***Titanic - Machine Learning from Disaster***

**Probelem Statement**

The sinking of the Titanic is one of the most infamous shipwrecks in history.

On April 15, 1912, during her maiden voyage, the widely considered “unsinkable” RMS Titanic sank after colliding with an iceberg. Unfortunately, there weren’t enough lifeboats for everyone onboard, resulting in the death of 1502 out of 2224 passengers and crew.

While there was some element of luck involved in surviving, it seems some groups of people were more likely to survive than others.

In this challenge, we ask you to build a predictive model that answers the question: “what sorts of people were more likely to survive?” using passenger data (ie name, age, gender, socio-economic class, etc).

**Statement Analysis**

* Supervised Machine Learning Problem.
* The Target Value is **Survived**.

**Goal**

Using this model, We will try to understand what sort of people were more likely to **survive** than others.

**Libraries:**

**os**

**numpy**

**pandas**

**matplotlib**

**seaborn**

**warnings**

**sklearn**

**Coounter**

**Collections**

**About Dataset**

This dataset is named Titanic Survival. The dataset contains a set of **891** records under **12** attributes:

**Column Name and their Description**

* **PassengerId**: Unique passenger ID.
* **Name** : name of passenger.
* **Pclass** : Ticket class (A proxy for **socio-economic status** (SES)).
* **Sex** : gender of passenger.
* **Age** : age of passenger.
* **SibSp**: number of siblings/spouse of passenger.
* **Parch** : number of parents of passenger.
* **Ticket** : ticket number.
* **Fare** :Passenger fare.
* **Cabin** : Cabin number.
* **Embarked** : Port of Embarkation.

**Variable Notes**

**Pclass**: A proxy for socio-economic status (SES)

* 1st = Upper
* 2nd = Middle
* 3rd = Lower

**Age**: Age is fractional if less than 1. If the age is estimated, is it in the form of xx.5

**sibsp**: The dataset defines family relations in this way...

* **Sibling** = brother, sister, stepbrother, stepsister
* **Spouse** = husband, wife (mistresses and fiancés were ignored)

**parch**: The dataset defines family relations in this way...

* **Parent** = mother, father
* **Child** = daughter, son, stepdaughter, stepson
* Some children travelled only with a nanny, therefore **parch=0** for them.

### **Data Exploration**

Lets combine both datasets since both are similier Below are the steps involved to understand, clean and prepare your data for building your predictive model:

* Missing values treatment
* Outlier treatment
* Variable Identification
* Univariate Analysis
* Bi-variate Analysis
* Variable transformation
* Variable creation

**Modeling:**

**Decision Tree Classifier:**