**Machine Learning Project Activity – III**

**Detailed Problem Definition**

**Name: B. THANIKAIVEL Roll no:1613419146 Date: 3/3/16**

1. **Dataset Name with link**

-Cloud Data Set

-https://archive.ics.uci.edu/ml/datasets/Cloud

1. **Dataset Description**

CLOUD DATABASE

The data sets we propose to analyse are constituted of 1024 vectors, eachvector includes 10 parameters. You can think of it as a 1024\*10 matrix.

To produce these vectors, we proceed as follows:

- we start with two 512\*512 AVHRR images

(1 in the visible, 1 in the IR)

- each images is divided in super-pixels 16\*16 and in each

super-pixel we compute a set of parameters.

visible: mean, max, min, mean distribution, contrast,entropy, second angular momentum

IR: mean, max, min

The set of 10 parameters we picked to form the vectors is a compromised betweenvarious constraints. Actually we are still working on the choice of parametersfor the data vectors. The data set I send you has not been normalized. Thenormalization of the data set is required by our classification scheme but thatmay not be true for yours.

To normalize the data we compute the mean andstandard deviation for each parameter on the entire data set then for eachparameter of each vector we compute:

Norm. value = (un-norm value - mean)/SD where

mean = mean value for this particular parameter over the data set

SD = standard deviation

1. **Tools (if any) – Description**

Weka is a collection of machine learning algorithms for data mining tasks. The algorithms can either be applied directly to a dataset or called from your own Java code. Weka contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization. It is also well-suited for developing new machine learning schemes.

Weka is open source software issued under the GNU General Public License.

"WEKA" stands for the Waikato Environment for Knowledge Analysis, which was developed at the University of Waikato in New Zealand. WEKA is extensible and has become a collection of machine learning algorithms for solving real-world data mining problems. It is written in Java and runs on almost every platform. WEKA is easy to use and to be applied at several different levels. We can access the WEKA class library from our own Java program, and implement new machine learning algorithms.

There are three major implemented schemes in WEKA.

(1) Implemented schemes for classification.

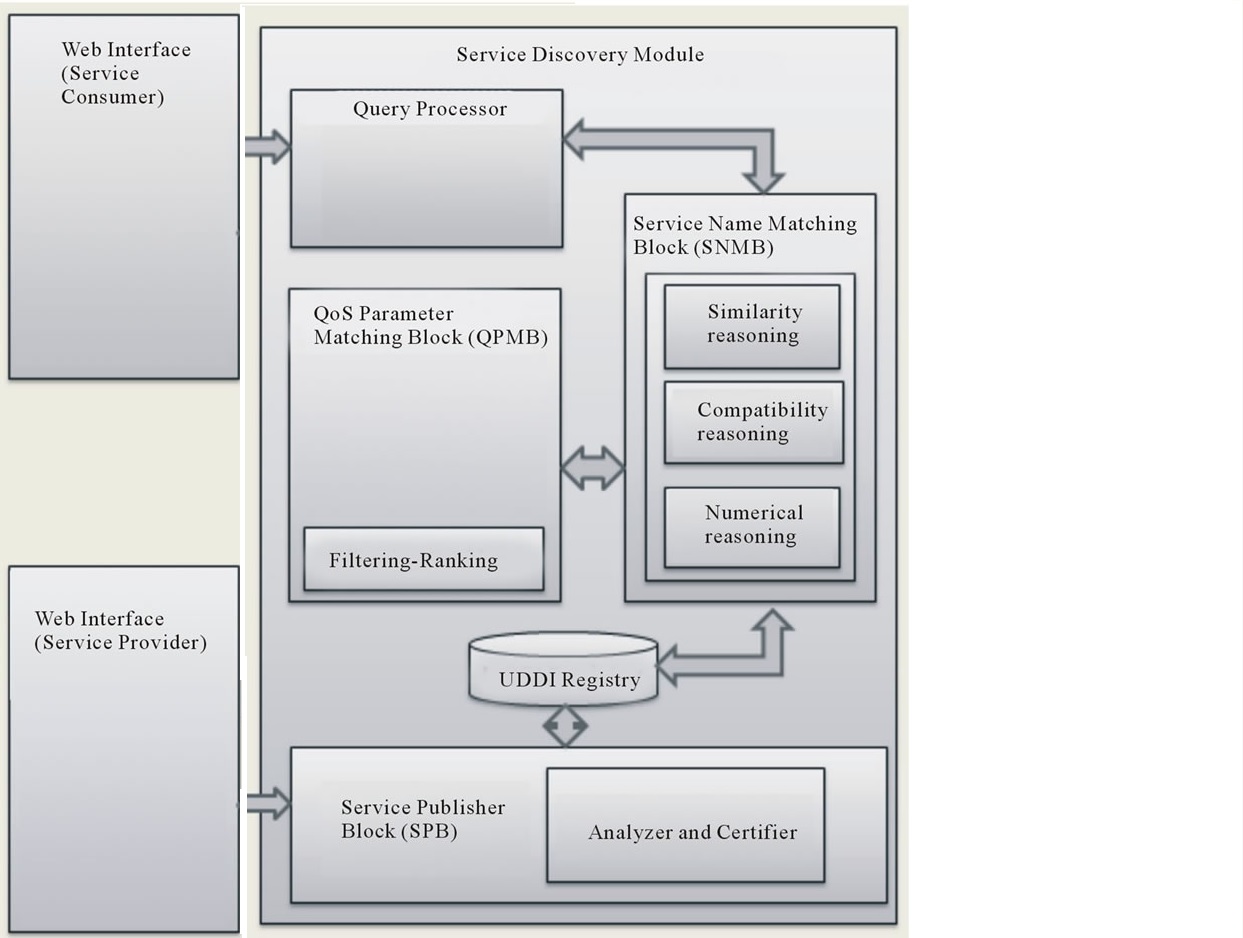
(2) Implemented schemes for numeric prediction.

(3) Implemented "meta-schemes”.

1. **Evaluation parameters to be used**

* The performance is often quite basic and associated with mean resource requirements of the applications
* The workload analysis and inference techniques in the cloud
* cloud systems an important role is played by resource pricing models.

**5. Detailed Block diagram**

****