# Task 1 : Classification

## Objective:

This project challenges you to develop a robust machine learning model to predict survival outcomes for cancer patients. Using a dataset containing patient demographics, diagnosis details, treatment history, and various examination results, your primary goal is to accurately classify the label(Survival_Status) for each individual, where 1 signifies 'Alive' and 0 signifies 'Deceased'.

## Process Overview:

Success in this project hinges on a methodical approach encompassing data exploration, preparation, modeling, and evaluation.

1. **Data Exploration & Understanding:** Begin by thoroughly analyzing the dataset. Investigate the relationships between features (like Birth_Date, Weight, Cancer_Type, Stage_at_Diagnosis, Treatment History, etc.) and the Survival_Status target variable to gain insights into factors influencing patient outcomes.
2. **Preprocessing & Feature Engineering:** Effective data preparation is crucial for building a high-performing model. Apply appropriate techniques such as handling missing values, encoding categorical features (Urban_Rural, Occupation, etc.), scaling numerical data, managing date information, and potentially engineering new features to better capture predictive signals.
3. **Model Development & Selection:** Experiment with various binary classification algorithms suitable for this task (e.g., Logistic Regression, SVM, Decision Trees). Compare their performance using appropriate validation strategies and select the model that yields the highest **Accuracy** on your validation set.
4. **Comprehensive Evaluation:** While **Accuracy** is the primary metric for final scoring, you must also calculate and log **Precision, Recall, and F1-Score** during your model development process. These metrics provide a more nuanced understanding of your model's strengths and weaknesses.

## Prediction & Submission:

Your final, trained model will be used to predict the Survival_Status for a separate test dataset where the true labels are withheld.

- **Output:** Submit your predictions as a **CSV file**.
- **Format:** The CSV must contain exactly two columns:
    - id: The unique identifier for each test instance.

- label: Your model's predicted value (0 or 1).

## Dataset Overview:

| Feature Name | Description |
|---|---|
| **Birth_Date** | The patient's date of birth. |
| **Weight** | The patient's weight measurement. |
| **Height** | The patient's height. |
| **Urban_Rural** | Indicates whether the patient lives in an urban or rural area. |
| **Occupation** | The patient’s profession. |
| **Insurance_Type** | The type of health insurance the patient holds. |
| **Family_History** | Indicates whether there is a family history of cancer. |
| **Cancer_Type** | Specifies the type of cancer diagnosed in the patient. |
| **Stage_at_Diagnosis** | Describes the cancer stage at the time of diagnosis. |
| **Diagnosis_Date** | The date when the cancer diagnosis was made. |
| **Symptoms** | The reported symptoms at diagnosis or during disease progression. |
| **Tumor_Size** | The size of the tumor. |
| **Surgery_Date** | The date on which surgery was performed. |
| **Chemotherapy_Drugs** | Lists the chemotherapy drugs administered to the patient. |
| **Radiation_Sessions** | The count of radiation therapy sessions the patient received. |

| | |
|---|---|
| **Immunotherapy** | Indicates whether the patient underwent immunotherapy treatment. |
| **Targeted_Therapy** | Indicates whether the patient received targeted therapy, designed to attack specific cancer cells. |
| **Recurrence_Status** | Indicates whether the cancer has recurred following initial treatment. |
| **Smoking_History** | Information about the patient's smoking habits. |
| **Alcohol_Use** | Details regarding the patient's alcohol consumption. |
| **label(Survival_Status)** | The current survival outcome of the patient (e.g., 1(alive) or 0(deceased)). |

To participate in the competition related to this task Click on the competition link: Kaggle Competition.

Good luck with the competition!