ISSN: 0974 - 7435 Volume 12 Issue 4

BioTechnology
An Indian Journal

Regular Paper

BTAIJ, 12(4), 2016

Support vector regression to predict survival of *Lactobacillus acidophilus* in concentrated yoghurt

Mozhgan Nasiri Shahri, Ali Mohamadi Sani*, Maryam Tavakoli Fadiheh
Young Researchers and Elite Club, Quchan Branch, Islamic Azad University, Quchan, (IRAN)
E-mail: mohamadisani@yahoo.com

ABSTRACT

In this paper we present a function to predict the survival of *Lactobacillus acidophilus* (LA) in concentrated yoghurt. For this purpose we used Artificial Intelligence tools based on Support Vector Machines for Regression (SVR). Various parameters including: pH, percentage of prebiotic compounds (inulin and oligo-fructose) and inoculum dosage of probiotic bacteria which are effective factors on LA survival were considered. Performance of developed model was evaluated by calculating the mean square error (MSE). The results showed that the mean square error on days 1, 7, 14 and 21 were $1.04x10^{-5}$, $1.08x10^{-5}$, $9.56x10^{-6}$, $7.73x10^{-6}$ respectively and defined model had the capacity of estimation accuracy for predicting survival of LA during storage in the refrigerator. Low values of MSE indicate that SVR is able to predict LA count in concentrated yoghurt.

KEYWORDS

Support Vector Regression (SVR);
Lactobacillus Acidophilus;
Concentrated yoghurt.

INTRODUCTION

Probiotic dairy products are very important due to their large effects on peopleís health. Probiotic bacteria prevent the growth of harmful bacteria and have a positive effect on gastrointestinal tract through positive impact on beneficial intestinal flora. They also increase food digestion and sustainability of body, enhance immune system function and resist against spread and infection of disease[15]. Probiotics also balance the intestine acidity through producing lactic acid, hydrogen peroxide and acetic acid and prevent the production of pathogenic bacteria. Some probiotics produce organic compounds called bacteriocin that is a natural antibiotic compound.

Yogurt can be a suitable environment for storage and transfer of probiotic bacteria in body due to specific physicochemical parameters[3]. Starters which have been commonly used in probiotic yogurts include *Streptococcus thermophilus* and *Lactobacillus bulgaricus* and different strains of probiotic bacteria such as *Lactobacillus acidophilus*, *Bifidobacterium lactis* and *Bifidobacterium bifidum*. Physicochemical parameters of yogurt such as pH, acidity, storage temperature, oxygen, incubation time and primary initiator can be effective on survival of probiotic bacteria during storage[3,7].

Support Vector Machines (SVM) are very specific class of algorithms, characterized by usage of kernels, absence of local minima, sparseness of the solution and

Regular Paper

capacity control obtained by acting on the margin, or on number of support vectors, etc. They were invented by Vladimir Vapnik and his co-workers, and first introduced at the Computational Learning Theory (COLT) 1992 conference with the paper[18]. Still it contains all the main features that characterize maximum margin algorithm: a non-linear function is leaned by linear learning machine mapping into high dimensional kernel induced feature space. The capacity of the system is controlled by parameters that do not depend on the dimensionality of feature space.

In the same way as with classification approach there is motivation to seek and optimize the generalization bounds given for regression. They rely on defining the loss function that ignores errors, which are situated within the certain distance of the true value.

Many researchers have used artificial intelligence in order to achieve desired objectives in food industry such as: predicting microbial growth, evaluating olive oil fraud, predicting persisting flour wheat in bakery, predicting the combined effect of temperature, pH and water activity on heat inactivity of bacteria, evaluating sensitive features of noodles, prediction of moisture as a function of thermal conductivity of food, predicting shelf life and sensory quality of instant coffee drink.

Examining various sources suggests that different models have been used to predict the growth of probiotic bacteria. Kiwi Hardjo *et al* used RSM models in order to optimize and determine coefficients of *B. langum* growth. Sofo and Ekinci[13] estimated shelf life of yogurt using artificial intelligence model.

To learn the function able to predict survival of probiotic bacteria we used SVM for regression (SVR). We discuss the use of different options to configure this algorithm using kernels functions. We used this technology since SVM (SVR) is acknowledged as the most powerful learning algorithms in many application fields.

