Operationalization means turning abstract conceptual idea into measurable observations:

* Machine Learning Challenge:
  + What are the generated predictions regarding the dataset test-1.json & train-1.json?
* Hypothesis:
  + What is the number going to be regarding citations?
* Testable Hypothesis:
  + How do the data differ from test-1.json and train-1.json?
    - Punctuations/brackets?
    - Integers/floats?
    - Stop words?
  + What variables do we possess? (test-1.json & train-1.json)
    - Quantitative data
      * Discrete (number of citations, year of issue)
      * Continuous (number of citations, references)
    - Categorical data
      * Binary (open access, abstract ..)
      * Nominal (author, fields of study)
      * Ordinal (title name, topics, etc…)
  + Segment out/measure each variable in the columns
    - Freq (hist), min, avg, max citation count?
    - Freq (hist), min, avg, max field of study?
    - Freq (hist), min, avg, max year of issue?
    - …..
* How do the data correlate from test-1.json and train-1.json?
* Operationalized research question:

How well will our prediction model predict the number of citations?