ISSN 2348 – 9928
Doi:01.0401/ijaict.2014.05.02 Published on 05 (1) 2016

A GENERIC PREDICTIVE KNOWLEDGE MANAGEMENT MODEL FOR FISHERIES WITH SPECIAL EMPHASIS CATCH OF OIL-SARDINE ALONG THE SOUTH-WEST COAST OF INDIA

Mr. Bruce Mathew
M. Phil Scholar,
Department of CS,
GRD College of Science,
Coimbatore, Tamilnadu, India

Prof. P. Sumathi
Assistant Professor,
Department of IT & CS,
GRD College of Science,
Coimbatore, Tamilnadu, India

Abstract— ***Knowledge Management is a very hot domain these days due to the increasing asset value for knowledge available within the organization. These days, knowledge is treated as a vital asset that can increase organization's competitive advantage. The potential that knowledge management has for improving fisheries management is increasingly being recognized. There are relatively few studies that have specifically addressed the challenges of knowledge management in fisheries. Knowledge about various fishery resources has become a major focal point of all stakeholders in the investment of time and effort. In this paper, we propose a predictive knowledge management model that is specific to the availability of Oil Sardine along the south-west coast of Kerala. This model can be applied for gaining business intelligence as well as improving competitive effectiveness. This model is processed for attribute reduction with the help of rough set theory and accordingly decision rules have been defined.***

Keywords—Knowledge, Rough Set Theory, Rule Induction, Reduction, Reducts, Local Covering.

I. INTRODUCTION

The ability to manage knowledge is crucial in today's knowledge economy. The creation and diffusion of knowledge have become increasingly important factors in competitiveness. An organization in the Knowledge age is one that learns, remembers, and acts based on the best available information, knowledge, and know-how. All these developments have created a strong need for a deliberate and systematic approach to capturing and sharing a company's knowledge base. In other words, in order to be successful in today's challenging organizational environment, companies need to learn from their past errors and not reinvent the wheel.

Knowledge management represents an effective approach to ensure the full utilization of the organization's knowledge base, coupled with the potential of individual skills, competencies, thoughts, innovations, and ideas to create a more efficient and effective organization [1].

The importance of knowledge management in fisheries is relevant in modern century. Fishery managers and stakeholders rely on the output of research, an important component of knowledge, when developing plans for sustainable fisheries. KM is also required to improve the understanding of the biology of fish stocks – their growth rates, their feeding patterns, where they migrate to, the impact that fishing has on the seabed, and how external factors such as temperature and oceanographic processes affect recruitment. In the fisheries domain, many decisions are made based on experience; there is a need to be able to make decisions which are to a greater extent based on more objective evidence which is the focus of any KM system.

The proposed problem stems from the relevance of KM and its application in fisheries domain to attain competitive advantage and business intelligence. There is a wide spread decline in the catch rate of oil sardine along the south west coast of India. Large efforts are expended to catch this species of fish. But the effort produced has no relationship with the catch received each day due to the varied environment and biological conditions. In this paper, we briefly study about the various climatic parameters that relate to the catch of oil sardine and specifically analyze the relationship of each attributes with respect to the catch of oil sardine. We also apply rough set theories to identify prominent and non prominent attributes out of the identified parameters; followed by generating rules for proposing a model for fish catch & effort with relation to climatic parameters identified. These rules are helpful for the fisherman community to design their fishing plans, which will enable them to improve their catch with minimum effort. The rules generated are of utmost importance in establishing the knowledge related with the

Corresponding Author: Mr. Bruce Mathew, GRD College of Science, Coimbatore, Tamilnadu, India.

ISSN 2348 – 9928
Doi:01.0401/ijaict.2014.05.02 Published on 05 (1) 2016

catch of Oil sardine and thus forms part of business intelligence of the domain.

The remainder of the paper is organized as follows. An introduction to the various basic terminologies related with this paper is given in section II, description of the data collected and the procedure adopted for processing the data is given in section III. In section IV, the work is explained along with the results. This is followed by the conclusion.

II. BASIC TERMINOLOGIES AND CONCEPTS

2.1 Knowledge

Knowledge is often a justified personal belief. Dictionary has defined knowledge as "the facts, information, and skills acquired through experience or education or the theoretical/ practical understanding of a subject". Knowledge is derived from information, but it is much richer and meaningful than mere information. Knowledge includes familiarity, awareness and understanding gained through experience or study, and results from making comparisons, identifying consequences, and making connections. In organizational terms, knowledge is generally thought of as being "know how", or "applied action".

