

CPSC 310-01 Project Proposal for Team Carbon (Pierce Fleming)

April 8, 2019

Content to be included in the proposal:

1. Names of team members: Pierce Fleming
2. Project name (you pick!): Team Carbon
3. Dataset description
 1. Describe the contents of the dataset: The dataset contains price data of diamonds along with their attributes.
 2. Where it can be located: <https://www.kaggle.com/shivam2503/diamonds>
 3. Its format: .csv file
 4. Include a brief description of the attributes and what you will use as class information (i.e., what attribute or attributes you will try to predict): Attributes include carat, cut, color, clarity, depth, table, length, width, depth, and price. I will classify on price using equal length bins to make it a classification problem.
4. Implementation/technical merit
 1. Anticipated challenges in pre-processing and/or classification: Anticipated challenges include figuring out what I failed to properly execute on past PA's.
 2. New data mining topic(s) learned: I will learn to use Tensorflow because it's widely used in the data science community and I anticipate it will come in handy in the future for personal/side projects.
5. Potential impact of the results
 1. Why are these results useful? These results are useful because they can help consumers make educated purchases of diamonds and educated transactions when selling their diamonds to get a fair price. Jewelers distributors can also use these results for appraisal to get a fair price on their end when buying diamonds from distributors and/or buying second hand diamonds from consumers. I could possibly monetize it this idea by selling it to corporations or consumers (monetizing is ambitious, I know, just a thought experiment for now).
 2. Who are stakeholders interested in these results? Diamond/Jewelry distributors, consumers, Professor Sprint, myself.