**Project Report: SDG 3 - Good Health and Well-being**

**Project Title:** Lung Cancer Prediction Model

* **Objective :**

The primary objective of this project is to develop a predictive model for lung cancer, which aligns with Sustainable Development Goal (SDG) 3: Good Health and Well-being. Lung cancer remains one of the most severe and prevalent health conditions worldwide, leading to significant mortality rates. With early detection, the chances of successful treatment and recovery improve considerably, thereby enhancing the quality of life for individuals at risk. This project focuses on using machine learning (ML) to analyze patient data and provide a reliable prediction of lung cancer occurrence. The model serves as a supportive tool for healthcare practitioners, allowing them to proactively identify high-risk patients and offer preventive interventions.

The predictive capabilities of ML in healthcare can greatly aid in overcoming some of the challenges in disease prevention and management. By contributing to early detection and risk assessment, this lung cancer prediction model has the potential to be part of larger health systems, where it can complement traditional diagnostics, support resource allocation, and inform personalized patient care strategies. This work specifically addresses the need for accessible, data-driven solutions in preventive healthcare, which is essential for achieving universal health coverage as outlined in SDG 3.

* **Sources :**

For this project, the lung cancer dataset was sourced from Kaggle, a widely recognized platform for data science and machine learning datasets. Kaggle provides a repository of datasets in various fields, including healthcare, which makes it a valuable resource for projects focused on developing ML models. This dataset includes a comprehensive set of attributes associated with lung cancer risk, such as demographic information, lifestyle factors, and medical history. By selecting data from Kaggle, the project benefits from a dataset with a robust structure and sufficient detail to build an accurate prediction model.

**Data Preprocessing:**

To ensure accuracy, the dataset underwent several preprocessing steps. This included data cleaning, handling missing values, and feature encoding. Encoding transformed categorical variables into a numerical format, which was necessary for the ML model to process the data effectively. Finally, the dataset was split into training and testing sets to facilitate an unbiased evaluation of the model’s performance.

* **Methodology :**

The methodology focused on selecting a model that would yield high accuracy while minimizing the risk of overfitting. A range of models was initially considered and tested, including Logistic Regression, K-Nearest Neighbors (KNN), and Support Vector Machine (SVM) classifiers. The purpose of testing multiple models was to determine which algorithm best captures the underlying patterns in the data, given the nuances of lung cancer prediction. After thorough testing, **Logistic Regression** emerged as the best-performing model, achieving an accuracy score of approximately 92%.

**Data Transformation and Splitting**:

* **Encoding**: Label encoding was applied to categorical variables, ensuring the dataset could be effectively processed by the chosen model.
* **Training and Testing Sets**: The dataset was divided into training and testing subsets, with an 80-20 split. This structure allowed for both model training and a reliable evaluation on unseen data, ensuring that the final model generalizes well to new cases.

**Model Evaluation**:

The performance of the Logistic Regression model was evaluated using a variety of metrics, primarily accuracy, to determine how well it predicts lung cancer risk. Logistic Regression was chosen because it provided the highest accuracy among the models tested, making it an optimal choice for our goal of early disease detection. Accuracy scores were further validated using confusion matrix visualizations, which helped confirm the model’s reliability and effectiveness in classifying high-risk individuals accurately.

* **Focus on SDG :**

This project is directly aligned with **SDG 3: Good Health and Well-being**, which emphasizes ensuring healthy lives and promoting well-being for all. Chronic diseases, such as lung cancer, have become a major public health concern, impacting individuals, families, and healthcare systems worldwide. The burden of chronic diseases underscores the importance of early diagnosis and prevention, which are fundamental to reducing mortality rates and enhancing quality of life.

**Contributions to SDG 3**:

By providing a predictive tool for lung cancer, this project supports the aims of SDG 3 in several ways:

* **Early Diagnosis**: Detecting lung cancer at an early stage can significantly improve treatment outcomes. This ML model serves as an early-warning system, helping healthcare providers identify individuals at high risk before the disease progresses.
* **Preventive Care**: In addition to identifying at-risk patients, the model can be incorporated into broader health strategies aimed at preventive care. By focusing on early detection, healthcare practitioners can offer targeted interventions that promote healthier lifestyles and potentially reduce the incidence of lung cancer.
* **Support for Healthcare Systems**: The model aids in resource management by enabling healthcare professionals to focus their efforts on high-risk populations. This helps optimize the allocation of healthcare resources, ensuring that preventive care reaches those most in need.

**Long-Term Impact:**

This project not only provides a practical solution for early lung cancer prediction but also demonstrates the potential of AI and machine learning in addressing global health challenges. The model exemplifies how technological innovation can be leveraged to achieve SDG 3 by supporting early disease detection, informed healthcare decision-making, and personalized patient care. These contributions collectively enhance health outcomes and promote well-being, which are core tenets of SDG 3**.**