Knowledge is closely linked to doing and implies know-how and understanding. The knowledge possessed by each individual is a product of his experience, and encompasses the norms by which he evaluates new inputs from his surroundings [2].

In organizations, knowledge often becomes embedded not only in documents or repositories, but also in organizational routines, practices and norms.

In general, there are two types of knowledge: Tacit knowledge and Explicit Knowledge. Tacit knowledge is difficult to articulate and difficult to put into words, text, or drawings. Explicit knowledge represents content that has been captured in tangible form such as words, audio recordings, or images. Tacit knowledge tends to reside within the one's head, whereas explicit knowledge is usually contained within tangible or concrete media.

Comparison of properties of tacit versus explicit knowledge

Properties of tacit knowledge

- Ability to adapt, to deal with new and exceptional situations
- Expertise, know-how, know-why, and care-why
- Ability to collaborate, to share a vision, to transmit a culture

Properties of explicit knowledge

- Ability to disseminate, to reproduce, to access and re-apply throughout the organization
- Ability to teach, to train
- Transfer knowledge via products, services, and documented processes

2.2 Knowledge Management

According to Davenport & Prusak [2], which states that KM "is managing the corporation's knowledge through a systematically and organizationally specified process for acquiring, organizing, sustaining, applying, sharing and renewing both the tacit and explicit knowledge of employees to enhance organizational performance and create value".

Knowledge management is about applying the collective knowledge of the entire workforce to achieve specific organizational goals. The aim of knowledge management is not necessarily to manage all knowledge, just the knowledge that is most important to the organization. It is about confirming that people have the knowledge they need, where they need it, when they need it – the right knowledge, in the right place, at the right time.

Knowledge management is based on the idea that an organization's most valuable resource is the knowledge of its people. This means that creating, sharing and using knowledge are among the most important activities of nearly every person in every organization and domain. The basic goal is to improve organizations knowledge assets to effectuate better knowledge practices, improved organizational behavior, getter decision and improved organizational performance.

2.3 Rough Sets

Rough set theory (RST) is a major mathematical method developed by Pawlak. This method has been developed to manage uncertainties from information that presents some inexactitude, incompleteness and noises. When the available information is insufficient to determine the exact value of a given set, lower and upper approximations can be used by rough set for the representation of the concerned set. In decision making, it has been confirmed that rough set methods have a powerful essence in dealing with uncertainties. Rough set philosophy is founded on the assumption that with every object of the universe of discourse some information (data, knowledge) is associated. Objects characterized by the same information are indiscernible (similar) in view of the available information about them. The in-discernibility relation generated in this way is the mathematical basis of rough set theory [3].

ISSN 2348 – 9928
Doi:01.0401/ijaict.2014.05.02 Published on 05 (1) 2016

2.4 Reducts

A reduct is a subset of attributes that are jointly sufficient and individually necessary for preserving a particular property of a given information table. The basic objective of reduct construction is to reduce the number of attributes, and at the same time, preserve the property that is required by the user.

2.5 Rule Mining

The rule mining is a technique used in data mining and is the defined as follows. Given a set of transactions, find rules that will predict the occurrence of an item based on the occurrences of other items in the transaction. The goal of rule mining is to generate rules that can analyze and predicting customer behavior. Every rule must satisfy two constraints: a measure of statistical significance called support and a measure of goodness called confidence.

Support(s) of an association rule mining is defined as the percentage/fraction of records that contain $X \cup Y$ to the total number of records in the dataset. Confidence of an association rule mining is defined as the percentage/fraction of the number of transactions that contain $X \cup Y$ to the total number of records that contain X, where if the percentage exceeds the threshold of confidence, an interesting rule X=>Y can be generated.
i.e. Confidence (X|Y)= Support (XY)/Support (X).

2.6 Local Covering

An information system $I = (U, A\cup\{d\})$, where $d \notin A$, is usually called a decision table. The elements of A are called conditional attributes and d is called the decision attribute. A decision table defines an information function $f : U\times A \rightarrow V$, where V represents the set of all attribute values. Let $a \in A$ and $v \in V$, and $t = (a, v)$ be an attribute-value pair. Then a block of t, denoted by [t], is a set of objects from U for which attribute 'a' has value v [11].

