ISSN NO: 9726-001X
Volume 13 Issue 02 2025
JOURNAL OF CURRENT SCIENCE

ML Enabled Regressor for Humidity LevelsPrediction for MOX Gas Sensors Data

Badavath Kailas[1], Bogari Jayanth [2], Ubbala Vineel[3], Mr. K. Rajeshwar[4]
[1,2,3] UG Scholar, Department of ECE, St. Martin's Engineering College, Secunderabad, Telangana, India, 500100
[4]Assistant Professor, Department of ECE, St. Martin's Engineering College, Secunderabad, Telangana, India, 500100
kailasbadavath2@gmail.com

Abstract:

Metal–oxide–semiconductor (MOS) gas sensors are widely used for gas detection and monitoring. However, MOS gas sensors have always suffered from instability in the link between gas sensor data and the measured gas concentration. In this paper, we propose a novel deep learning approach that combines the surface state model and a Decision Tree-based regression to enhance the analysis of gas sensor data. The surface state model provides valuable insights into the microscopic surface processes underlying the conductivity response to pulse heating, while the DTR model effectively captures the temporal dependencies present in time-series data. Over the past decade, machine learning (ML) and artificial intelligence (AI) have attracted great interest in research and various practical applications. Currently, smart, fast, and high sensitivity with excellent selectivity is becoming increasingly interesting due to the high need for environmental safety and medical applications. The main challenge is to improve sensor selectivity, which requires the combination of interdisciplinary research areas to successfully develop smart gas/chemical sensing devices with better performance. In this review, we present a few principles of gas sensing based on low-cost interdigital electrodes (IDEs), such as electrochemical, resistive, capacitive, and acoustic sensors. In addition, the most important current methods for improving gas sensing performance, the different materials, the different techniques used to fabricate IDE gas sensors, and their advantages and limitations are presented. In addition, a comparison between different ML and AI algorithms for pattern recognition and classification algorithms is also discussed. The discussion then establishes application cases of smart ML algorithms, which provide efficient data processing methods, for the design of smart gas sensors that are highly selective. In addition, the challenges and limitations of ML in gas sensor applications are critically discussed. The study shows the importance of ML with the need for structural optimization to develop and improve smart, sensitive, and selective sensors.

Keywords: Machine Learning, Regression Models, Humidity Prediction, MOX Gas Sensors, Environmental Sensing, Data-Driven Modeling.

1. INTRODUCTION

Gas sensing technology has evolved significantly over the years, driven by increasing environmental concerns and industrial safety requirements. Metal–oxide–semiconductor (MOS) gas sensors have played a crucial role in detecting harmful gases such as carbon monoxide (CO), nitrogen oxides (NOx), methane (CH_4), and volatile organic compounds (VOCs). In India, rapid urbanization and industrial growth have contributed to rising air pollution levels, with cities like Delhi, Mumbai, and Kolkata experiencing dangerously high air quality index (AQI) levels. According to reports, air pollution contributes to nearly 1.67 million deaths annually in India. Traditional gas sensors often suffer from low selectivity, cross-sensitivity, and stability issues, leading to inaccurate readings. To address these limitations, machine learning (ML) and artificial intelligence (AI) techniques have been introduced to improve gas sensor performance, ensuring accurate, real-time monitoring.

2. LITERATURE SURVEY

Thomas et al. [1] conducted a comparative study to estimate the prevalence of dementia among elderly populations in Canada by utilizing data from the Canadian Study of Health and Aging and the National Population Health Survey. The research highlighted significant differences in prevalence estimates between the two surveys, attributing the variations to methodological differences in data collection and analysis. The study emphasized the importance of standardizing methodologies in epidemiological research to ensure accurate prevalence estimates, which are critical for planning and resource allocation in elderly care. The findings underscored the growing burden of dementia and the need for comprehensive national strategies to address it.

