1. **Use the J48 Decision tree learning scheme to analyze the bolts data (bolts.arff without the TIME attribute):**

* **The dataset describes the time needed by a machine to produce and count 20 bolts. (More details can be found in the file containing the dataset.)**
* **Analyze the model produced. What adjustments (if you were to make any) would have the greatest effect on the time to count 20 bolts (i.e. what is the most important/selective attribute/value pair in the tree)?**
* **According to each classifier, how would you adjust the machine to get the shortest time to count 20 bolts?**

To answer these questions and after discussing with instructors the fact that the specs of the system are not completely clear, I will make the following assumptions:

* RUN (as it is an ID, unique attribute) and (as mentioned in question, which makes sense) TIME have been excluded from the analysis.
* I assume that TOTAL number of bolts is pre-set. That is, it is the amount of bolts required to do the pre-established job (probably to fill a box). This means that TOTAL depends on requirements outside of this processing machine. Therefore, this attribute helps us understand the efficiencies and dependencies of the machine, but most likely its value will not be one of the elements to adjust.
* I am unclear about what drives the switching from the first speed to the second speed. Is it NUMBER2 ('the last few bolts') a pre-set amount that drives the switching of speeds? That is, if NUMBER2 is preset to 2, and the total amount of bolts to package is 10, after 8 bolts are on their way to the packing box, the speed changes to the second speed to process the last two bolts? And when NUMBER2 is instead set to 0, if TOTAL = 10, all 10 bolts will be processed at the first speed, which will remain at that speed until all 10 bolts are processed?
* I am assuming that NUMBER2 is part of TOTAL. That is, if TOTAL = 10 and NUMBER2 = 2, then 8 bolts have to be processed at SPEED1 and 2 at SPEED2.
* SENS is the sensitivity of the electronic eye to insure that the correct number of bolts are counted. However, the documentation says "From an engineering standpoint if the correct number of bolts is counted, the sensitivity \*\*\* should \*\*\* not affect the time to count bolts". But if that is the case, what is the point of including this element in our dataset? As it is unclear whether ‘should’ means that in fact SENS does not affect the T20BOLT, I have analyzed this attribute too.

To do the analysis, I discretized the attributes to apply J48. I did some preliminary analysis with rule schemes, applied J48 in all possible attribute combinations, and came to the following suggestions:

* If I was working for that company I would request a much larger sample before making a decisions, as well as a much balanced sample, where for instances the middle value for SPEED1 and TOTAL had as much representation as the extreme values.
* As often increasing machine speed implies additional use of energy (and therefore cost), and SPEED1 seems to produce better final results at lower speeds, I would recommend keeping SPEED1 processing bolts at the lowest speed (2).
* Although the documentation explains that SENS ‘should’ not affect time, I would recommend keeping SENS at the highest level, as it seems to increase the time efficiency (this would have to be reviewed if there is a difference on energy consumption for the optical device that produces SENS and the energy consumption increases as SENS increases).