

# CS-5312 Big Data Analytics

## Project Title

Name1 {Roll # }, Name2 {Roll # }, Name3 {Roll # }

## 1 Problem Description

- Write briefly about the problem you're addressing in your project
- Why this problem is so important to be addressed?
- Is this problem un-addressed or you're solving it in some other "better" way?
- Source? How and why did you think of this problem?
- What will be the input and output of the project as a "final product"? This in the context of end-user.

## 2 Motivation

- What is the motivations behind your chosen project?
- Who are the potential audience end-user of the outcome of this project (who can you sell the outcome of this project as a product)? for example universities, hospitals, software developers, HR manager, tourists, news websites...
- What will be the potential applications of your project as a product (and in what domains)?

## 3 (Potential) Tools and Techniques

- What are the potential tools you'll be using? like language(s), toolkit(s) etc.
- What are the techniques you'll use?
- If you're solving the existing (solved) problem with "better" technique(s), how will it be better? Also briefly describe the technique(s) already used?
- If you are using existing technique but for different data (data used in your project), how can you claim that that techniques are well suited for the problem and data? For example the problem described by your fellows in "Electricity Theft Detection" be will solved using the technique of their paper reading (for ref. recall last class).

## 4 Related/Background Work

- What is related/background work? You can give some paper, book chapter even video lecture example.
- Briefly discuss issues/limitations of each existing work, which can improve in your work.

## 5 Your idea and approach

- Briefly describe your idea(s) about for the project
- Think about the limitation and possible extensions of related work

## 6 Data Sets

- Briefly describe all dataset descriptions you're using in project
- Describe the sources, sizes (both on disk and in dimensions), statistics, types, etc.
- If you're using synthetic datasets, what is the generation method?
- If you are doing project selected from some platform like Kaggle, are you intended to add some other datasets except their provided ones? If yes, give the detail of those datasets also.
- Is the data clean? If not what type of data cleaning/preprocessing is required?

## 7 Evaluation Criteria/Metrics

- How will you evaluate your results? (which measures/metrics will be adopted to report goodness of your work)
- How will the evaluation of your work will be compared with the existing/related work?

## 8 Potential Conference/Publication

- Have you done some home work for the potential conference(s)?
- What are the deadlines for paper submission in the relevant conference?
- How is/are the selected conferences well suited for your work?