Kalache and Gatti [2] explored the concept of active aging as a policy framework, emphasizing the need for policies that support the health, participation, and security of the aging population. Their work argued that active aging policies are essential in mitigating the effects of an aging population on social and healthcare systems. The framework advocated for a holistic approach to aging, where the focus is not only on prolonging life but also on enhancing the quality of life through active participation in social, economic, and cultural activities. This research has influenced global aging policies, highlighting the importance of creating environments that enable older adults to remain active contributors to society.

The World Health Organization (WHO) [3] provided a comprehensive guide on what individuals and societies need to know about aging, focusing on the challenges and opportunities associated with an aging population. The document emphasized the importance of preparedness at both individual and societal levels, advocating for early planning to address the health, economic, and social implications of aging. It highlighted key strategies for promoting healthy aging, including maintaining physical activity, social engagement, and access to healthcare. The WHO's guidelines serve as a foundational reference for policymakers and healthcare providers aiming to improve the well-being of the elderly population.

Kulik et al. [4] examined the implications of aging populations on management practices, particularly in the workplace. The study discussed the challenges that an aging workforce poses to traditional management models and explored strategies to accommodate older workers while maintaining productivity. It highlighted the need for flexible work arrangements, continuous skill development, and policies that support work-life balance for older employees. The research also addressed the potential benefits of an aging workforce, such as experience and institutional knowledge, suggesting that organizations can gain a competitive advantage by effectively managing age diversity.

The World Health Organization (WHO) [5] published guidelines on disability and health, addressing the intersection of aging and disability. The report emphasized the growing prevalence of disability among the elderly and the need for healthcare systems to adapt to this demographic shift. It called for integrated care models that address both the medical and social needs of elderly individuals with disabilities. The WHO's recommendations include promoting accessibility, preventing disability through early intervention, and ensuring that health services are inclusive of older adults with disabilities.

The Centers for Disease Control and Prevention (CDC) [6] provided an overview of chronic diseases, focusing on their impact on the aging population. The report highlighted the high prevalence of chronic conditions such as heart disease, diabetes, and arthritis among older adults and the significant burden these diseases place on healthcare systems. The CDC advocated for preventive measures, including lifestyle modifications and regular screenings, to manage chronic diseases and improve the quality of life for the elderly. The report also emphasized the importance of public health initiatives in reducing the incidence and impact of chronic diseases among aging populations.

The Ontario government [7] released a budget report focusing on strengthening healthcare, with particular attention to the needs of the elderly population. The report outlined investments in healthcare infrastructure, long-term care facilities, and home care services, aimed at addressing the growing demand for elderly care. It emphasized the importance of accessible, high-quality care for the elderly, particularly in rural and underserved areas. The budget also included provisions for improving mental health services and support for caregivers, recognizing the critical role they play in the healthcare system.

Venkat [8] provided a global outlook on the healthcare industry, identifying key trends and challenges, particularly those related to aging populations. The report highlighted the increasing demand for healthcare services as populations age and the need for innovation in healthcare delivery to meet this demand. It discussed the role of technology, including telemedicine, wearable devices, and AI, in transforming healthcare and making it more accessible and efficient for the elderly. The report also addressed the economic implications of aging populations on healthcare systems worldwide and the need for sustainable funding models.

The Population Ageing Projections report [9] presented global and regional projections of population aging, emphasizing the rapid increase in the number of elderly individuals worldwide. The report provided detailed statistical analyses of aging trends, highlighting significant regional variations and the implications for social, economic, and healthcare systems. It called for urgent policy responses to address the challenges posed by an aging population, including the need for age-friendly environments, improved healthcare services, and social protection systems that can support the elderly.

Gelineau [10] reported on a study that ranked Canada fifth globally in terms of the well-being of its elderly population. The study assessed various factors contributing to elderly well-being, including health care access, income security, and social inclusion. The report highlighted Canada's strengths, such as its comprehensive healthcare system and social support networks, while also identifying areas for improvement, such as the need for better mental health services for the elderly. The findings underscored the importance of holistic approaches to elderly care that address both physical and mental health needs.

