Lecture #1

What is DS?

* In history:

Counting stars, that is only things people can rely on(?)

Counting crops – how to divide or distribute

* Make machines to describe the phenomena (using data)

Theoretical approaches to drive equations to phenomena

Computational approaches

* Then DS

Both computer science, Math, Business…

Inter-disciplinary, data and task, resources

* Potential of DS

Disease diagnosis – detecting malaria

Drug discovery

Urban planning,

Precision agriculture

Negative: Gender bias, racial bias

DS process:

* Ask an interesting question
* Get the data
* Explore the Data
* Model the Data
* Communicate / Visualize the Results

What we are covering in the class.

Goals:

In theory, learning key machine learning concept, import metrics for evaluation, handling different kinds of data, extracting insights from analysis of the models.

In practice, implement ML and deep learning models using python libraries. Using free online tools and resources for data science.

Impact, solving real-life problems using DS, evaluating the social impacts.

Lecture #2

**What is data??**

* Not digital format (not should be digital)
* Not should be output of sensing device

Factual information used as a basis for reasoning discussion or calculation.

Datum: single piece of information, observation (just seeing one basket ball)

Data ( plural ): multiple observation

Dataset : multiple game recorded

Everything can be data – Facebook, alarm clock set for news, car speed, trips etc,

Question

**Aspects of data: formats, scope, biases, etc:**

Format difficulty – for computer and for human

For computer, something structured are easy to compute- chart, QR code

Comprehensive data – have access to all the data, observation possible,

Sample data,

Biases

* Gallup polls
* IMDb Movie Ratings

Format (structured)

* Plain text
* XML
* JSON

**Asking precise question:**

**Parsing data with Regular Expressions:**

* Cleaned data set
* If the pattern is found, do sth and return