CSE 464 / DATS 501

Introduction to Data Science and Big Data Final Project

Due: 12-MAY-2020 - Submit to COADSYS before the class.

The main purpose of the project is to enable you to apply data science algorithms for real-world tasks by using **Python notebook** and thus to make you qualified in data science.

- Your first task is to choose a project topic. Many topics and datasets can be obtained from the Kaggle website [1], UCI Machine Learning Repository [2], and Stanford Machine Learning Projects [3].

  For example: Predict Future Sales, https://www.kaggle.com/c/competitive-data-science-predict-future-sales

- Once you have identified your project topic and project group (**max 3 people**), please inform (email) your instructor about your project topic, dataset (specifically URL) and project members **until 14th April.**

- It can be useful to look up existing research on relevant topics by searching related keywords on an academic search engine such as: http://scholar.google.com.

Evaluation Criteria

- The technical quality of the work. Are the proposed algorithms or applications clever and interesting?

- Each technique used in the study will be scored.

  - Feature Engineering
  - Exploratory Data Analysis
  - Data Preparation (Data Cleaning, Handling Outliers, etc.)
  - Text Vectorization (CountVectorizer, TfidfVectorizer, Word2Vec, etc)
  - Model Building with at least 3 different machine learning models
  - Ensemble Learning with the models built (Voting Classifier)
  - Parameter Tuning
  - Comparative Performance Analysis (Confusion Matrix, Accuracy, F-Measure, AUC, etc.)

- Significance. (Did the authors choose an interesting or a real" problem to work on, or only a small toy" problem?)
- Only the projects of the students presenting the project and submitting their reports will be graded.

Good Luck!

References

[1] Kaggle website, http://www.kaggle.com

[2] UCI Machine Learning Repository, https://archive.ics.uci.edu/ml/datasets.php

[3] Stanford Machine Learning Projects, http://cs229.stanford.edu/projects.html