The Canadian Institute for Health Information [11] published a report on national health expenditure trends from 1975 to 2014, with a focus on the increasing costs associated with aging populations. The report provided a detailed analysis of healthcare spending, highlighting the significant portion of the budget allocated to elderly care. It also discussed the challenges of maintaining sustainable healthcare funding in the face of rising demand for services. The report called for reforms to improve the efficiency of healthcare delivery and ensure that resources are allocated effectively to meet the needs of the aging population. The Economist [12] published an article on the shifting demographics of the working-age population, discussing the economic implications of an aging workforce. The article explored how countries with aging populations are facing challenges such as labor shortages and increased pension costs. It also highlighted strategies that governments and businesses can adopt to mitigate these challenges, such as encouraging later retirement, promoting lifelong learning, and adapting workplaces to be more age-friendly. The article emphasized the need for policies that support both the elderly and the economy as populations age.

Anderson and Knickman [13] discussed the need for transforming the chronic care system to better meet the needs of aging populations. The article highlighted the limitations of the current healthcare system in addressing chronic diseases, which are prevalent among the elderly. The authors proposed a shift towards a more patient-centered approach that integrates medical care with social services and emphasizes prevention and management of chronic conditions. They argued that such a transformation is essential for improving the quality of care for the elderly and reducing the long-term costs of healthcare. AmeriGlide [14] provided an analysis of the advantages and disadvantages of nursing homes, focusing on the implications for elderly care. The article discussed the benefits of nursing homes, such as access to professional medical care and social activities, but also highlighted the drawbacks, including the loss of independence and the potential for social isolation. It emphasized the importance of considering individual needs and preferences when making decisions about elderly care, and suggested that alternative options, such as home care or assisted living, may be more suitable for some individuals. The article contributed to the ongoing debate about the best approaches to elderly care, advocating for a personalized approach that respects the dignity and autonomy of the elderly.

3. PROPOSED METHODOLOGY

Step 1: MOX_Gas Dataset

The MOX_Gas dataset is a collection of sensor readings obtained from metal-oxide (MOX) gas sensors. These sensors measure environmental parameters such as temperature, humidity, and gas concentration levels. The dataset contains multiple features that provide insights into gas sensor behavior under varying atmospheric conditions. This dataset is used as the foundation for training machine learning models to predict humidity levels. The data is loaded into a structured format, typically a CSV file, and then analyzed for further processing.

Step 2: Data Preprocessing (Handling Null Values, Description, Unique Values)

Once the dataset is loaded, the next step is data preprocessing, which ensures that the data is clean and structured for training machine learning models. The first step is identifying and handling missing values, as null values can impact model performance. Duplicate records are also removed to prevent redundancy. Statistical analysis of the dataset is performed to understand feature distributions, summary statistics, and unique values in each column. Features that are irrelevant or highly correlated are either transformed or removed. Standardization techniques such as feature scaling are applied to ensure consistency across numerical attributes. This preprocessing step is crucial for improving model accuracy and efficiency.

Step 3: Existing Ridge Regressor

Ridge Regression is implemented as a baseline model for predicting humidity levels. Ridge Regression is a type of linear regression that introduces a penalty term to prevent overfitting. The algorithm minimizes the squared error while adding a regularization parameter to reduce model complexity. This helps in handling multicollinearity and improving generalization performance. The Ridge model is trained on the preprocessed dataset, and once trained, it makes predictions on the test dataset. The performance of the Ridge model is evaluated using standard regression metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R^2 score.

Step 4: Proposed Decision Tree Regressor

A Decision Tree Regressor is introduced as an improved predictive model for humidity level estimation. This algorithm works by recursively splitting the dataset into subsets based on the feature that maximally reduces variance in the target variable. Unlike Ridge Regression, which assumes a linear relationship, Decision Tree Regression can capture non-linear dependencies, making it a powerful choice for gas sensor data. The model is trained using the training dataset and validated using test data. The Decision Tree model is expected to perform better in handling complex patterns and outliers in gas sensor readings.