Let $x\in U$ and $B\subseteq A$. An elementary set of B containing x, denoted by $[x]_B$ and is defined by the following set.

$$\cap\{[a, v] \mid a\in B, f(x, a) = v\}$$

Elementary set are subsets of U consisting of all cases from U that are indistinguishable from x while using all attributes from B [12]. Elementary sets are also known as information granules, which represents the building blocks of knowledge about U. When subset B is restricted to a single attribute, elementary sets are blocks of attribute-value pairs defined by that specific attribute. Elementary sets can also be defined by using the idea of in-discernibility relation, the mathematical basis of rough set theory.

There are two main approaches to data mining from complete data sets based on RST. In both approaches decision tables are used. In the first approach, the entire attributes are used for analysis and hence the approach is known as global. The second approach is known as local, in which blocks of attribute-value pairs are used for analysis. To define the concept of a local covering, the idea of a minimal complex is introduced first. A minimal complex corresponds to a single rule [4].

Let X is a concept. Let t be an attribute-value pair (a, v) and let T be a set of attribute-value pairs. Then a block of t, denoted as [t] is the set of objects for which attribute 'a' has value v. Set X is said to depend on set T of attribute-value pairs if and only if:

$$\emptyset \neq \cap\{[t] \mid t\in T\} \subseteq X$$

A set T is a minimal complex of X if and only if X depends on T and no proper subset T′ of T exist such that X depends on T′.

A local covering corresponds to a rule set describing a concept. Let L be a non-empty collection of non-empty sets of attribute-value pairs. Then L is a local covering of X if and only if the following conditions are satisfied.

- Each member T of L is a minimal complex.
- $\cup\{[T] \mid T \in L\} = X$ and
- L is minimal, i.e., L has smallest possible number of members.

2.7 ROSE2 (Rough Sets Data Explorer)

This is the software tool that is widely used throughout this work. ROSE2 (Rough Sets Data Explorer) is a software implementing basic elements of the rough set theory and rule discovery techniques. It has been created at the Laboratory of Intelligent Decision Support Systems of the Institute of Computing Science in Poznan, and is based on a fourteen-year experience in rough set based knowledge discovery and decision analysis.

ROSE2 is an interactive system running on 32-bit Microsoft Windows operating systems (Windows 95 and better). Core modules were written in C++ (ANSI standard), while the interface modules were developed using Borland C++ Builder and Borland Delphi.

The system consists of a graphical environment and a set of separate computational modules. The modules are platform-independent and can be recompiled for different operating systems, including Linux.

Corresponding Author: Mr. Bruce Mathew, GRD College of Science, Coimbatore, Tamilnadu, India.

ISSN 2348 – 9928
Doi:01.0401/ijaict.2014.05.02 Published on 05 (1) 2016

The environment acts as an overlay on all computational modules. So it is quite easy to add new modules to the system and that is an important characteristic. This guarantees greater expandability of the system in the future [5].

2.8 R - The statistical computing tool

For the data preprocessing, we have used R as the tool. R is a language and environment for statistical computing and graphics. R provides a wide variety of statistical (linear and nonlinear modeling, classical statistical tests, time-series analysis, classification, clustering, etc) and graphical techniques, and is highly extensible. R is an integrated suite of software facilities for data manipulation, calculation and graphical display. It includes an effective data handling and storage facility, a suite of operators for calculations on arrays, in particular matrices and a large, coherent, integrated collection of intermediate tools for data analysis [6].

III. DESCRIPTION & PROCESSING OF EXPERIMENTAL DATA

Data set used for this experiment is collected from two sources –the primary data pertaining to the catch of Oil sardine from Central Marine Fisheries Research Institute, Kochi and secondary data from the International Comprehensive Ocean Atmospheric Data-Set (ICOADS), National Oceanic & Atmospheric Administration (NOAA), U.S. Department of Commerce. The location that was selected for the study was along the south west coast of India lying between 8° to 16° North latitude and 72° to 77° East longitude. This was the area where the catch of Oil sardine is higher in India. At the time of data collection, a number of attributes were considered and data was collected accordingly. The data collected had values on a monthly basis of a period of 3 years starting from January 2010 to December 2013 (refer Appendix I). The collection of the data set was merely based on the study related to the climatic parameters of the landings of Oil sardine published by the Fisheries Department of Canada [7]. Of the various climatic parameters, we have randomly selected six environmental parameters and data was collected accordingly. These six parameters were SST (surface sea temperature), AT (atmospheric temperature), RH (relative humidity), SLP (sea level pressure), TC (Total cloudiness) and East Wind. Studies have found that these six parameters have a great role in the growth of pelagic fishes, especially the oil sardine [13].

