Exploration of process dynamics and parametric sensitivity on Decolorization of dye wastewater using C4.5 classifiers

P.Anantha Narayanan, Ashutosh Das, Mukesh Goel*and P. Parthiban
Centre for Environmental Engineering,
Ponnaiyah Ramajayam Institute of Science and Technology (PRIST),
Thanjavur, Tamil Nadu, India – 613403
Corresponding author:
Dr. Mukesh Goel
Centre for Environmental Engineering,
PRIST University, Thanjavur, Tamil Nadu,
India – 613403

Abstract

The increasing database associated with wastewater treatment systems necessitates their proper utilization to enhance the system characterization, parametric optimization and decision support systems. In the present study, there has been a definitive effort to use the extensive database for adsorbents (i.e., the materials use to retain unto itself) for colouration of dye-wastewater, for tree-based classification (to be specific, C4.5) using Weka-3.8. The adsorbents used in this study are jatropha deoiled cake (JDC) and its pre-treated counterparts (solvent extraction by water, methanol, acid, alkali, hexane and toluene as well as thermal activation). The database, segregated with regard to the adsorbents used for treatment, being thus modelled, was compared with ZeroR model and integrated model for comparison. The C-4.5 classifier used in the study, with optimized pruning and minimum number of objects, was used to generate the tree, used for interpretation of the process domain. It was found that low-pH is the most crucial parameters for increased colour removal in all the best-fitted models (i.e., Methanol-extracted, Toluene-extracted and Activated JDC adsorbents), also secondarily so (next to dosage) even in least-fitted models (Acid, Alkali & Hexane-extracted JDC adsorbents). It was also observed in all the cases that the increase in dosage lead to decrease in the colour removal capacity of the adsorbents, contrary to common expectations. The results obtained in the study are useful not only in understanding and optimizing the system performance, but also provide a means for exploring complex systems in processing and optimization systems.

Keywords:Adsorbent; dye wastewater; treatment; Weka; C 4.5 classifier.

1. Introduction

India’s textiles sector is one of the oldest industries in Indian economy dating back several centuries. Even today, textiles sector is one of the largest contributors to India’s exports with approximately 11 % of total exports. The textiles industry is also labour intensive and is one of the largest employers. In fact, the textile industry employs about 40 million workers and 60 million indirectly. India's overall textile exports during FY 2015-16 stood at US$ 40 billion. As reported by Ministry of Textiles, Indian Textile Journal, Department of Industrial Policy and Promotion, Press Information Bureau(201). Textile industries are found in most countries and their numbers have increased enormously. Textile industries consume substantial volumes of water and chemicals associated with dyeing process. These industries have shown a significant increase in the use of synthetic complex organic dyes as colouring material. Due to the increasing demand for textile products, textile mills and their wastewater have been increasing accordingly, thus causing a major problem of pollution. Various treatment options are available such as biodegradation, adsorption, etc. [1, 2].

Adsorption has been proven to be efficient and cost effective method of chemical removal from aqueous effluents. Reduced sludge generation and better decontamination efficiency are the major advantages of this method. However, conventional adsorbents like activated carbon are quite expensive. Natural materials available in large quantities, few industrial solid wastes that have adsorption potential can be exploited for this purpose. Many solid wastes like rice husk, saw dust, fly ash, bagasse pith, chitin, etc., have been studied for their adsorption potential, and had shown promising results. There had been extensive studies on use of agro-residue as adsorbent[3, 4].

Data mining is increasingly gaining attention amongst scientists and researchers to analyze the wide variety of data generated in present information age. In fact bioinformatics has become intrinsic part of computational biology and genome analysis is being done using bioinformatics. Data generation rate is so fast that large array of data can be generated from a single analysis. This necessitates finding suitable algorithms and tool to decipher the pattern present in such data and use it for prediction. Such prediction will be useful in experimental analysis, can be integrated with a model and can lead to scale up [5].

Data classification is a standard data mining tool to develop the models from given data. Decision tree classifiers, Bayesian classifiers, k-nearest-neighbor classifiers are some of more common classification techniques. Decision tree classification is used a lot because of its simplicity and economic nature [6].

Objective of this work is to develop adsorbent-based classifier systems (C4.5), and evaluate their performance with regard to base-line classifier (zeroR) and the combined classifier. This work also aims at exploring the levels of contribution and threshold values of the process parameters used (namely, pH and dosage).