From a mathematical point of view, given a classification (or regression) learning task, a SVM (SVR) solves a convex optimization problem. The solution gives rise to a hypothesis able to predict unseen cases drawn with the same distribution of the initial learning task. The advantage of the convexity in this context is that it guarantees that there exists only one optimal solution. Therefore the optimizer of SVM (or SVR) will not return a local minimum instead of the best one as happens, for instance with Artificial Neural Networks.

In this study, viability of a probiotic bacteria (*Lactobacillus Acidophilus*) was reviewed during days 1, 7, 14 and 21 after production and its value is predicted using artificial intelligence models.

Support Vector Regression (SVR)

The foundations of Support Vector Machines (SVM) have been developed by Vapnik (1998) and are widely used due to many attractive features and

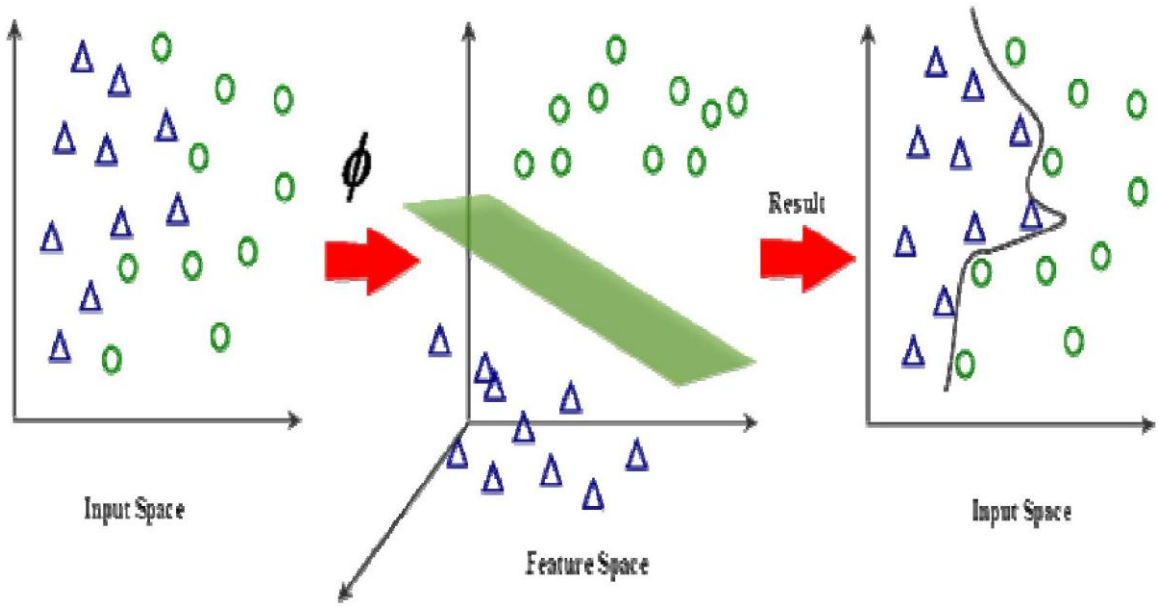

Figure 1 : Converting the input space to feature space

promising empirical performance. SVMs were developed to solve classification tasks, and then they have been extended to handle regression tasks (Smola, 1996; Vapnik, 1998), in this case these algorithms are called Support Vector Regression (SVR).

The formal presentation of SVR starts with a dataset $S = f(x_1 ; y_1), \ddot{O}., (x_n ; y_n)$

Consisting of instances described by pairs $(x_i ; y_i)$, where $x_i \in R^d$ and yi $\in$ R. Each y_i is the desired target or output value for the input vector x_i. A regression model is learned from these patterns and used to predict the target values of unseen input vectors.

Each y_i is the desired target or output value for the input vector xi. A regression model is learned from these patterns and used to predict the target values of unseen input vectors.

Among the various types of SVR, the most commonly used is ε-SVR[17]. The goal is to find a function f(x) that has at most ε deviation from the actually obtained targets y_i for all the training data, and at the same time is as at as possible. In other words, we do not care about errors as long as they are inside the ε-insensitive band (ε-tube). See Figure 1.