Once the data was collected, it was found incomplete due to lack of data pertaining to many attributes. This led to the introduction of rough set theory which plays a great role in cases where there is inconsistency and incompleteness in dataset. Since the data that was collected was on a continuous basis and which was not suitable for rough set algorithms, the data was discretized which was a form of preprocessing. Discretization of continuous attributes is one of the important data preprocessing steps of knowledge extraction. The goal of discretization is to find a set of cut points to partition the range into a small number of intervals that have good class coherence, which is usually measured by an evaluation function [8].

TABLE 1: Domain range for attributes after Discretization process using CAIM Algorithm

Class	SST	AT	SLP	TC	EAST WIND	TOTAL
1	27.47044 to 27.56866	26.9874 to 27.2605	1007.252 to 1008.006	2.6833 to 2.7087	-1.7618 to -1.5853	2968.5 to 6763.9
2	27.56867 to 27.73025	27.2606 to 27.3380	1008.007 to 1008.192	2.7088 to 2.8126	-1.5852 to -0.6908	6764 to 8816.2
3	27.73026 to 28.36893	27.3381 to 27.4600	1008.193 to 1009.053	2.8127 to 2.8754	-0.6907 to -0.6258	8816.3 to 10503.75
4	28.36894 to 28.37480	27.4601 to 27.4751	1009.054 to 1009.661	2.8755 to 3.5463	-0.6257 to 0.2781	10503.76 to 18176.45
5	27.37481 to 28.38404	27.4752 to 27.4895	1009.662 to 1009.705	3.5464 to 3.6820	0.2782 to 2.0447	18176.46 to 19872.76
6	28.38405 to 28.75955	27.4896 to 28.0133	1009.706 to 1010.031	3.6821 to 4.4684	2.0448 to 2.8098	19872.77 to 41995.78
7	28.75956 to 28.92879	28.0134 to 28.2072	1010.032 to 1010.124	4.4685 to 4.5193	2.8099 to 3.3691	41995.79 to 42549.54
8	28.92880 to 28.98595	28.2073 to 28.3297	1010.125 to 1010.240	4.5194 to 4.5907	3.3692 to 3.9743	42549.55 to 46032.91
9	28.98596 to 29.27877	28.3298 to 28.4523	1010.241 to 1010.459	4.5908 to 5.0155	3.9744 to 6.2574	46032.92 to 48760.83
10	29.27878 to 29.30431	28.4524 to 28.9949	1010.460 to 1012.311	5.0156 to 5.6426	6.2575 to 7.1881	48760.84 to 52168.3
11	29.30432 to 29.42249	28.9950 to 29.5891	1012.312 to 1013.107	5.6427 to 6.5142	7.1882 to 7.8436	52168.4 to 54146.63
12	29.42250 to 30.74936	29.5892 to 30.2920	1013.108 to 1013.454	6.5143 to 6.6061	7.8437 to 8.6373	54146.64 to 92154.68

Though there were various methods for discretization, we used the CAIM discretization algorithm for preprocessing [9]. The data was discretized and was classified into 12 classes and now, it was ready for further rough set processing.

Corresponding Author: Mr. Bruce Mathew, GRD College of Science, Coimbatore, Tamilnadu, India.

ISSN 2348 – 9928
Doi:01.0401/ijaict.2014.05.02 Published on 05 (1) 2016

During the next stage where a ROSE2 was involved, the discretized dataset was processed using a proportional rough set (PRS) relevance method for attribute reduction. The aim was to reduce the number of attributes and thus in turn reduce the rules that would be generated at the later stage. The PRS relevance method is an effective Rough Set based method for attribute selection [10]. The concept of reducts is used as the basic idea for the implementation of this approach. The main idea behind the reduct based feature selection approach is, more frequent a condition attribute appears in the reducts, the more relevant the attribute is. Hence the number of times an attribute appears in all reducts and the total number of reducts determines the significance (priority) of each attribute in representing the knowledge contained in the dataset. This idea is used for measuring the significance of various features in reduct based PRS relevance feature selection approach.

Consider a decision table $T = \{U, A, d\}$, where U is the non-empty finite set of objects called the Universe, $A = \{a_1, a_2, .. a_n\}$ be the non-empty finite set of conditional attributes and d is the decision attribute. Let $\{r_1, r_2, ..., r_m\}$ be the set of reducts generated from T. Then, for each conditional attribute $a_i \in A$, reduct based attribute priority $\beta(a_i)$ is defined as [10]:

$$\beta(a_i) = \frac{\left|\{r_j | a_i \in r_j\}\right|}{m}, \ j = 1,2,3,...,m$$

The discernability matrix approach of ROSE2 generated five reducts out of the identified attributes. These five reducts or prominent attributes were SST, AT, SLP, TC & East_Wind. RH was identified as non-significant among the dataset and so it was omitted. These reducts had significance percentage of 60%, 80%, 80%, 100% and 100 % respectively. The non significant factor RH had only 40 % and so we identified it as non-significant.