2. Materials and methods

2.1 Adsorption Studies

Reactive blue dye was obtained from Merck India. Jatropha Deoiled Cake (JDC) was collected, dried for overnight, at 105^{o}C in an electric oven, followed by crushing and sieving to obtain the particles having an average diameter of 125 μm. The sample, thus obtained, was dried, desiccated and preserved for subsequent analysis and kept in air-tight containers. Using various pretreatment, eight kinds of adsorbents were obtained. They are Raw JDC, methanol extracted JDC, hydro extracted JDC, acid extracted JDC , alkali extracted JDC, hexane extracted JDC, toluene extracted JDC and activated JDC. Adsorption experiments were conducted for three different pH's, namely, 2, 7 and 10. Furthermore, adsorbent dosage was varied from 1 g/l to 8 g/l. All the experiments were conducted at the same temperature.

2.2 Data-Mining Classification:

The response variable (i.e., percentage reduction of colour through adsorption) were classified to nominal variable using the following ranges : Lowest (<20%), Lower Medium(≥ 20 to < 40%), Medium (≥ 40 to < 60%), Higher-Medium (≥ 60 to < 80%), and Highest (>80%). Using Weka (3.8), we used the dataset obtained from extensive adsorption experiments (converted to csv-file format) for data-mining classification (Table 1).

Table-1 Real number attribute description

S.No	Attribute	Description	Data Type
1.	pH	Hydrogen ion concentration	Numeric
2.	Dosage	Amount of adsorbent used (gpl)	Numeric
3.	Reduction	% Reduction = Initial concentration- Final concentration/ Initial concentration Range: Lowest, 0-25 (Lower medium), 26-50 (Medium), 51-75 (Higher), 76-100 (Highest)	Textual

ZeroR classifier was used to establish the baseline (using batch size 100) for classification of the colour-reduction. The most common tree-classifier, namely C 4.5 classifier (known as J48 in Weka) was used, with confident factor 0.25 and minimum number of object 5 (to avoid overfitting).The various classifier performances are compared to zeroR and with differential combination of pruning, binary split and collapse time to obtain the most optimized performance parameters, using 10-fold validation

3. Results and Discussions:

As evident from Fig 1, primary factor that determines the colour-removal capacity of an adsorbent is pH in all adsorbent, except in case of alkali & hexane-extracted-JDC's adsorbents, where pH plays a secondary role, next to the dosage). In fact, these are the two of three adsorbents (third one being acid-extracted-JDC adsorbent), who show lower accuracy and higher RMS error than the rest of the adsorbents, in relation to combined modeling using all adsorbents.

In all the cases, lower pH is found to be associated with better colour-removal capacity of the adsorbents. Contrary to common understanding, lower dosage of the adsorbents used are also found to aid better colour removal, probably because of the increase in dosage beyond the threshold leading to suspension thereof, thereby contributing to colour again.

Acid extracted-JDC is the least-accurately classified adsorbent (where precision is lower than that of ZeroR, and is the only adsorbent with lower TP (True Positive) rate than FP (False Positive) rate, lowest precision, lowest recall and lowest F-measure. The next two lowest performing models are the classifiers for Alkali & Hexane-extracted JDC adsorbent's. In all these cases the FP-rates are higher than all other adsorbents, with significantly low precision,

Table 2 : Performance Parameters of Different Pre-Treated Adsorbents using C 4.5 Classifier

Pre-treatment (Jatropha-De-oiled cake)	**Accuracy (%)**	**RMS error**	**TP (True Positive) Rate**	**FP (False Positive) Rate**	**Precision**	**Recall**	**F-Measure**	**ROC**
Control (Raw)	76.19	0.38	0.762	0.125	0.766	0.762	0.761	0.801
Hydro-extracted	71.43	0.38	0.714	0.381	0.619	0.714	0.636	0.555
Methanol-extracted	90.48	0.24	0.905	0.038	0.871	0.905	0.883	0.896
Acid-extracted	42.86*#	0.53	0.429	0.442	0.409	0.429	0.414	0.556
Alkali- extracted	57.14#	0.49	0.571	0.341	0.513	0.571	0.54	0.521
Hexane-extracted	62#	0.43	0.619	0.411	0.5	0.619	0.551	0.595
Toluene-extracted	90.95	0.33	0.81	0.165	0.736	0.81	0.77	0.776
Activated	89.66	0.34	0.667	0.13	0.643	0.667	0.654	0.87

Classifier used- C 4.5

*Accuracy lesser than benchmark classifier (i.e., ZeroR)

Accuracy lesser than combined modeling

recall, F-measure and ROC.

