Special Issue - 2015

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
NCACS-2015 Conference Proceedings

To Build A Water Effluent Treatment in Bleaching Industry using Decision Tree Technique

G. D. Praveen Kumar[1],
[1]M.Phil Research scholar ,
Kaamadhenu arts and science college,

Abstract:- **A research area are working on data mining techniques in bleaching industry, knowledge discovery and data mining have found numerous application in business and scientific domain. Valuable knowledge can be discovered from application of data mining techniques in bleaching industry. There are four attribute used in the dataset. The dataset are based on the effluent treatment plan processing in bleaching industry .This data model was tested in WEKA3-7 software. The decision tree was implemented to classify the pollution class of water. The class are divided into three parts.**

Key Words: knowledge discovery and data mining, data set, decision tree

INTRODUCTION

Thirupathy Bleaching is one of the small scale industry in erode. It producing a white thread. Bleaching industry are water intensive industry from the point achieving zero discharge deserve special mention and appreciation. Only source water here is ground water. The state pollution control board has restricted the extraction of ground water ,so that water are pretreatment in effluent treatment plan(ETP).

Data mining

Data mining is an approach for information extraction from huge amount of data stored in database[3]. This process is performed to find hidden pattern and relationship in a present data. The overall objective of data mining process is to extract information from a large dataset and transform it into a comprehensible structure for future use. Generally, the task of data mining into two types[2].

1. Descriptive Data Mining: In descriptive data mining ,the data set is summarized in a concise manner and present interesting properties of data.
2. Predictive Data mining: The ultimate goal is prediction, which is the most common application of data mining, in this behavior of the future data set is predicted.

The following process in a data mining are,

1. Selection
2. Preprocessing
3. Transformation
4. Data mining
5. Evaluation

DECISION TREE

A decision tree is classification scheme which generate a tree and set of rules representing the model of different classes from a given dataset[1]. The former is used to deriving the classifier while the latter is used to measure the accuracy of the classifier. The accuracy of the classifier is determine by percentage of the test data that are correctly classified.

Categorize the attribute of the record into two different type attribute domain is numerical are called numeric attribute and the attribute domain is not numeric are called categorical attribute. There is one distinguished attribute called the class label[7]. The goal of classification is to build a concise model that can be used to predict the class of the record whose class label is unknown.

A decision tree allow the calculation of forward and backward because of that correct decision automatically. Decision tree is a classifier in the form of tree structure where each node is either:

1. Leaf node.
2. Decision node.

WEKA SOFTWARE

The WEKA software was developed in the university of new Zealand. A number of data mining methods are implemented in theWEKA software. Some of them are based on the decision tree like the J48 decision tree. Some are rule based like zeroOR and decision table and some of them are based on probability and regression like the Naïve Bayes algorithm[5]. WEKA use the J48 algorithm, which is WEKA implementation of C4.5. decision tree algorithm. J48 is actually a slight improved tool and the latest version of C4.5. it was the last public version of this family of algorithm before the commercial implementation of C5.0 was released.

EXPERIMENTAL DATASETS

In this research we have used purified water quality of Thirupathy Bleaching in Erode from 2005 to 2014. The data set after preprocessing is given Table 1.

Special Issue - 2015

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
NCACS-2015 Conference Proceedings

EXPERIMENTAL SETUP

In the experiment whole datasets was training as a training set for developing a model . J48 is an open source java implementation of the C4.5 algorithm in the weka data mining tool.C4.5 is a program that creates a decision tree based on a set of labeled input data. This algorithm was developed by Ross Quinlan. The decision tree generated by C4.5 can be used for classification and for this reason, c4.5 is often referred to as statistical classifier[6].

DATA PREPARATION:

The process of data cleaning and preparation is highly dependent on the specific data mining algorithm and software chosen for the data mining task. To prepare the data according to requirement of the selected data mining software[4].