Step 5: Performance Comparison Graph

To compare the efficiency of both models, a performance evaluation graph is plotted. The graph visualizes various performance metrics such as MAE, MSE, RMSE, and R^2 score for both Ridge Regression and Decision Tree Regression. The scatter plots of actual versus predicted values help in understanding which model aligns better with the ground truth. The Decision Tree model is expected to demonstrate lower errors and higher predictive accuracy compared to Ridge Regression. This step highlights the strengths and weaknesses of both models and justifies the selection of Decision Tree Regression for final predictions.

Step 6: Prediction of Output from Test Data with Decision Tree Regressor Algorithm Trained Model

The final step involves using the trained Decision Tree Regressor model to make predictions on new test data. The test dataset is loaded, and necessary preprocessing transformations are applied to ensure consistency with the trained model. The trained model then generates predictions for humidity levels, which are appended to the dataset for further analysis. The predicted values are displayed in a structured format, allowing for validation against expected outcomes. This final step demonstrates the real-world applicability of the developed model in forecasting humidity levels using MOX gas sensor data as shown in figure 1.

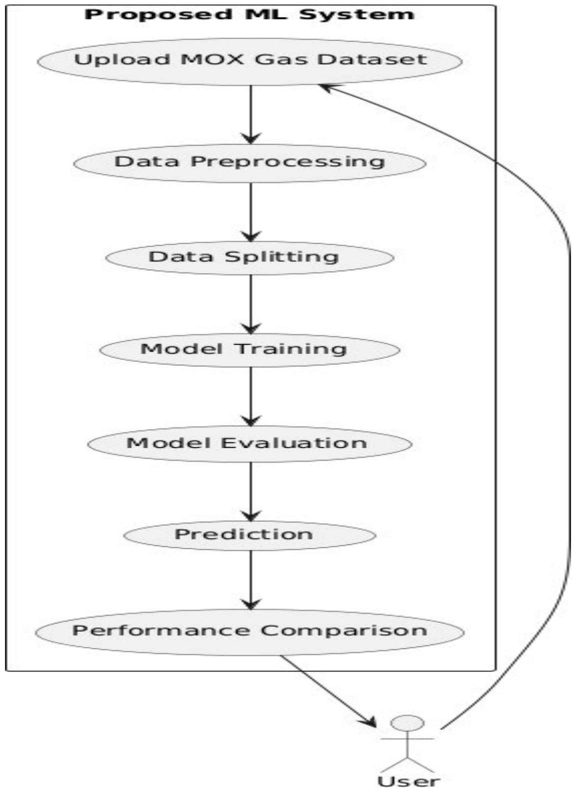

Fig. 1: Architectural Diagram

4. EXPERIMENTAL ANALYSIS

The figure 2 illustrates the process of uploading the MOX gas dataset into the graphical user interface (GUI). It displays the dataset loading mechanism and presents an overview of the uploaded data, including sensor readings, environmental conditions, and operational parameters. The interface enables users to inspect dataset attributes, ensuring that all relevant columns, such as CO concentration, temperature, humidity, and sensor resistance values, are correctly loaded before proceeding with further analysis.

