Data Mining for Pesticides Dietary Exposures- A Review Article

Rucha Patel
MCA Department
PARUL UNIVERSITY, VADODARA, GUJARAT

Abstract- *In this paper an attempt has been made to review the research studies on application of data mining techniques in the field of pesticides dietary exposures. Some of the techniques, such as Clustering (Recursive Noise Removal), Classification, artificial neural networks and support vector machines applied in the field of pesticides were presented. Data mining in application in pesticides is a relatively new approach for forecasting / predicting of pesticides dietary exposures. This article explores the applications of data mining techniques in the field of pesticides dietary exposures.*

Keywords- Pesticides, Dietary exposures, Clustering, Artificial Neural Networks

I. INTRODUCTION

Agriculture is backbone business in India. It contributes 10-15% GDP into the India economy. India has an agrarian economy, where the 1012.4 million populations are dependent on agricultural commodities from 124.07 million hectares cropped area cultivated by 110.7 million producers. Thus, agricultural growth can help alleviate poverty and uplift our economy, thereby proving beneficial for everyone. It is needed optimization or reduction of pest usage in India. To the optimization of usage of pesticides and dietary exposure, data mining is a process that uses a variety of data analysis tools to discover patterns and relationships in data that may be used to make a valid prediction. Discovering previous unknown patterns, predicting new trends and behaviour is the two outstanding characteristics of data mining. The two primary goals of data mining tend to be a prediction and description. Prediction involves using some variables or fields in the data set to predict unknown or future values of the other variables of interest and description focuses on finding patterns describing the data that can be interpreted by humans. Several data mining techniques have been proposed to date such as decision tree, genetic algorithm, artificial neural network, nearest neighbour method, rule induction etc. are used to evaluate the when and how much pesticides will use.

II. LITERATURE REVIEW

The presented literature review is described about the consumption of pesticide usage in India and also it is displayed the comparison between India and world on the basis of pesticide usage.

India is one of the five major cotton growing countries in the world. Almost 70% of world cotton is produced in China (Mainland), India, Pakistan, USA and Uzbekistan. India is the worlds 7th most populous country; anticipating population growth, in 60's and 70's pesticides were identified as a means for increasing production, as a positive correlation is believed to exist between yield and pesticide usage. However, FAO (2001) has reported the existence of a negative correlation between pesticide use and yield in India (Figure 1).

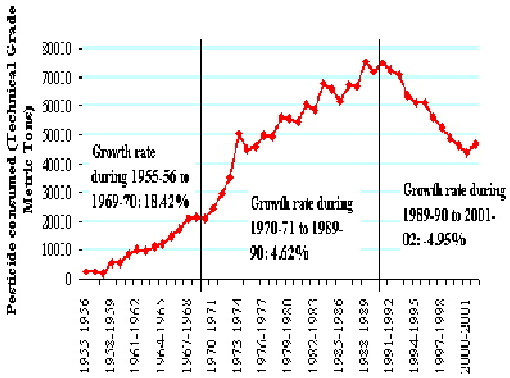

Figure.1. Consumption of Pesticides in India

A marked increase in yield loss while the pesticide usage is on the rise has created a complex situation. Excessive use of pesticides is harmful in multiple ways. Here study about usage of pesticides in India compare to world, it helps in collecting the pest scouting data for further evaluation.

Another case study for pesticide dietary exposures describes that, though God is not the only source of exposure to the more than one billion pounds of pesticides used annually in the United States, pesticide residues are a part of most meals. Almost 80 percent of the 21,807 samples of conventionally grown fresh fruits tested byUSDA's Pesticide Data Program (PDP) from 1993-2002 contained one or more pesticide residues. Nearly one-half of the samples of

conventional crops tested by USDA in the last decade contained multiple residues. In the 2002 testing, peaches contained an average of 4.2 different residues and celery contained 3.5 residues see in below figure.

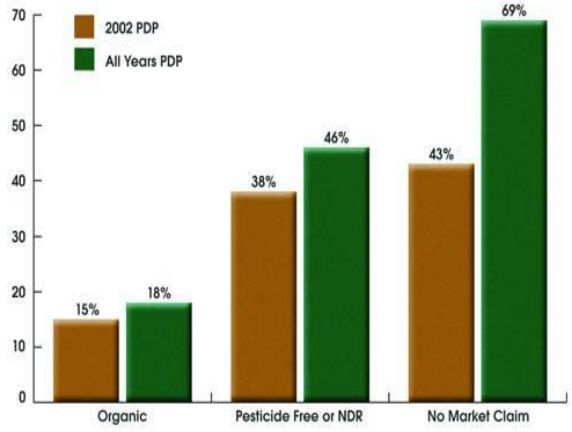

Fig.2. Pesticides Consumption Statistics

From the above described literature review, usage of the pesticides will increase the dietary exposures issues day by day and Pesticides have contaminated almost every part of our environment. Pesticide residues are found in soil and air, and in surface and ground water across the nation, and urban pesticide uses contribute to the problem. Pesticide contamination poses significant risks to the environment and non-target organisms ranging from beneficial soil microorganisms, to insects, plants, fish, and birds.

