|  |
| --- |
| QSAR Title |
| Provide the title of the model. The title should include keywords such as: endpoint modelled and name of modeller. One sentence. |
| Software |
| Specify the name and version of the software and libraries that were used implemented the model. Python version and libraries used. Likely ~10 packages. |
| Abstract |
| A free text description of the context and background of the QSAR model including where the data has come from. ~100 words. |
| Date |
| Date |
| Endpoint |
| Endpoint: describe the endpoint that is modelled. One sentence. |
| Endpoint Units |
| Endpoint units: Specify the units of the endpoint measured. |
| Dependent Variable |
| Dependent variable: Specify the relationship between the dependent variable being modelled and the endpoint measured, i.e. any mathematical transformation. One sentence. |
| Type of Model |
| Type of model. One sentence. |
| Explicit Algorithm |
| Explicit algorithm: Describe succinctly the algorithm for generating predictions from the descriptors. Include a reference to a scientific paper where the algorithm is described in detail in the references. Up to ~300 words.  Submit your complete Jupyter notebook to CANVAS. |
| Descriptors in the Model |
| Descriptors in the model: Identify the number and the name or identifier of the descriptors included in the model. A couple of sentences. |
| Software name and version for descriptor generation |
| Software name and version for descriptor and algorithm generation: Specify the name and the version of the software used to generate the descriptors and the algorithm. If relevant, report the specific settings chosen in the software to generate a descriptor or the algorithm. A couple of sentences. |
| Availability of the data |
| Availability of the data, e.g. website, paper etc. and the format such as SMILES for molecules and numeric for the dependent variable. |
| Statistics for goodness-of-fit for training set |
| Statistics for goodness-of-fit statistics for training set including r2 and RMSE. |
| Test set |
| How the test set was generated and how this has avoided data leakage. Couple of sentences. |
| Statistics for goodness-of-fit for test set |
| Statistics for goodness-of-fit statistics for training set including r2 and RMSE. |
| Model descriptor analysis |
| The descriptors that are important in the model. Where possible provide a chemical interpretation and mechanistic interpretation in a biological context. Up to ~200 words. |
| References |
| Reference(s) to main scientific papers and/or software package: List the main bibliographic references (if any) to original paper(s) explaining the model development and/or software implementation. References in RSC format. |
|  |