

GROUP PROJECT - BIA 5302

MACHINE LEARNING

MILESTONE 1

INSTRUCTOR: SARAMA K

Once you have decided the MACHINE LEARNING problem you would like to solve, the first step is to deep dive into the hypothesis and find a possible solution to your problem.

MILESTONE 1: PROBLEM STATEMENT, CHOICE OF STATISTICAL METHODS AND DATA COLLECTION

THE OSEMN pipeline is to be followed for the semester project.



Within the problem statement, please make sure you:

- Identify needs within the broader public or a specific business that help define problem/product
- Develop key requirements for data product/solution
- Use various resources to define features/aspects necessary for a minimal viable product/solution
- Articulate the impact the proposed product/solution could have within the specified public/business context

- Formulate the public/business problem/product as a well-defined and coherent analytics problem
- Translate the public/business problem/product into an equivalent analytics problem
- Establish a clear definition of success / outcome you want to achieve.

CHOICE OF MODEL

- Identify the relevant analytical and statistical methods necessary for delivering the product/solution
- Outline a full data science pipeline that a practitioner in the field could reasonably expect to solve/create the relevant problem/product.

DATA COLLECTION

- Collect and manage data to devise solutions to their research question, problem or hypothesis
- Collect ONE dataset to determine if data are appropriate for a given hypothesis and either mention a hyperlink / data file.
- Explain the features of these datasets and mention preferred datasets, why do you prefer using it.
- GO TO MILESTONE 2.