III. APPLICATION OF DATA MINING TECHNIQUES IN PESTICIDES DIETARY EXPOSURES

In Indian agriculture, the volume of data is enormous. The data when become information is highly useful for many purposes. The conventional and traditional system of data analysis in agriculture is purely dependent on statistics.

Data mining is a modern data analysis technique. It has a wide range of applications in the field of agriculture. Data mining is the process of extracting vital and useful information from large sets of data. Data mining is the process of discovering previously unknown and potentially interesting patterns in large data sets. The mined information is used for representation as a model for prediction or classification. Datasets from the agricultural domain appear to be significantly more complex than the datasets traditionally used in machine learning. Data mining is mainly categorized as descriptive and predictive data mining. But in the agricultural area, predictive data mining is mainly used. Different techniques of data mining have been used in this field. Though, there are lots of techniques available in the data mining, few methodologies such as Artificial Neural Networks, classification and clustering, visual data mining, fuzzy sets. Here is an overview of some of the methods according to past usage. coating (TBC) system, such as edges and corners Methods:

A) Artificial Neural Networks [ANN]:

An artificial neural network is a non-linear predictive model that learns through training and resembles biological neural networks in structure. ANNs as data driven empirical models have been successfully applied in all fields of Agriculture including Optimization of Pesticide usage.

B) Classification [Through Classification and regression Tree-CART]:

In this study, our aim was to identify households with the highest burden of pesticide residues for selected pesticides. Classification and regression tree (CART) analysis is one approach that can provide this identification. CART is a nonparametric procedure that has been applied in several disciplines including environmental health studies to build predictive models. CART works by disaggregating data into homogeneous subsamples and is applicable in settings with small sample sizes and multivariate comparisons as well as multiple interactions between variables. The CART analysis provides a qualitative classification, a potential limitation of the method in some settings. However, from a risk management perspective, we are more concerned with identifying high-risk households than with capturing the quantitative relationships among variables.

C) Clustering [Recursive Noise Removal]:

This technique is used to research on the pest scouting, pesticide usage and meteorological data from the Gujarat agriculture field. Pest scouting data by itself can be termed as “Gold mine” of data and coupling it with pesticide usage and meteorological data can provide an excellent insight into the dynamics of past situations and their outcomes.

Other than these methods we can also search on the pesticide usage and how to optimize its like fuzzy sets, visual data mining, clustering etc. They are included in further proposed research work. By use of these methods best optimization technique will be developed.

Effect of pesticides on humans can’t be directly checked because of the poisonous nature of pesticides, therefore the usage of pesticides on food causes dietary exposures has been taken into consideration for this purpose.

Here mentioned techniques can not only be used for measure pesticide dietary exposures data, but also possesses the flexibility to deal with any numeric data which contained by dietary exposures databases.

VI. CONCLUSIONS

There are a growing number of techniques of data mining for predicting pesticides dietary exposures and a growing amount of data that are currently available from many data mining databases. This is relatively a novel research field and it is expected to grow in the future. There is a lot of work to be done on this emerging and interesting research field. The multidisciplinary approach of integrating computer science with agriculture will help in forecasting/ predicting pesticides dietary exposures, pest scouting effectively.

REFERENCES

[1] Abdullah, A. and Brobst, S. (2003): Clustering by recursive noise removal. Proc. Atlantic Symposium onComputational Biology and Genome Informatics,USA, pp. 973-977.
[2] Han and Kamber (2000): Data Mining: Concepts and Techniques. Morgan Kaufmann Publishers.
[3] FAO (2001), Policy and Strategy for Rational use of Pesticides. Food and Agriculture Organization.UNDP
[4] Rhona Julien, Jonathan I. Levy, Gary Adamkiewicz, Russ Hauser, and John D. Spengler,Pesticides in Urban Multiunit Dwellings: Hazard Identification Using Classification and Regression Tree (CART) Analysis, ISSN:1047-3289 J. Air & Waste Manage. Assoc. 58:1297–1302
[5] S.S. Baskar, L. Arockiam, V. Arul Kumar and L. Jeyasimman, Brief Survey of Application of Data Mining Techniques to Agriculture (2010)
[6] Vijesh V. Krishna, N.G. Byju, S. Tamizheniyan, Integrated Pest Management In Indian Agriculture: A Developing Economy Perspective, 2002.
[7] Chopra, K. 1993.Sustainability of agriculture. Indian Journal of Agricultural Economics, 48(3): 527-537
[8] US Environmental (July 24, 2007), What is a pesticide?epa.gov. Retrieved on September 15, 2007.
[9] "Types of Pesticides". US Environmental Protection Agency. Retrieved 20 February 2013.
[10] Henrik Boström, Feature vs. Classifier Fusion for Predictive Data Mining – a Case Study in Pesticide Classification [2009]
[11] Hastie T., Tibshirani R. and Friedman J., The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Springer-Verlag (2001)