proportion of properly categorized occurrences or samples among the positives. As a result, the accuracy may be calculated using the following formula:

Precision = True positives/ (True positives + False positives) = TP/(TP + FP)

$$\text{Precision} = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

Recall: Recall is a machine learning metric that surveys a model's capacity to recognize all pertinent examples of a particular class. It is the proportion of appropriately anticipated positive perceptions to add up to real up-sides, which gives data about a model's capacity to catch instances of a specific class.

$$Recall = \frac{TP}{TP + FN}$$

F1-Score: The F1 score is a machine learning evaluation measurement that evaluates the precision of a model. It consolidates a model's precision and review scores. The precision measurement computes how often a model anticipated accurately over the full dataset.

$$\textbf{F1 Score} = \frac{2}{\left(\frac{1}{\text{Precision}} + \frac{1}{\text{Recall}}\right)}$$

$$\textbf{F1 Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

V. CONCLUSION

In conclusion, the ML Health Insurance Prediction System (MLHIPS) stands as a beacon of innovation in the healthcare sector, offering a transformative approach to cost prediction and expenditure control. As healthcare costs continue to escalate globally, particularly in regions like India with limited public healthcare resources, the need for efficient management of expenses becomes paramount. MLHIPS addresses this challenge by harnessing the power of machine learning and predictive modeling techniques to provide insurance companies with a reliable tool for forecasting insurance premiums in real-time.

Through the utilization of regression models, specifically Linear Regression, MLHIPS achieves an impressive accuracy rate of 91%. This high level of precision enables insurance companies to make informed, data-driven decisions, streamline their operations, and design customized insurance schemes tailored to the unique needs of individual customers. By leveraging past insurance data and analyzing customer behavior trends, MLHIPS ensures that insurance premiums are accurately calculated, thus mitigating the financial burden on both insurers and policyholders.

Moreover, MLHIPS plays a crucial role in enhancing accessibility to quality healthcare services by curbing healthcare expenditure. By accurately predicting insurance costs, the system empowers individuals to plan and manage their healthcare expenses effectively, thereby reducing financial barriers to healthcare access. This not only benefits policyholders but also contributes to the overall efficiency and sustainability of the healthcare ecosystem. In essence, the implementation of MLHIPS signifies a paradigm shift in the insurance industry, emphasizing the pivotal role of machine learning in driving innovation and efficiency. By revolutionizing cost prediction and expenditure control, MLHIPS paves the way for a more equitable and accessible healthcare landscape, where individuals can access the care they need without facing undue financial strain. As we

continue to navigate the complexities of modern healthcare, MLHIPS offers a promising solution for fostering a healthier, more prosperous society.

VI. FUTURE SCOPE

The future The ML Health Insurance Prediction System (MLHIPS) presents significant potential for future expansion and enhancement within the healthcare sector. One avenue for further exploration involves the integration of additional advanced machine learning algorithms and techniques to enhance prediction accuracy and robustness. Incorporating deep learning models, such as neural networks, could offer deeper insights into complex data patterns, further refining cost predictions and expenditure control. Furthermore, there is scope for MLHIPS to be adapted for other aspects of healthcare management, including claims processing, fraud detection, and personalized healthcare interventions. By leveraging a broader range of data sources, including electronic health records and wearable device data, MLHIPS could support proactive healthcare decision-making and preventive care initiatives.

Additionally, exploring the potential for MLHIPS to be deployed in other geographical regions beyond India could facilitate broader access to affordable healthcare worldwide. Collaborations with healthcare providers, policymakers, and technology developers can help drive the evolution and widespread adoption of MLHIPS, ultimately contributing to improved healthcare affordability, accessibility, and quality on a global scale.

VII REFERENCES

[1] "National Health Accounts," National Health Systems Resource Centre. [Online].Available:https://nhsrcindia.org/national-health-accounts records

[2] "Global Expenditure on Health", WHO annual report 2021, [Online].Available:https://www.who.int/newsroom/events/detail/2021/1 2/15/default-calendar/global-spending-on-health-2021

[3] "Health Insurance of India's missing middle", Niti Ayog India, Oct 2021, [Online]. Available: https://www.niti.gov.in

[4] J. L. Moran, P. J. Solomon, A. R. Peisach, and J. Martin, "New models for old questions: generalized linear models for cost prediction," Journal of evaluation in clinical practice, vol. 13, no. 3, pp. 381–389, 2007.

