### Title: Design and develop a Health Recommendation System

**Problem statement:**

You can find the medical dataset in the link below and/or explore other available datasets:

<https://github.com/DigitalBiomarkerDiscoveryPipeline/Digital_Health_Data_Repository>

You have to design and implement an application using the data sets, which will recommend the persons that are more susceptible to critical medical conditions. The application should roll the names of such persons continuously on the client dashboard. You can use the data set to train the system and pick some not in the training set to test. It is left up to you how you pick necessary features and build the training that creates this recommendation. Or find other approaches to identify those that are likely to develop medical conditions.

The project has to be developed as a Web Application. Front-end technologies can be - HTML, CSS, and JavaScript etc (it can be anything of your choice). The backend server should involve an appropriate database for storing the data. You are free to use any existing backend frameworks like Spring, struts, Hibernate, etc.

For storing the data, you can use any type of database (MySQL, Oracle, NOSQL, etc.).

You may use Apache Tomcat Server, Microsoft Internet Information Server, or any other Web server of your choice for hosting the web application locally.

These are the suggested steps you can follow to develop a typical recommendation system:

Step 1: Setup a Hadoop cluster where the data sets should be stored on the set of Hadoop data nodes.

Step 2: Implement a content-based recommendation system using MapReduce, i.e. given a job description you should be able to suggest a set of applicable courses.

Step 3: Execute the training step of your MapReduce program using the data set stored in the cluster. You can use a subset of the data depending on the system capacity of your Hadoop cluster. You have to use an appropriate subset of features in the data set for effective training.

Step 4: Test your recommendation system using a set of requests that execute in a distributed fashion on the cluster. You can pick a set of 3-5 job descriptions in the data set to show how they are executed in parallel to provide corresponding course recommendations.

**Evaluation rubric and submission material:**

1. Document the design of your logic and approach adopted to develop the system. Include some screenshots of your implementation and Code files with comments for your implementation. Submit a PDF file with details. **(4 marks)**
2. Front-end Implementation **(4 marks)** and Back-endImplementation **(8 marks)**
3. Demonstrate the working of your application. Submit a short video of the execution steps. Do a screen recording that is clear and shows working code. **(4 marks)**

You will not be evaluated for accuracy of your recommendation engine but mainly on the data engineering or storage and system implementation aspects.

**General instructions:**

1. Your submission should be a zip containing PDF document, working code files and the demo video.

2. The name of the assignment file must adhere to the format:  GRP\_<group number>.zip  
3. The document should have full names of the group members along with the BITS Registration no. of each group member.

4. Only one copy of the file to be uploaded by any member with the filename as mentioned in (2) for the entire group. Multiple submissions by the same group will incur a penalty.

5. Plagiarism will be strictly dealt with and if found, will result in cancellation of the Assignment and 0 marks being awarded to all the group members.  
6. The last date of submission will not be extended in any case given evaluation deadlines.