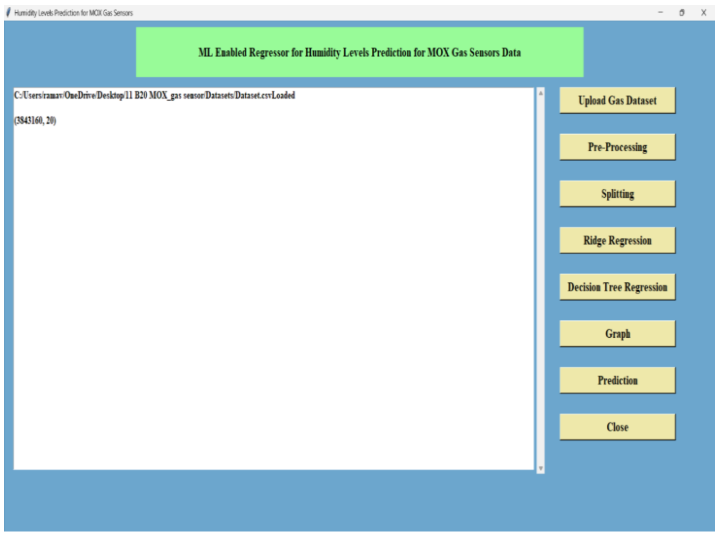

Fig. 2: Upload of MOX Gas Dataset and Its Analysis in the GUI Interface

ISSN NO: 9726-001X

Volume 13 Issue 02 2025

JOURNAL OF CURRENT SCIENCE

-------Pre Processing-----

0	Time (s)	CO (ppm)	Humidity (%r.h.)	...	R12 (MOhm)	R13 (MOhm)	R14 (MOhm)
count	3.843160e+06	3.843160e+06	3.843160e+06	...	3.843160e+06	3.843160e+06	3.843160e+06
mean	4.545202e+04	9.900479e+00	4.600410e+01	...	2.629418e+01	2.328358e+01	2.919879e+01
std	2.624242e+04	6.427533e+00	1.246778e+01	...	1.937179e+01	1.807001e+01	2.261727e+01
min	0.000000e+00	0.000000e+00	1.634000e+01	...	3.270000e-02	3.210000e-02	3.130000e-02
25%	2.272724e+04	4.440000e+00	3.657000e+01	...	9.817900e+00	7.825300e+00	9.804800e+00
50%	4.545279e+04	8.890000e+00	4.709000e+01	...	2.640040e+01	2.204970e+01	2.740270e+01
75%	6.817693e+04	1.556000e+01	5.576000e+01	...	4.046330e+01	3.595290e+01	4.550490e+01
max	9.091013e+04	2.000000e+01	8.381000e+01	...	1.299261e+02	9.820950e+01	1.294220e+02

[8 rows x 20 columns]

Fig. 3: Data Preprocessing in the GUI

The figure 3 represents the data preprocessing steps implemented within the GUI. It includes operations such as handling missing values, feature scaling, data normalization, and train-test splitting. The figure highlights how raw sensor readings and environmental parameters are processed to prepare the dataset for machine learning models. The interface provides options to standardize and clean data, ensuring optimal performance for the regression models.

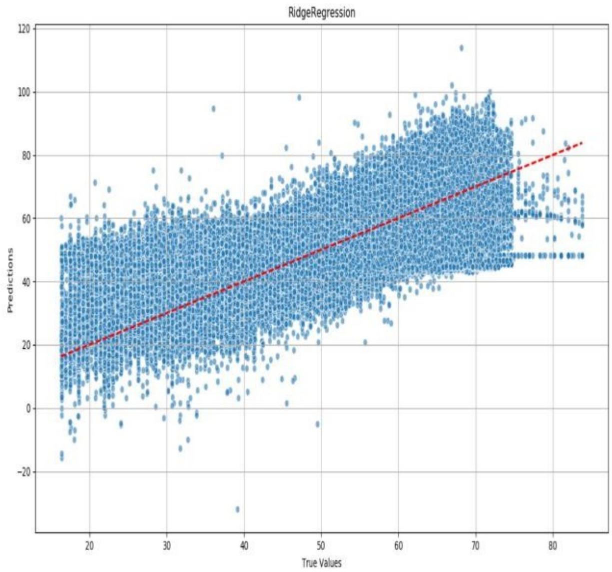

Fig. 4: Performance Metrics and Regression Scatter Plot of Ridge Regression Model

The figure 4 presents the performance metrics of the Ridge Regression model, including:

- Mean Absolute Error (MAE): 7.385
- Mean Squared Error (MSE): 88.496
- Root Mean Squared Error (RMSE): 9.407
- R-squared (R^2): 0.429

Additionally, it includes a scatter plot visualizing the predicted vs. actual values for the Ridge Regression model. The plot provides insights into how well the model generalizes and highlights any deviation from the ideal regression line. The relatively lower R^2 score indicates that Ridge Regression does not capture complex relationships within the dataset effectively.