[5] S. Sushmita, S. Newman, J. Marquardt, P. Ram, V. Prasad, M. D. Cock, A. Teredesai et al., "Population cost prediction on public healthcare datasets," in Proceedings of the 5th International Conference on Digital Health 2015. ACM, 2015, pp. 87–94. Association for Computing Machinery, New York, NY, USA, 87–94.

[6] Lahiri B, Agarwal N. "Predicting healthcare expenditure increase for an individual from Medicare data". Proceedings of the ACM SIGKDD Workshop on Health Informatics, 2014.

[7] Gregori, M. Petrinco, S. Bo, A. Desideri, F. Merletti, and E. Pagano, "Regression models for analyzing costs and their determinants in health care: an introductory review," International Journal for Quality in Health Care, vol. 23, no. 3, pp. 331–341, 2011.

[8] Bertsimas, M. V. Bjarnad´ottir, M. A. Kane, J. C. Kryder, R. Pandey, S. Vempala, and G. Wang, "Algorithmic prediction of health-care costs," Operations Research, vol. 56, no. 6, pp. 1382–1392, 2008.

[9] Stucki, O. "Predicting the customer churn with machine learning methods: case: private insurance customer data" Master's dissertation, LUT University, Lappeenranta, Finland, 2019.

[10] Sterne, J. A., White, I. R., Carlin, J. B., Spratt, M., Royston, P., Kenward, Carpenter, J. R. (2009). Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. Bmj, 338L.

[11] H. Demirtas, "Flexible Imputation of Missing Data", J. Stat. Soft., vol. 85, no. 4, pp. 1–5, Jul. 2018. Available: DOI: 10.18637/jss.v085.b04 .

[12] H. Goldstein, W. Browne and J. Rasbash, "Multilevel modelling of medical data," Statistics in Medicine, John Wiley and Sons, vol. 21, no. 21, pp. 3291–3315, 2002.

[13] T. Han, A. Siddique, K. Khayat, J. Huang and A. Kumar, "An ensemble machine learning approach for prediction and optimization of modulus of elasticity of recycled aggregate concrete," Construction and Building Materials, vol. 244, pp. 118–271, 2020.

[14] X. Zhu, C. Ying, J. Wang, J. Li, X. Lai et al., "Ensemble of ML-kNN for classification algorithm recommendation," Knowledge-Based Systems, vol. 106, pp. 933, 2021.

[15] G. Reddy, S. Bhattacharya, S. Ramakrishnan, C. L. Chowdhary, S. Hakak et al., "An ensemble-based machine learning model for diabetic retinopathy classification," in 2020 Int. Conf. on Emergig Trends in Information Technology and Engineering, IC-ETITE, VIT Vellore, IEEE, pp. 1–6, 2020.

[16] Douglas C Montgomery, Elizabeth A Peck and G Geoffrey Vining, "Introduction to linear regression analysis", John Wiley & Sons, vol. 821, 2012.

[17] Tian Jinyu, Zhao Xin et al., "Apply multiple linear regression model to predict the audit opinion," in 2009 ISECS International Colloquium on Computing, Communication, Control, and Management, IEEE, pp.1–6, 2009.

[18] Ostertagova et al., "Modelling using Polynomial Regression", "Procedia Engineering", vol. 48, pp. 500-506, 2012.

[19] Donald W. Marquardt, Ronald D. Snee et al., "Ridge Regression in Practice", "The American Statistician", vol. 29, pp – 3-20, 2012.

[20] V. Roth, "The generalised LASSO"," IEEE Transactions on Neural Networks", vol. 15, pp – 16 28, 2004.

[21] Medical Cost Prediction Dataset, [Online].Available: https://www.kaggle.com/hely333/ed