Case Study-3: Online News Popularity

Due date: April 5th 2021

Utilize "OnlineNewsPopularity.csv" to perform supervised learning using K-Nearest Neighbor algorithm to predict the number of shares (target variable 'shares') in social networks that demonstrate the popularity of an article published by Mashable in a period of two years.

According to UCI Machine Learning repository (https://archive.ics.uci.edu/ml/datasets/Online+News+Popularity), this dataset summarizes a heterogeneous set of features about articles published by Mashable in a period of two years. Number of Attributes: 61 (58 predictive attributes, 2 non-predictive, 1 goal field)

To-Do 1: Build a decision support system (using *K*NN) to predict the shares. You are free to modify the data including "shares" attribute to a discrete range or any other modifications that you think will be useful.

To-Do 2: If you were the CEO of this company, what do you propose from the lessons you learned through To-Do 1?

Teams presenting in the class: **5, 6, 7, 8** (https://piazza.com/class_profile/get_resource/kjyews4yluh32w/kkhkvngzu654um)

Try to be as creative as possible. If you are the team that is presenting, please prepare a 10-minute presentation which includes important insights that you want to show off!

Every team regardless of whether you are presenting or not should submit a report (max 8 pages) on your work and visualization. Report will be evaluated based on these 5 points presented with clarity:

1. Basic details about the dataset that you worked on;
2. Interesting insights about your data (a visualization is preferred);
3. Explain why the insights you obtained are interesting using *K*NN approach;
4. why you think your proposed insights are important? Is there any information that you recommend to ignore?

5. Approaches you utilized (for example, you used *K*NN with different data processing techniques and values of *k*).
6. What do you propose moving forward as a future work?

For any questions, please contact the instructor via Piazza or email.

Deadline to submit the report: **April 5th 2021 11:59 PM via LMS**