PREPROCESSING OF THE DATA:

Before data mining algorithms can be used a target data set must be assembled. As data mining can only uncover pattern s actually present in the data the target datasets must be large enough to contain these pattern while remaining concise enough to be mined with an acceptable time limit[7]. A common source for data is a data mart or dataware house. Preprocessing is essential to analyze the multivariate dataset before data mining. The target set is then cleaned. Data cleaning removes the observation containing noise and those with missing data.

YEAR	FS	LIME	DO	NH3-N	CLASS
2005	10.00	8	8.1	0.39	A
2006	8.2	7.2	7.8	0.12	C
2007	7.2	6.1	7.9	0.13	B
2008	9	8.2	8.2	0.14	A
2009	8.2	7.5	7.7	0.11	B
2010	9	8.4	7.5	0.06	C
2011	6.3	5.5	5.3	0.49	A
2012	7.8	7	7	0.31	A
2013	9.2	8	8.2	0.12	B
2014	9.7	8.2	8.3	0.13	C

Table 1-purified water quality dataset after preprocessing
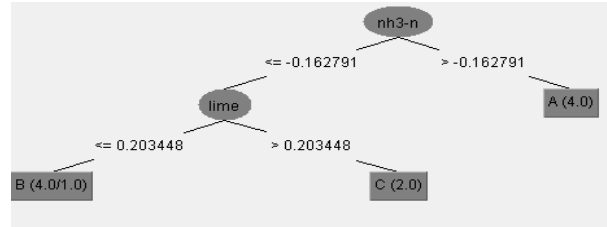
MODEL:

Many parameter can influence the surface water quality. Four parameters are selected to tested in ETP.

Table 2-purified water quality parameter

S.NO	ATTRIBUTE	ABBREVATION
1	FS	Ferrous Sulphate
2	LIME	Lime Powder
3	DO	Dissolve oxygen
4	NH-3	Ammonia nitrogen
5	class (polluated class)	A,B,C

DECISION TREE TECHNIQUE IN ETP

Generally water class are divided into three classes: class A, industry waste are fully recycled in ETP . class B, turbocirculator followed by pressure sand filter. Turbo circulator is basically a flash mixer. Class C ,sledge waste water.

J48 pruned tree
Rule1: IF nh3-n <= 0.13
Rule 2: IF lime <= 8: B (4.0/1.0)
Rule3:IF lime > 8: C (2.0)
Rule 4:IF nh3-n > 0.13: A (4.0)

Number of Leaves : 3
Size of the tree : 5
Time taken to build model:0.03 seconds

EVALUATION OF TRAINING SET

Correctly Classified Instances   9   90   %

Incorrectly Classified Instances 1   10   %

CONCLUSION

Now, days most of industries have need to treat water to obtain very high quality water for demanding purpose. Water treatment produces organic and mineral sludge from filtration and sediment an . In this result we have implemented purified water quality data with decision tree technique. We have used j48 classifier in weka3-7 data mining tool. Experiment used four attribute of purified water quality data which affect the accuracy of water. The correctly classification of instance is 90% and incorrectly classified instance is 10%, the total time taken to test model on training data is 0.01 seconds.

REFERENCES

1. Bhargava neeraj et al,2013-"decision tree analysis on j48 algorithm for data mining"-international journal of advance research in computer science and software engineering,vol-3,issue-6.
2. Han, j and M.kamber, data mining: concept technique morgan kaufmann,2001.
3. Han j and kamber m. 2001.data mining:concepts and techinques. San Francisco, morgan Kauffmann publisher.
4. Dehaiwar kavita and singh jagdish,2012"water resource management and water quality,case of Bhopal",international conference on chemical, ecology and environment sciences(ICEES'2012)S MARCH 17-18, 2012 BANGKOK.
5. Fayyad U.M., piatetsky-Shapiro G.,Smyth P.,from data mining to knowledge.
6. Decision tree 2003 http://en.wikipedia.org/wiki/decision_tree c4.5 algorithm 2005.
7. Arun K Pujari, data mining techniques, university press 2001.