| **CSE355 - DATA SCIENCE PROFESSIONAL CERTIFICATION****CA4 - PROJECT REPORT****LIFE EXPECTANCY PREDICTION USING ARTIFICIAL NEURAL NETWORKS****BY****E0122015 - SAIRAM.R** |
| --- |
| **TABLE OF CONTENTS**  | **S.NO** | **SECTION** | **PAGE NO.** | | --- | --- | --- | | **1.** | **EXECUTIVE SUMMARY** | **3** | | **2.** | **INTRODUCTION** | **3** | | **3.** | **LITERATURE REVIEW** | **4** | | **4.** | **METHODOLOGY** | **4** | | **5.** | **RESULTS** | **5** | | **6.** | **DISCUSSION** | **6** | | **7.** | **CONCLUSION** | **6** | | **8.** | **REFERENCES** | **7** | | **9.** | **ACKNOWLEDGMENTS** | **7** | |

| **EXECUTIVE SUMMARY** This project investigates the use of Artificial Neural Networks (ANNs) to predict life expectancy using a rich global dataset that includes socioeconomic, health, and demographic indicators. After preprocessing, normalization, and encoding, a sequential ANN model was trained and evaluated. The model achieved a promising level of accuracy and offers a scalable solution for public health forecasting. **INTRODUCTION** Life expectancy is a critical measure of public health and societal progress. Predicting life expectancy accurately can support policy-making, healthcare investments, and targeted interventions. This project uses machine learning—specifically, an ANN—to model complex, nonlinear relationships among factors such as GDP, immunization rates, alcohol consumption, and healthcare expenditures. |
| --- |

| **LITERATURE REVIEW** Numerous studies have explored statistical models and regression-based methods to estimate life expectancy. Traditional approaches often fall short in capturing nonlinear dependencies. Recent advancements in deep learning have shown improved performance in similar prediction tasks. This project aims to build upon these insights by applying an ANN to predict life expectancy using WHO-provided global health data. **METHODOLOGY****Dataset**  * Source: Kaggle - Life Expectancy Data * Records: 2,938 rows * Features: 22 (e.g., Country, Year, Status, Adult Mortality, BMI, GDP, etc.)  **Steps Followed**  1. **Data Cleaning**: Handled missing values using mean/mode imputation. 2. **Encoding**: Label encoding applied to categorical variables like 'Status'. 3. **Feature Scaling**: MinMaxScaler used to normalize all features. 4. **Model Building**: ANN using Keras Sequential API.    * Layers: Input → Hidden (ReLU) → Hidden (ReLU) → Output (Linear)    * Loss Function: Mean Squared Error    * Optimizer: Adam    * Epochs: 100    * Batch Size: 32 5. **Train-Test Split**: 80% training, 20% testing.  **RESULTS** **Train Loss**: ~0.0025  **Test Loss**: ~0.0028  **Mean Absolute Error (MAE)**: ~0.041  **Visualization**: Plots show good convergence between predicted and actual life expectancy values.  The ANN model generalizes well and shows no signs of overfitting. |
| --- |

| **DISCUSSION** The model demonstrates that deep learning can effectively capture the complex relationships in health and demographic data. Factors like GDP, adult mortality, immunization rate, and schooling were found to be highly correlated with life expectancy. The neural network approach offers flexibility and robustness over traditional regression models. **CONCLUSION** The ANN-based model achieves strong predictive performance and supports its use in public health analytics. The project confirms that neural networks can model life expectancy accurately using real-world data. Future work can focus on LSTM or ensemble techniques and incorporating more temporal or geospatial features. |
| --- |

| **REFERENCES**   1. Tawsifur Rahman, "Life Expectancy Data", Kaggle Dataset.  https://www.kaggle.com/datasets/kumarajarshi/life-expectancy-who 2. Chollet, François. Deep Learning with Python. Manning Publications, 2018. 3. WHO Global Health Observatory (GHO) Data 4. Geron, Aurélien. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow. O’Reilly, 2019.  **ACKNOWLEDGMENTS** Thanks to the course instructors and peers for their guidance.  Gratitude to the creators of the dataset and open-source libraries like scikit-learn, pandas, and matplotlib. |
| --- |