***Proposal 1***

***The titanic problem on kaggle (***<https://www.kaggle.com/c/titanic/data>***)***

**Research Question to be addressed:-**

“Analysis of what sorts of people were likely to survive. In particular, we ask you to apply the tools of machine learning to predict which passengers survived the tragedy.”

**Problem Description :-**

“The sinking of the RMS Titanic is one of the most infamous shipwrecks in history.  On April 15, 1912, during her maiden voyage, the Titanic sank after colliding with an iceberg, killing 1502 out of 2224 passengers and crew. This sensational tragedy shocked the international community and led to better safety regulations for ships.

One of the reasons that the shipwreck led to such loss of life was that there were not enough lifeboats for the passengers and crew. Although there was some element of luck involved in surviving the sinking, some groups of people were more likely to survive than others, such as women, children, and the upper-class.” ( Kaggle)

**Data Set Description :-**

The data has been split into two groups:

* training set (train.csv)
* test set (test.csv)

The training data set is used to develop the model and the test data set is used to test the model performance

**Notes :-**

**pclass**

describes the class for example 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.

***PROPOSAL 2***

***Instacart :- Market Basket Analysis (Kaggle Competition)***

**Company Description:**

“Whether you shop from meticulously planned grocery lists or let whimsy guide your grazing, our unique food rituals define who we are. Instacart, a grocery ordering and delivery app, aims to make it easy to fill your refrigerator and pantry with your personal favorites and staples when you need them. After selecting products through the Instacart app, personal shoppers review your order and do the in-store shopping and delivery for you” – Instacart

**Note:-** Rules of associations can be used here such as apriori algorithm. To check frequently ordered item sets.

**Problem Statement:**.

Predict which products will a user buy again for the first time or add to the cart during a separate session.

**Link:**

https://www.kaggle.com/c/instacart-market-basket-analysis

**File Descriptions:**

**Aisles.CSV :-**

Contains a unique Aisle id and the description of the various aisles in a store.

Example :- [ 1, prepared soup salad ]. Where aisle id is one and prepared soup salad is the aisle description

**Department.CSV :-**

This file contains the department names and their unique id

**Order\_Products\_\_\*.csv :-**

These files specify which products were purchased in each order. Order\_Products\_\_Prior.csv contains previous order contents for all customers. 'Reordered' indicates that the customer has a previous order that contains the product.

Note:- That some orders will have no reordered items. You may predict an explicit 'None' value for orders with no reordered items.

**Orders.csv :-**

This file tells to which set (prior, train, test) an order belongs. Prediction is for reordered items only for the test set orders. 'order\_dow' is the day of week.

**Products.csv :-**

This file contains the product\_id,product\_name,aisle\_id,department\_id

***Proposal 3***

***Mercedes Benz kaggle competition (https://www.kaggle.com/c/mercedes-benz-greener-manufacturing/data)***

**Problem statement:-**

Reduce the time the car spends on the test site . Optimizing the speed of their testing system for so many possible feature combinations with a concise algorithmic approach.

**Problem Description :-**

“Tackle the curse of dimensionality and reduce the time that cars spend on the test bench. work with a dataset representing different permutations of Mercedes-Benz car features to predict the time it takes to pass testing. Algorithms that will contribute to speedier testing, resulting in lower carbon dioxide emissions without reducing Daimler’s standards “

**Data Set description :-**

The dataset contains an anonymized set of variables, each representing a custom feature in a Mercedes car. For example, a variable could be 4WD, added air suspension, or a head-up display. The ground truth is labeled ‘y’ and represents the time (in seconds) that the car took to pass testing for each variable.

* train.csv - the training set to develop a concise model
* test.csv - the test set, one must predict the 'y' variable for the 'ID's in this file and test your model