Moreover, to make the learning method more robust, the image of the input data does not need to lie strictly on or inside the ε-tube. Instead, the images which lie outside the ε -tube are penalized and slack variables are introduced to take into account for these situations (analogously to the soft margin in SVM for classification). The objective function and constraints are typically given as follows:

$$\text{maximize} \quad \begin{cases} -\frac{1}{2}\sum_{i,j=1}^{\ell}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)\langle x_i, x_j\rangle \\ -\varepsilon\sum_{i=1}^{\ell}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{\ell} y_i(\alpha_i - \alpha_i^*) \end{cases}$$

$$\text{subject to} \quad \sum_{i=1}^{\ell}(\alpha_i - \alpha_i^*) = 0 \text{ and } \alpha_i, \alpha_i^* \in [0, C]$$

$$f(x) = \sum_{k=1}^{N}(\alpha_k - \alpha_k^*)K(x, x_k) + b$$

Where:

$$K\left(x_i^T, x_j\right) = \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\times\sigma^2}\right)$$

That name Gaussian kernel function.

In many problems the relation between outputs and input components is nonlinear and then kernel functions are needed. The idea of the kernel function is to enable operations to be performed in the input space rather than in the potentially high dimensional feature space. An inner product in the feature space has an equivalent kernel in the input space,

One of the most widely adopted kernel function is the radial basis function (RBF) which is defined as inner products in a feature space (Figure 1).

MATERIAL AND METHODS

Materials

Pasteurized milk (Laban Dasht, Iran), Milk powder (Golshad, Iran), Inulin (Beneo-*Orafti*, Belgium), Oligu Fructose (Beneo-*Orafti*, Belgium), Starter Culture including traditional yoghurt bacteria,*Lactobacillus delbrueckii subsp*. *Bulgaricus* and *Streptococcus thermophilus* (Christian Hansen, Denmark), Probiotic bacteria, *Lactobacillus Acidophilus* (Christian Hansen, Denmark), MRS agar (Merck).

Yoghurt preparation

Concentrated symbiotic yoghurt was produced by whey less method. Prebiotic compound included Inulin and Oligo fructose added to milk with other powder material at 3 levels (0, 1/5 and 3%). Probiotic microorganism was *Lactobacillus acidophilus* that added to milk with traditional yoghurt bacteria. Physicochemical properties and survival of probiotic microorganism was assessed at times 1, 7, 14 and 21 after production. TABLE 1 shows different treatments in formulation of concentrated symbiotic yogurt.

Data selection

After preparing concentrated symbiotic yogurt sample (including *Lb. acidophilus*, inulin and oligo-fructose fibers), they tested on days 1, 7, 14 and 21 after production for pH, syneresis and count of the probiotic bacteria. These parameters were assayed because they are important factors affecting on survival of the probiotic bacteria. So these variables were set as input of model (independent variables) and prediction of LA count as output of the model (dependent

Regular Paper

TABLE 1 : Used treatment for sample producing

Treatment	Inoculation (%V/V)	Oligofroctose (%W/V)	Inulin (%W/V)
1	0/5	0	3
2	0/3	1/5	1/5
3	0/5	0	0
4	0/5	3	0
5	0/1	1/5	1/5
6	0/5	3	3
7	0/5	1/5	1/5
8	0/3	1/5	1/5
9	0/3	1/5	0
10	0/3	1/5	1/5
11	0/3	1/5	1/5
12	0/3	3	1/5
13	0/3	0	1/5
14	0/1	3	3
15	0/3	1/5	1/5
16	0/3	1/5	1/5
17	0/1	0	0
18	0/1	0	3
19	0/1	3	0
20	0/3	1/5	3

variables).

Data cleansing and preparation

The second step is cleansing and preparation of data. In this step, data that independent variables do not present due to incomplete information or do not calculated are removed. Among independent selected variables, 12 variables were selected using conducted tests according to standard methods.