After the reduction process, all unwanted attributes were avoided and a refined decision table was created which was ready for rule induction. For the rule induction, we used the LEM2 (Learning by Examples-Module 2) algorithm, which was also implemented in ROSE2. The pseudo code of the standard LEM2 algorithm is as follows [4]:

Algorithm LEM2(X)

```
// X represents a set of objects representing lower/upper
// approximation of the concept selected.
// t represents an attribute-value pair.
// The algorithm returns a single local covering L of X.
{
  G := X;
  L := ∅;
  while (G≠∅) do
  {
    T := ∅;
    T(G) := {t | [t] ∩ G ≠ ∅};
    while ((T = ∅)or(not([T] ⊆ X)))
    {
      Select a pair t∈T(G) such that |[t] ∩ G| is maximum.
      if a tie occurs arbitrarily select any one pair.
        T := T ∪ {t};
        if ([t] ∩ G ≠ ∅)
          {
            G := [t] ∩ G;
            T(G):={t |[t]∩G ≠ ∅; t∈T(G)}
            T(G) := T(G) – T;
          }     }
    for each t in T do
    {       if ((T – {t}) ≠ ∅) then
              if ([T – {t}] ⊆ X) then
                  T := T – {t};
    }
    L := L ∪ {T};
    G := X – ⋃_{T∈L} [T];
}
for each T ∈ L do
      if (∪_{P∈L-{T}}[P] = X) then
            L := L – {T};}
```

IV. DATA ANALYSIS AND RESULTS

After the rule mining process using the LEM2 algorithm, decision rules for the knowledge management system is generated. The LEM2 algorithm now generates number of certain rules and possible rules.

The idea of lower and upper approximations is used for rule mining. Elementary sets of the decision attribute are called concepts. For any concept, rules induced from its lower approximations are certainly valid and hence such rules are called certain rules. Rules induced from upper approximation of the concept are possibly valid are called possible rules. After computing the lower and upper approximations, select those attribute-value pairs (a, v) satisfying the condition $[(a, v)] \cap X \neq \emptyset$, where X represents the lower or upper approximation of the concepts as the case may be.

These attribute-value pairs are then used in the LEM2 algorithm to generate a local covering for the set X. Each member of the local covering is a minimal complex, which corresponds to a single rule. Hence the local covering represents a set of association rules generated from the decision table satisfying the set X. By changing the set X, all

Corresponding Author: Mr. Bruce Mathew, GRD College of Science, Coimbatore, Tamilnadu, India.

ISSN 2348 – 9928
Doi:01.0401/ijaict.2014.05.02 Published on 05 (1) 2016

the decision rules satisfying the lower and upper approximations of the remaining concepts are also generated separately by using the same algorithm [11].

The rules thus generated are classified into certain rules and possible rules. The rules generated are as follows:

- ***Certain rules***

Rule 1. (SST = 7) & (SLP = 1) => (Total = 1);
Rule 2. (AT = 10) & (East_wind = 10) => (Total = 1);
Rule 3. (East_wind = 11) => (Total = 2);
Rule 4. (SST = 2) => (Total = 2);
Rule 5. (SST = 1) => (Total = 2);
Rule 6. (SST = 7) & (AT = 8) => (Total = 2);
Rule 7. (SLP = 11) & (TC = 2) => (Total = 2);
Rule 8. (AT = 12) & (SLP = 6) => (Total = 2);
Rule 9. (AT = 10) & (East_wind = 12) => (Total = 2);
Rule 10. (SLP = 10) & (East_wind = 9) => (Total = 2);
Rule 11. (SLP = 3) & (TC = 10) => (Total = 3);
Rule 12. (AT = 3) & (East_wind = 10) => (Total = 3);
Rule 13. (SST = 6) & (TC = 9) => (Total = 4);
Rule 14. (SLP = 7) => (Total = 5);
Rule 15. (TC = 6) & (East_wind = 4) => (Total = 5);
Rule 16. (SST = 10) => (Total = 7);
Rule 17. (SST = 6) & (TC = 4) & (East_wind = 4) => (Total = 8);
Rule 18. (TC = 3) => (Total = 9);
Rule 19. (AT = 4) => (Total = 10);
Rule 20. (SLP = 5) & (East_wind = 7) => (Total = 11);
Rule 21. (SST = 6) & (East_wind = 5) => (Total = 11);
Rule 22. (SST = 12) & (TC = 4) => (Total = 11);
Rule 23. (SST = 11) => (Total = 12);
Rule 24. (SST = 3) & (AT = 1) => (Total = 12);
Rule 25. (SLP = 12) => (Total = 12);
Rule 26. (SLP = 10) & (TC = 2) => (Total = 12);
Rule 27. (SST = 9) & (TC = 9) => (Total = 12);
Rule 28. (TC = 5) => (Total = 12);
Rule 29. (SLP = 8) => (Total = 12);
Rule 30. (East_wind = 1) => (Total = 12);
Rule 31. (SLP = 5) & (East_wind = 9) => (Total = 12);
Rule 32. (AT = 8) & (East_wind = 4) => (Total = 12);