Methanol-extracted, Toluene-extracted and Activated JDC adsorbents, on the other hand, show maximum accuracy, TP’s, precision, recall, F-measure and ROC, with minimum RMS, compared to rest of the adsorbents, where pH ≤ 2 clearly delineates the highest colour removal (Table 2).

Conclusion:

One of the essential criteria responsible for any treatment process is the optimization of the operational parameters and estimation their threshold values. The data-mining tool comes very handy dealing with such systems, in understanding the system performance, using the model performance. Use of a robust classifier (namely C 4.5), taking in account two crucial parameters for colour-removal of wastewater, it was possible to decipher the higher and lower-precision systems, comparing their threshold performance matrices with the base-line modeling (i.e. Zero-R), which is estimated only using the class value (i.e., response variable), without referring to the independent variables. Methanol-extracted, Toluene-extracted and Activated JDC adsorbents are found to be best classified models whereas, Acid, Alkali & Hexane-extracted JDC adsorbents are least classified ones. It was also observed that in all models, lower the pH and lower the dosage, there is preferential increase in colour removal. The study also opens of the expansive scope of use of data-mining in optimization and decision-support systems for process-sectors.

References:

1. J. S. Sudarsan, V. T. Deeptha, Deepak Maurya, Mukesh Goel , K. Ravi Kumar and Ashutosh Das, (2015), Study on Treatment of Electroplating Wastewater Using Constructed Wetland, Nature Environment and Pollution Technology, 14, 95-100.

2. IrfanNausath;Mukesh Goel *, Arul Prakash; Soundaryan R; Alagar Raj; Ravikumar Kanapuram; ThePotential Of Using Low Cost Naturally Available Biogenic Substrates For Biological Removal Of Chlorophenol, accepted for publication in Bioresource Technology

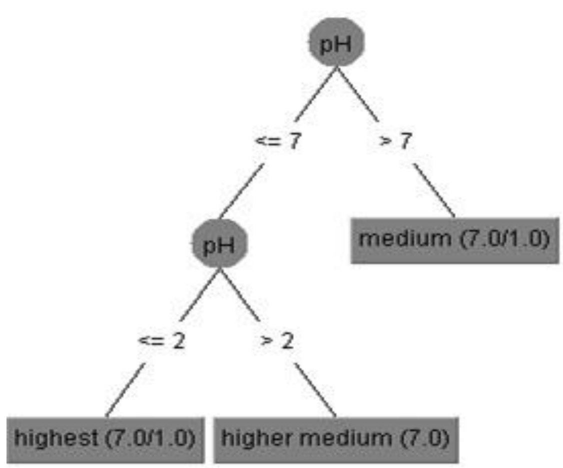
3. Ashutosh Das, B. Subramanyam, Mukesh Goel , B. Ram Kiran, K. Sudharshan, K. Ravi Kumar, (2014), "Alternative Usage of Edible Deoiled Cake for Decolonization of Reactive Red Dye", Desalination and Water Treatment, 1-7.

4. Tzung-I tang,Gang Zheng, Yalou huang,Guangfu Shu,Pengtao wang. A comparative study of medical data classification methods based on decision tree and system reconstruction analysis. IEMS vol.4,no.1,pp-102-108,june 2005

