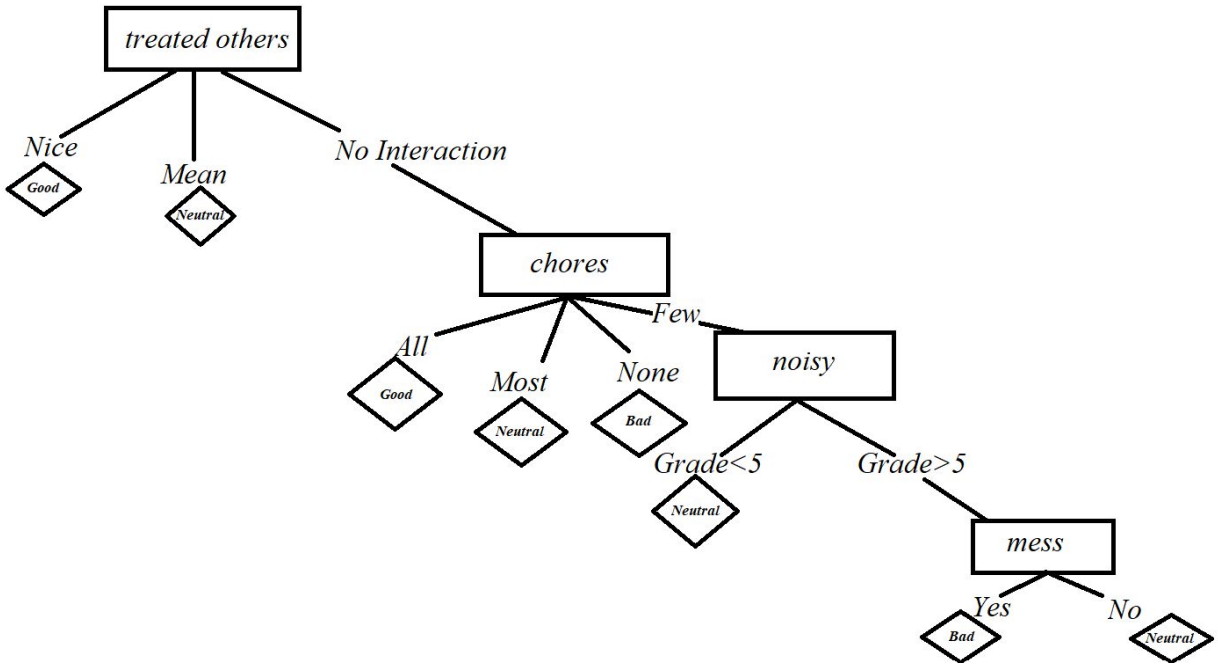
Assignment 4.3

Explain how a possible decision tree for the following problem description may look like. Discuss on choices for features in the decision tree and possible challenges.
Children should be classified based on their behavior into three classes of Good, Neutral and Bad. Each child is evaluated daily on the following points: how many **chores** they fulfilled (All, Most, Few, None), how **noisy** they are (a grade between 1 and 10), whether they made a **mess** (Yes or No) and how they **treated other** children (Nice, Mean, No Interaction). Additionally, for each child you have information about them, meaning their name, birth date, supervisor and the income of the parents.
It has been found out, that the classification of a child mostly depends on their age, their noisiness and how they treated others. Additionally, some supervisors grade children worse than others.



Feature Selection:

- The *'**treated other**'* attribute is selected based upon the observation provided in the problem statement and also due to the possibility of a 3-way split.
- Based upon '*No Interaction*' we select ***'chores'*** as the following attribute with the 4-way split.
- The attribute ***'noisy'*** being a multi-valued attribute, is basically discretized into a 2-way split as *grade<5 & grade>5*.
- Due to *grade>5*, the next attribute ***'mess'*** is recognized as a binary split and the labels are assigned accordingly.

Challenges:

- For attribute *'noisy'* there are possibilities of different multiway splits but, should be restricted to only a 3-way split considering the available classes at hand.
- Due to the aforementioned point, there can be the possibility of the alternative tree leading to an alternate hypothesis.
- It can lead to overfitting as, if the current hypothesis has a larger error as compared to an alternative hypothesis (the solution is to follow *Occam's razor*).
- If unnecessary splits are made then there may be the need for pruning the tree.