- ***Possible rule***

Rule 33. (East_wind = 8) => (Total = 2) OR (Total = 6);

These rules can be read as follows. In the case of rule 1, when we say that SST=7 and SLP=1, then the predicted catch =1, means that if SST falls in the range 28.75956 to 28.92879, and SLP falls in the range 1007.252 -1008.006, then total catch will be in the range 2968.5 to 6763.9 tones of Oil sardine. Likewise rule 24 is read as if SST=3 and AT=1, then total=12. This implies that if SST falls in the range 27.73026 to 28.36893 and AT in the range 26.9874 to 27.2605, then the predicted total is 54146.64 to 92154.68. This representation is the same with all other rules and the actual range can be referred from table 1.

V. CONCLUSION

This work focused on creating a novel knowledge management model for fisheries domain with special emphasis to the catch of oil sardine. The model produced number of decision rules from the available data using rough set theory. In this paper, we have used the LEM2 algorithm for generating decision rules. These decision rules are very much useful for a knowledge system and hence, this rule forms the backbone logic for prediction and thus satisfies the role of a knowledge system. The accuracy of the rules were analyzed from the actual data collected and found correct. This work had limitations on the number of attributes studied. More accuracy on rules can be attained if the number of attributes considered could be increased. Further, since the raw data was continuous in nature, fuzzy logic can be used for preprocessing which could have significantly improved the results.

References

[1] King W.R, Knowledge Management & Organizational Learning, Annals of Information Systems 4, Springer Science+BusinessMedia LLC.

[2] Davenport, T.H., & Prusak, L. (2000),Working Knowledge: How Organizations Manage What They Know, Harvard Business School Press, Boston, MA.

[3] Pawlak, rough Set theory and its applications, Journal of telecommunications and Information Technology, 3/2002.

[4] Sabu & Raju, Rule Induction using Rough Set Theory-An applicationin Agriculture, 2013.

[5] Rough Set explorer System, [Online], http://idss.cs.put.poznan.pl /site/rose.html (Accessed: 08 September 2014).

[6] The R project for statistical computing [Online], http://www.r-project.org/ (Accessed: 08 September, 2014)

[7] Xu C. & Boyce S. M. ,Oil sardine (Sardinella longiceps) off the Malabar Coast: density dependence and environmental effects, Fish. Oceanogr. 18:5, 359–370, 2009

[8] Sotiris Kotsiantis, Dimitris Kanellopoulos, Discretization Techniques: A recent survey, GESTS International transactions on Computer Science and Engineering, Vol.32 (1), 2006, pp. 47-58.

[9] Kurgan & cios, Caim Discretization Algorithm, Ieee Transactions On Knowledge And Data Engineering, vol. 16, no. 2, february 2004.

[10] Salamó and López-Sánchez, Rough Set ased approaches to feature selection for Case-Based Reasoning classifiers, , 2011

[11] Jiye Li, Nick Cercone, “Discovering and Ranking Important Rules” KDM Workshop, Waterloo, Canada, Oct 30-31, (2006).

[12] Jerzy W. Grzymala-Busse: Rough Set Theory with Applications to Data Mining.

[13] Manjusha, Jayasankar, Remya, Ambrose and Vivekanandan, Influence of coastal upwelling on the fishery of small pelagic fishes off Kerala, south-west coast of India, CMFRI, Kochi.

Corresponding Author: Mr. Bruce Mathew, GRD College of Science, Coimbatore, Tamilnadu, India.