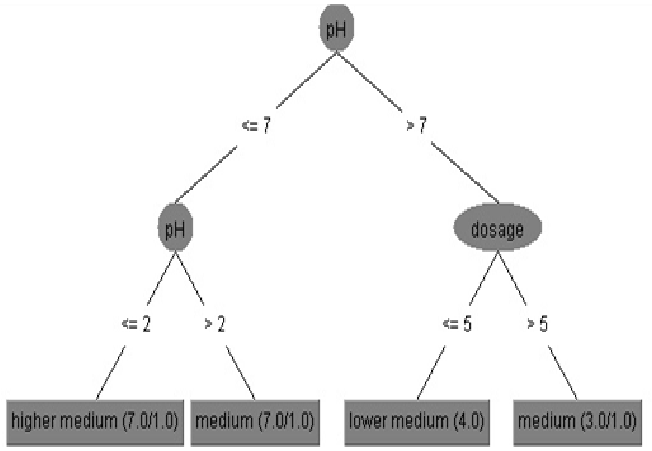
5. Xu,M, wang, J. Chen, T. (2006). Improved decision tree algorithm: ID3+, intelligent computing in signal Processing and pattern recognition, Vol. 345, PP.141-149

Appendix 1: Figures

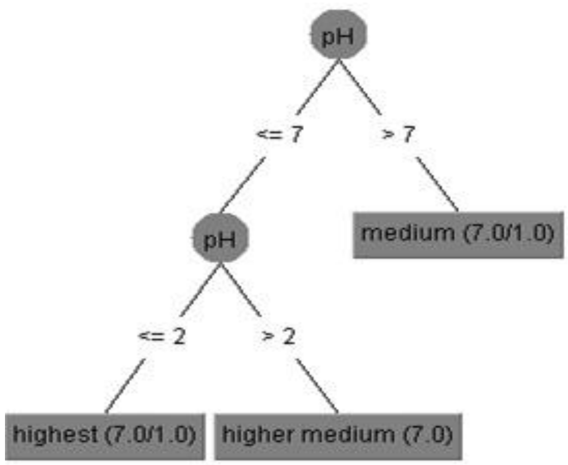
Figure 1: Tree Classification of the Colour Removal capacity of Various adsorbents using C-4.5 classifier