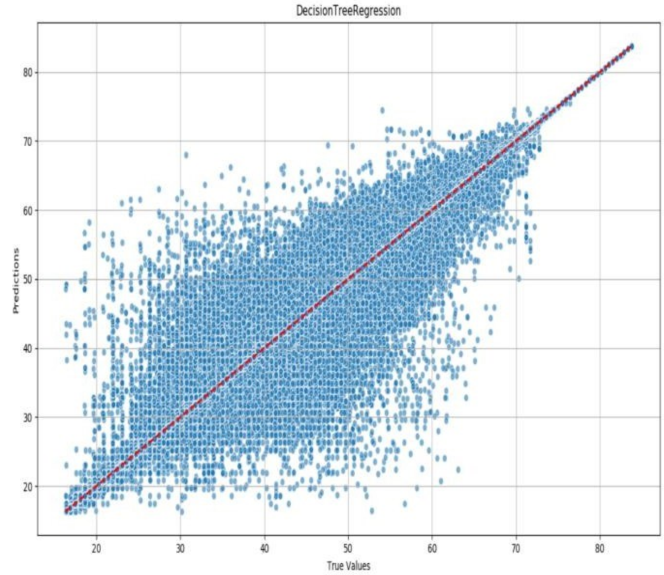

Fig. 5: Performance Metrics and Regression Scatter Plot of Decision Tree Regression Model

The figure 5 displays the performance metrics of the Decision Tree Regression model, including:

- Mean Absolute Error (MAE): 0.469
- Mean Squared Error (MSE): 2.938
- Root Mean Squared Error (RMSE): 1.714
- R-squared (R^2): 0.981

A regression scatter plot is also included, showing the relationship between predicted and actual values. The higher R^2 score (0.981) indicates that the Decision Tree model achieves a much better fit, capturing non-linear patterns and complex dependencies in the dataset.

-------prediction-------

	Time (s)	CO (ppm)	Humidity (%r.h.)	Temperature (C)	...	R12 (MOhm)	R13 (MOhm)	R14 (MOhm)	predection
0	82228.793	4.44	47.56	25.50	...	51.5393	45.0834	56.2820	58.5539
1	35429.873	17.78	56.29	26.22	...	14.7288	11.7522	15.0600	58.5539
2	84252.139	20.00	35.77	22.70	...	15.3540	11.5033	13.8807	58.5539
3	74220.712	6.67	51.65	25.54	...	37.6308	31.9919	40.5330	58.5539
4	84377.264	20.00	35.02	25.98	...	21.0475	16.4073	19.3970	58.5539
..	...	...	...	...	...	...	...	...	...
95	81825.533	0.00	66.38	24.74	...	58.8007	61.0589	75.5478	58.5539
96	76773.660	13.33	36.75	21.90	...	36.3548	30.1552	35.1372	58.5539
97	24788.846	13.33	30.83	27.30	...	0.1039	0.0966	0.1049	58.5539
98	83896.549	20.00	43.50	26.26	...	0.1087	0.1026	0.1075	58.5539
99	40166.502	2.22	17.39	23.94	...	63.1158	67.2546	76.8882	58.5539

[100 rows x 21 columns]

Fig. 6: Model Prediction on the Test Data

The figure 6 presents the predictions made by the trained models on test data. It includes actual vs. predicted values to evaluate how well the models generalize on unseen data. The visualization helps in assessing whether the models produce accurate predictions, particularly for different CO concentration levels and sensor resistance values.

