STATEMENT OF WORK

NATURE OF PROJECT:	ML Challenge
PROJECT REFERENCE NAME:	*Cancer Detection*

Section 1.

Description of Services to be completed by:

1. Project: ***Cancer Detection***

   Breast Cancer Wisconsin (Diagnostic)

   Human resource needs

   ● Python developer
   ● Data Scientist
   ● MLOps

   The project should be completed in several stages:

   - **Stage 1** Download data set
     - **Data Set Information**

       This endpoint will receive as input:

       - Attribute
       - Domain

   - **Stage 2** Will create a models

     - **Preprocessing**

       This endpoint will receive as input data set result in a data frame

       - Data info
       - Data describe

- Data values and counts
- Nulls and others characters
- Data normalization

- **Models**

  This endpoint will receive as input parameters

  - Construction models
  - End Date (date time)
  - Taring model
  - Cross validation
  - Test model

- **Metrics**

  - Precision
  - Recall
  - F1-Scores
  - Confusion Matrix
  - ROC

- **Stage 3 Create a pipeline to MLOps**

- The MLOps part will be done with mlflow performing the following tasks

  - Log metrics
  - Model signatures
  - Save the plot and log it as an artifact
  - Tracking url (localhost)
  - Run MLOps

Section 2.

Other Notes and Considerations:

- Use a personal account per team to connect to github
- Best Practices matters
- Code should be continuously uploaded to a list of users will be provided to be added as collaborators to the repo.
- Validate input parameters of all requests
- Exception handling is excepted

Section 3.

Documentation

Expected documentation should be in English and will include:

- Web Scraping
- Model Construction
- MLOps
- Instructions on how to pull the code from github, compile, run the solution and post a request. These instructions are targeted to technical.

Extra documentation (nice to have):

- Architecture Diagrams