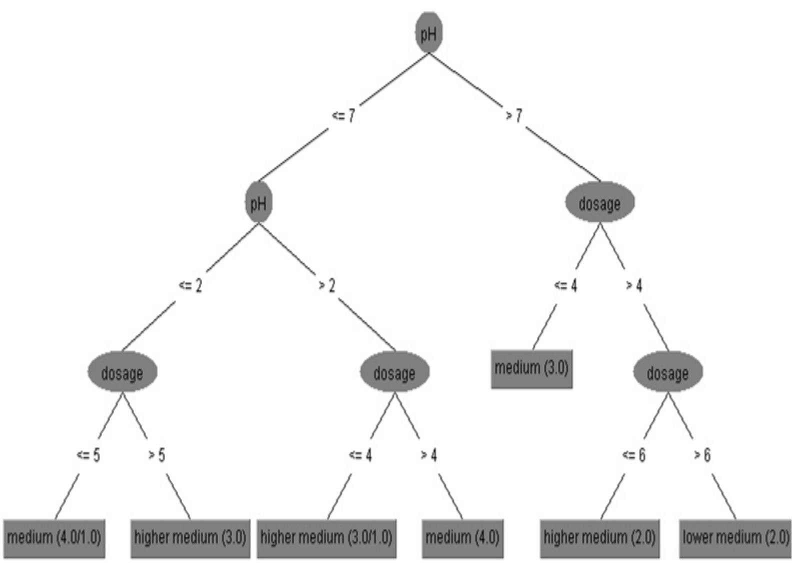
(a) Jatropha- Hydro-extracted Adsorbent



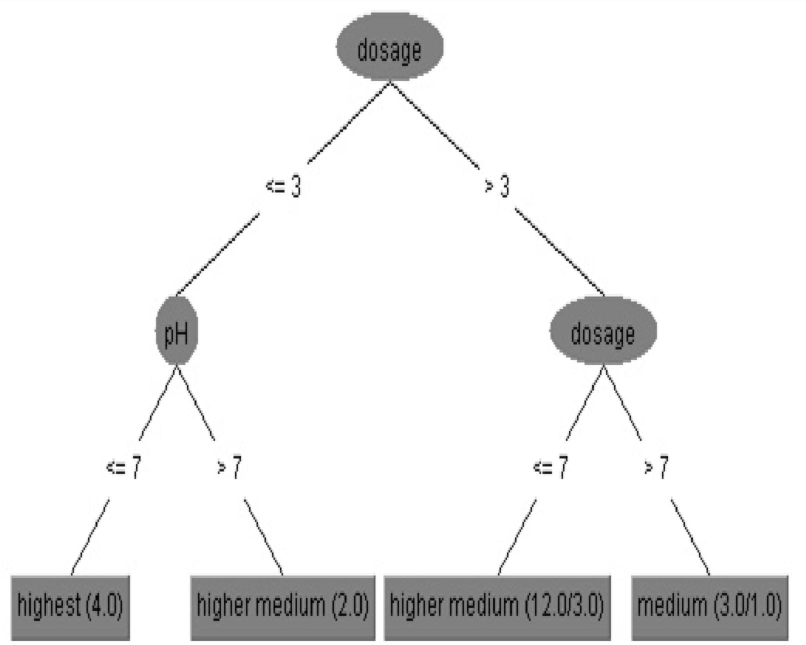
(b) Jatropha- Raw Adsorbent



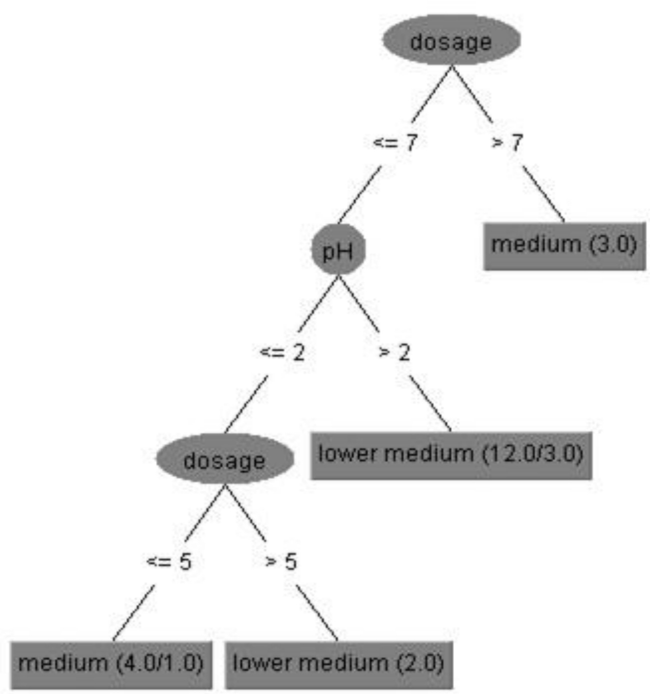
(c) Jatropha-Methanol-extracted Adsorbent



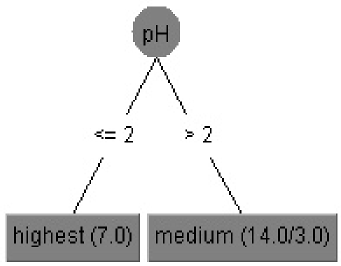
(d) Jatropha-Acid-extracted Adsorbent



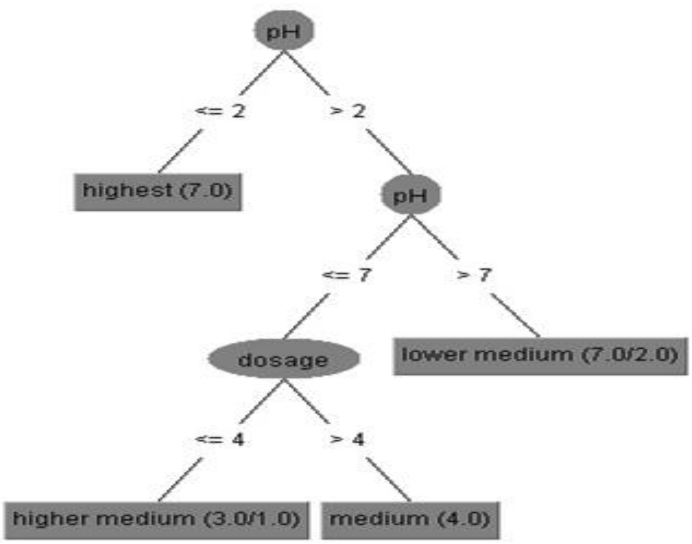
(e) Jatropha-Hexane-extracted Adsorbent



(f) Jatropha-Alkali-extracted Adsorbent



(g)Jatropha-Toluene-extracted Adsorbent



(h)Activated Jatropha Adsorbent