Response to Reviewer-2:

The manuscript "Tree-Based Machine Learning and Nelder-Mead Optimization for Optimized Cr(VI) Removal by Indian Gooseberry Seed Powder" by L. R. Kalabarige describes a series of statistical/Machine Learning methods to predict the removal of Cr(VI) from wastewater using the crushed powder of an indian seed. The work may be of interest for people working in the field, though it is probably a bit too specialized (but I guess Waters audience is quite specialized as well.

Author Response: The authors would like to thank the Editor and Reviewers for the time they spent reviewing our manuscript. The suggestions provided were constructive to enhance the quality of the manuscript. We have tried to answer all your questions/remarks/suggestions to the best of our knowledge. We hope you find the answers satisfactory and the manuscript of better quality.

Reviewer #2 Concern #1: In some parts of the paper the thread is difficult to follow and should be made clearer. The first question regards the necessity of building a "synthesized database". The experimental dataset shown (Table 1) contains 56*4 = 224 points (=compound + properties, like pH etc) is then converted into a larger one (2000*4) by some kind of extrapolation/generation, that is used for model training.

Response: We are thankful to the reviewer for raising a valid point regarding the necessity of a synthesized dataset when we have an initial dataset of 55 data points.

The experimental data provides valuable insights with 55 instances. However, it has limitations for training robust ML models. The substantial amount of data is required for the effective training of ML models to understand hidden patterns, and complex non-linear relationships between features. Hence, it is difficult to have an effective training of ML models to capture complex patterns, understand full range of interactions between the variables, and achieve optimal model performance. Furthermore, a larger dataset with 2000 instances, allows the model to learn from a broader range of scenarios. This enhances the model's generalizability, meaning it can perform well on unseen data not included in the training set. Moreover, the randomization approach employed to built the synthesized dataset expands the range of feature values beyond those observed in the experiments (Table 1). This enables the exploration of a wider parameter space, potentially leading to the identification of optimal conditions that might not have been captured in the original experimental design.

Reviewer #2 Concern #2: Excessive use of digits (e. g. 99.24284408 mg/L, pH of 4.96268861), is very frequent in the text. They are certainly unphysical. Also, numerical descriptors (like R2) are given with variable numbers of digits.

Response: As reviewer suggested, the excessive and variable length digits after decimal is reduced and made uniform.

Reviewer #2 Concern #3: "...ML modelling and ML based Nelder-Mead optimization reported 11.56% of increase in Cr(VI) removal percentage than the experimentation results for different absorptions which are not tested during experimentation. , tree-based regression models have not been previously employed. The part after THAN is quite unreadable,

Response: As suggested by reviewer, the sentence framing is not appropriate and place for the below paragraphs is not appropriate. Hence, we have corrected the entire paragraph and removed from introduction section and added in result section by removing duplication in writeup.

The removed paragraphs from introduction section are as follows:

Notably, the ETR approach outperformed all other models, showcasing a remarkable R^2 score. Furthermore, the MAE, MSE, and RMSE of the ETR indicated significantly lower error rates compared to the other models. Moreover, the percentage of Chromium (VI) removal for Initial concentration of Cr(VI) of 99.55 mg/L, pH of 4.0, and Adsorbent dosage of 8.0(g/L) gives highest Cr(VI) removal percentage such as 78.21\% and 78.107 respectively by DTR and RFR is as shown in Table 6. The removal efficiency of ETR model was 85.11\% for Initial concentration of Cr(VI) of 99.24284408 mg/L, pH of 4.96268861, and Adsorbent dosage of 9.61957019(g/L). Furthermore, the optimization results presented in Table~\ref{tab6} reported optimized values obtained by ML based Nelder-Mead optimization approach which obtain maximum Cr(VI) removal percentage. That is, the DTR-Nelder-Mead, RFR-Nelder-Mead and ETR-Nelder-Mead optimization approaches were able to obtain maximum of 78.107\% to 85.11\% of Cr(VI) removal which reported increase in Cr(VI) removal percentage ranging from 4.66\% to 11.56\% than the Cr(VI) removal percentage of 73.55\% obtained by experimentation for Initial concentration of Cr(VI) of 20 mg/L, pH of 2.0, and Adsorbent dosage of 8.0(g/L).

In summary, based on the comprehensive evaluation, it is concluded that the ML models exhibited strong performance overall. ETR, in particular, stood out as the top performer, demonstrating superior accuracy with minimal error rate and obtains higher Cr(VI) removal percentage than all other models. Furthermore, The ML based Nelder-Mead optimization results indicated that the maximum of 85.11\% chromium VI removal is possible with Initial concentration of Cr(VI) of 99.24284408 mg/L, pH of 4.96268861, and Adsorbent dosage of 9.61957019 (g/L). These results indicated that the maximum Cr(VI) removal percentage by ML modelling and ML based Nelder-Mead optimization reported 11.56\% of increase in Cr(VI) removal percentage than the experimentation results for different absorptions which

are not tested during experimentation. Furthermore, tree-based regression models have not been previously employed. Hence, this study ventures to apply three different models and ML based optimization to analyze and compare their performance on experimental data.

The corrected paragraph is as follows and placed in result section:

Similarly, the DTR-Nelder-Mead, RFR-Nelder-Mead and ETR-Nelder-Mead optimization approaches were able to obtain maximum of 78.107\% to 85.11\% of Cr(VI) removal which reported increase in Cr(VI) removal percentage ranging from 4.66\% to 11.56\% than the Cr(VI) removal percentage of 73.55\% obtained by experimentation for Initial concentration of Cr(VI) of 20 mg/L, pH of 2.0, and Adsorbent dosage of 8.0(g/L). In summary, based on the comprehensive evaluation, it is concluded that the ML models exhibited strong performance overall. ETR and ETR-Nelder-Mead optimization, in particular, stood out as the top performer, demonstrating superior accuracy with minimal error rate and obtains higher Cr(VI) removal percentage than all other models.

Reviewer #2 Concern #4: Page 4: "than the experimentation results for different absorptions which are not tested during experimentation. , tree-based regression models have not been previously employed."

Response: This matter was corrected and removed from introduction section and kept in result section as mentioned in response to concern3.

Reviewer #2 Concern #5: Please enlarge equation 1

Response: As suggested, it was rewritten as follows and enlarged the equation using its acronyms. The same is highlighted in manuscript with blue color.

Percentage removal of $Cr(VI) = ((a * IC) + (b * pH) + (c * AD))$

Reviewer #2 Concern #6: Obscure writing:

Page 6

“This conversion process not bounded to certain range.”

Response: As suggested by reviewer, it was identified that the above statement “This conversion process not bounded to certain range” doesn’t carry any meaning. Hence, it was removed from the manuscript.

"Finally, tested each model with testing data."

Response: The following paragraph under section 3.0.4 was modified by removing “Finally, tested each model with testing data." The revised manuscript highlighted the modified content in blue color.

The old write-up:

The proposed work applied tree-based regression models (as reported in Figure~\ref{fig1}) on synthesized dataset generated through curve-fitting on experimental data to build and train an ML model. Finally, tested each model with testing data. This work divided synthesized data into two portions such as training (85\%) and testing (15\%).

The modified writeup:

The proposed work employed tree-based regression models (as reported in Figure~\ref{fig1}) to attain higher prediction accuracy for Cr(VI) removal. The synthesized dataset was divided as training (85\%) and testing (15\%) sets.

Thank you again for reviewing our paper and providing valuable suggestions to improve the quality of the paper.