Project review

- Team: ISZ_8
- Authors: Karolina Woźniak, Maciej Górnik
- Reviewed by: Gustaw Cyburt, Artur Mzyk

1. Problem formulation [0-5]:
    a. Is the problem clearly stated?
    Yes, the problem is clearly stated [1]
    b. What is the point of creating model, are potential use cases defined?
    There are valid use cases defined [1]
    c. Where does data come from, what does it contain?
    Data comes from web scraped automotive webpage, content description is provided [1]
    d. DAG has been drawn?
    Yes, DAG has been drawn [1]
    e. Was confounding described?
    No, confounding wasn't described [0]
2. Data preprocessing [0-2]:
    a. Is preprocessing clearly described?
    Yes, preprocessing (including data standardization) is well-described [1]
    b. Reasoning and types of actions taken on dataset have been described?
    Yes, reasoning has been described thoroughly [1]
3. Model [0-4]:
    a. Are two different models specified?
    Yes, however they have the same type of distribution (linear) [1]
    b. Is the difference between two models explained?
    Yes, in the second model there is another variable added (mileage) [1]
    c. Is the difference in the models justified (e.g., does adding additional parameter makes sense)?
    Yes, Authors wanted to show how adding highly corelated parameter to the model will affect price estimation [1]
    d. Are models sufficiently described (what are formulas, what are parameters, what data is required)?
    Yes, Stan models are described in the report [1]
4. Priors [0-4]:
    a. Is it explained why particular priors for parameters were selected?
    No, that information wan not provided [0]
    b. Have prior predictive checks been done for parameters (does parameters simulated from priors make sense)?
    Yes, parameters have been checked [1]
    c. Have prior predictive checks been done for measurements (does measurements simulated from priors make sense)?
    Yes, there is a partly overlap between real and simulated data [1]

d. How prior parameters were selected?
The prior parameters were chosen through a semi-empirical process [1]

5. Posterior analysis (model 1) [0-4]:
    a. Were there any issues with the sampling? If yes, what kind of ideas for mitigation were used?
    Authors did not mention about difficulties being present during sampling [0]
    b. Are the samples from posterior predictive distribution analysed?
    Yes, they are described [1]
    c. Is the data consistent with posterior predictive samples and is it sufficiently commented (if they are not then is the justification provided)?
    Yes, it is consistent with briefly description [1]
    d. Has parameters marginal distributions been analysed (histograms of individual parameters plus summaries, are they diffused or concentrated, what can we say about values)?
    No, Authors did not mention about marginal distributions [0]
6. Posterior analysis (model 2) [0-4]:
    a. Were there any issues with the sampling? If yes, what kind of ideas for mitigation were used?
    Authors did not mention about difficulties being present during sampling [0]
    b. Are the samples from posterior predictive distribution analysed?
    Yes, they are described [1]
    c. Is the data consistent with posterior predictive samples and is it sufficiently commented (if they are not then is the justification provided)?
    Yes, it is consistent with briefly description [1]
    Has parameters marginal distributions been analysed (histograms of individual parameters plus summaries, are they diffused or concentrated, what can we say about values)?
    No, Authors did not mention about marginal distributions [0]
7. Model comparison [0-4]:
    a. Has models been compared using information criteria?
    Yes, WAIC and LOO criteria [1]
    b. Has result for WAIC been discussed (is there a clear winner, or is there an overlap, were there any warnings)?
    Yes, it was briefly discussed [1]
    c. Has result for PSIS-LOO been discussed (is there a clear winner, or is there an overlap, were there any warnings)?
    Yes, it was briefly discussed [1]
    d. Was the models' comparison discussed? Do authors agree with information criteria? Why in your opinion one model is better than another
    There is lack of descriptive models' comparison. Model with additional parameter is better. [0]

Overall: 20/27 -> 74%