**Exploratory Data Analysis and Machine Learning on Titanic Disaster Dataset**

RMS Titanic was a British cruise ship said to be the largest cruise ever made in the history of world. It collided with an iceberg during its maiden journey across the Pacific Ocean from Southampton to New York City. With more than 2200 passengers on board, nearly half of them died after the unprecedented mishap. Of the estimated 2,224 passengers and crew aboard, more than 1,500 died, making the sinking one of modern history's deadliest peacetime commercial marine disasters. RMS Titanic was the largest ship afloat at the time she entered service and was the second of three Olympic-class ocean liners operated by the White Star Line. The infamous incident compels researchers to dig into the dataset. This research is aimed at achieving an exploratory data analysis and understand the effect or parameters key to the survival of a person had they been on the ship. The survival prediction has been done by applying various algorithms like Logistic Regression, K - nearest neighbors, Support vector machines, Decision Tree. Towards the end, accuracies of the algorithms based on features fed to them has been compared in a tabular form.

**EXISTING SYSTEM:**

This dataset has been studied and analyzed using various machine learning algorithms like Naive bayes, Linear Regression etc. Various languages and tools are used to implement these algorithms including Weka, R, and Java etc. The approach existing methods are centered on R and java for executing algorithms- Nave Bayes, Linear Regression. The prime objective of the research is to analyze Titanic disaster to determine a correlation between the survival of passengers and characteristics of the passengers using various machine learning algorithms.

**DISADVANTAGES OF EXISTING SYSTEM:**

* According to Eric Lam and Tang used the Titanic problem to compare the algorithm accuracy which they took the feature of gender. There are no significant differences in accuracy between the their methods they used (naïve Bayes and Linear Regression)
* Based on likelihood of the survival of passengers.
* The most important conclusion provided by them is that more features utilized in the models do not necessarily make results better.
* **Algorithm**: Naive bayes, Linear Regression.

**PROPOSED SYSTEM:**

Supervised Learning or predictive modeling uses an atomic or group of columns to predict a variable. The predictive model’s purpose is to allow the programmer to predict a value that may be unknown or predict something that might happen in the future. The target variable in this case is ‘survived’ which is a binary value. This value is also called a categorical variable where we will use classification techniques. Had the variable been continuous, we would have used regression techniques to find a particular value for a given unknown value. The nature of the target variable is what led to the first predictive model to be Logistic regression.

**ADVANTAGES OF PROPOSED SYSTEM:**

* Reasons behind people who survived and who they are.
* The economic class have to do with the chances of survival
* The gender of the person influence the chances of survival.
* Age of someone a crucial parameter in determining the survival.

**Algorithm**: Logistic Regression, Decision Tree, Decision Tree with Hypertuning, K - Nearest Neighbors and Support Vector Machines

**SYSTEM REQUIREMENTS:**

**HARDWARE REQUIREMENTS:**

* System : Intel Core i5.
* Hard Disk : 1 TB.
* Monitor : 15’’ LED
* Input Devices : Keyboard, Mouse
* Ram : 8 GB.

**SOFTWARE REQUIREMENTS:**

* Operating system : Windows 10.
* Coding Language : Python
* Tool : PyCharm, Visual Studio Code
* Database : SQLite

**REFERENCE:**

Karman Singh, Renuka Nagpal, Rajni Sehgal Department of CSE Amity University, Uttar Pradesh India " **Exploratory Data Analysis and Machine Learning on Titanic Disaster Dataset** " International Conference on Cloud Computing, Data Science & Engineering Date Added to IEEE Xplore: 09 April 2020 INSPEC Accession Number: 19515191 DOI: 10.1109/Confluence47617.2020.9057955