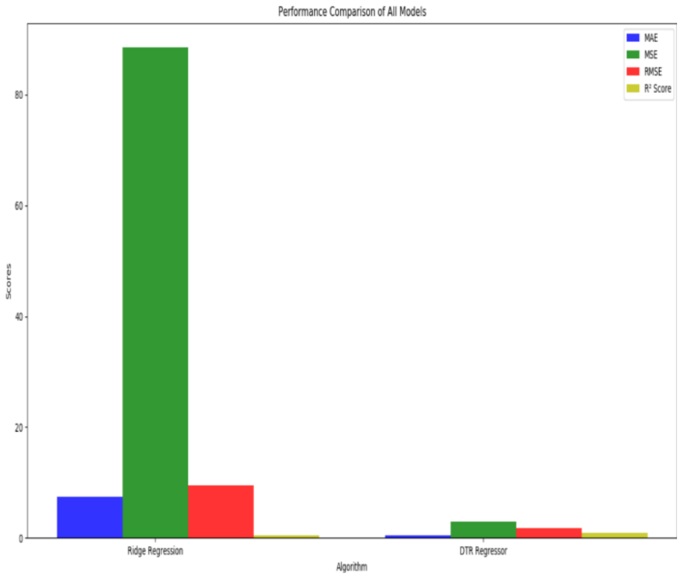

Fig. 7: Performance Comparison Graph of All Models

The figure 7 illustrates a comparative performance analysis of all the regression models used in the project. Metrics such as MAE, MSE, RMSE, and R^2 scores are plotted to highlight the efficiency of each model. The comparison shows that the Decision Tree Regression model outperforms Ridge Regression, achieving significantly lower error values and higher prediction accuracy. The graph provides a clear representation of model performance, guiding the selection of the best regression approach for this dataset.

5. CONCLUSION

This research successfully implemented machine learning models to predict humidity levels using MOX gas sensor data. The Ridge Regression and Decision Tree Regression models were evaluated for their effectiveness in handling the given dataset. Ridge Regression, which applies L2 regularization, helped reduce overfitting but struggled with capturing complex, non-linear relationships. In contrast, the Decision Tree Regressor demonstrated superior performance by effectively learning intricate patterns and interactions between features, leading to improved predictive accuracy. Performance metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared (R2R^2) score indicated that the Decision Tree Regressor outperformed Ridge Regression in this application. The project highlights the importance of selecting appropriate models based on dataset characteristics and problem complexity.

REFERENCES

[1] Thomas, V.S.; Darvesh, S.; MacKnight, C.; Rockwood, K. Estimating the prevalence of dementia in elderly people: A comparison of the Canadian Study of Health and Aging and National Population Health Survey approaches. Int. Psychogeriatr. 2001, 13, 169–175.

[2] Kalache, A.; Gatti, A. Active ageing: A policy framework. Adv. Gerontol. Uspekhi Gerontol. Akad. Nauk. Gerontol. Obs. 2002, 11, 7–110.

[3] World Health Organization (WHO). Are you ready? What You Need to Know about Ageing. Available (accessed on 11 May 2017).

[4] Kulik, C.T.; Ryan, S.; Harper, S.; George, G. Aging Populations and Management. Acad. Manag. J. 2014, 57, 929–935.

[5] World Health Organization. Disability and Health. Available online (accessed on 11 May 2017).

[6] Centers for Disease Control and Prevention. Chronic Disease Overview. Available online (accessed on 11 May 2017).

[7] Ontario. Budget in Brief: Strengthening Health Care. Available online (accessed on 23 May 2017).

[8] Venkat, R. Global Outlook of the Healthcare Industry; Frost & Sullivan: San Antonio, TX, USA, 2015

[9] Population Ageing Projections. Available (accessed on 4 June 2017).

[10] Gelineau, K. Canada Ranks Fifth in Well-Being of Elderly: Study. The Globe and Mail. Available (accessed on 4 June 2017).

[11] Canadian Institute for Health Information. National Health Expenditure Trends, 1975 to 2014. Available online (accessed on 4 June 2017).

[12] The Economist. Working-Age Shift. Available online (accessed on 4 June 2017).

[13] Anderson, G.; Knickman, J.R. Changing the Chronic Care System to Meet Peoples Needs. Health Aff. 2001, 20, 146–160

[14] Advantages & Disadvantages of Nursing Homes. AmeriGlide Stair Lifts and Vertical Platform Lifts. Available online accessed on 11 May 2017).