Data division using 10-Fold Cross-Validation

We need to divide data into two categories of education- validation and assessment data before entering data into models for this aim 10-Fold Cross-Validation method is used. In this method, data set (test set) is randomly divided into k equal parts so that there are 2 samples for study data in each section that are totally 20 treatments which are selected randomly among 20 samples. K pairs of $\{x_i, y_i\}_{i=1}^{K}$ are extracted randomly where x_i is independent variable and y_i is the dependent variable of i[th] sample. In conducting first part of 10 parts, a part is used for evaluation and 9 remaining parts are used for learning data. Among 9 learning parts, 1 part is used for validation data and remaining is used for evaluation data. For example, in Figure 2, 10th piece is used as evaluation data and first piece is used as validation data and 2 to 9[th] pieces are used as training data for first conduction. In the second conduction, other part of 10 parts is used for evaluation and 9 remaining parts are used for training-validation. For example, according to Figure 2 in the second conduction, 9[th] piece is used as evaluation data and 2[th] piece as validation data and 1[th] and 3[th] to 8[th] and 10[th] pieces have been used as training data. 10 algorithms are conducted with the same procedure. Figure 2 shows the data segmentation in 10 iterations.

One error rate is calculated per iteration for learning and assessment data and finally, the average error rate obtained will be assigned as error rate of learning and assessment data.

Finally, the average error rate obtained will be assigned as error rate of learning and assessment data. The reason for using this method is that Error Rate is one of the criteria for evaluating a classifier/ regressor which includes different types, generally, comparing calculated error on learning data we cannot perform good judgment about ability of algorithms. The error rate on learning data is usually less than error rate on data which have not been seen in learning process. Based on this argument, we cannot use learning error for comparing two algorithms. This reason is that for more complex models, classifications that have usually more parameters will have more complex borders. The complex border will decrease error rate on learning data compared to simpler models. So, a set of data is required for test in addition to learning data sets. In the case of neural networks, we need a set of data as validation data in addition to learn and test data because of over-fitting phenomenon which is selected

Regular Paper

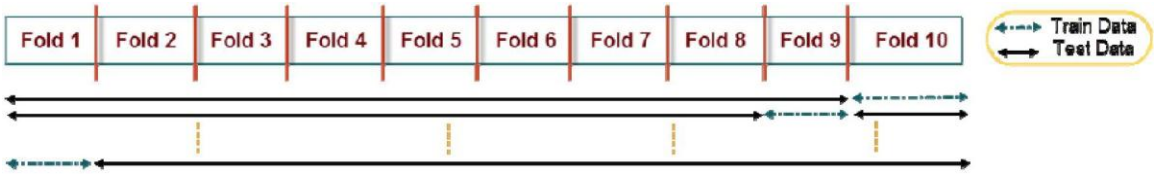

Figure 2 : Selection steps of learning and test data set with k = 10

from learning data set (Over-fitting phenomenon is one of the biggest problems in learning process and one way to avoid that is to use validation data). Thus, each data set is divided into three independent subsets of learning data, validation data and test data. Learning data is used for model training; validation data is used for appropriateness of model parameters and prevent from over-fitting. Test data is used to calculate algorithms error rate (accuracy of model prediction) on data that has not seen. Conducting an algorithm is not sufficient for appropriateness of test. Algorithms usually tend to close the estimated error rate to actual error rate (errors that occur in the real world) and it is possible through implementing and evaluating learning and test processes repeatedly. So when a data set is provided, part of it is set aside for final test and others are used for validation and learning and again three sets are changed and the model is re-tested. One of the common ways to do this is called K-Fold Cross Validation.

Model evaluation and training process

The model is trained when the sample is divided into two categories of learning data (Education Accreditation) and evaluation data. To assess the performance of the predictions, we used absolute differences.

So, if Sí = {(xí$_1$, yí$_1$),Ö., (xí$_m$, yí$_m$)}

is a testing dataset, the performance of a regressor f will be measured by MSE (mean square error), NMSE (normalized mean square error), MAE (mean absolute error), and SMAPE (mean absolute percentage error) defined as follows:

$$\mathbf{MSE} = \frac{1}{n}\sum_{i=1}^{n}(\quad - \quad)^2 \qquad y_i \quad d_i$$

TABLE 2 : Estimated error of training for predicting total count of probiotic bacteria at storage time

	TRAIN											
	MSE				MAE				NMSE			
	1	7	14	21	1	7	14	21	1	7	14	21
1	1.01E-05	1.01E-05	9.38E-06	8.05E-06	0.0025	0.0015	0.0025	0.002	3.86E-05	3.52E-05	3.59E-05	3.63E-05
2	1.11E-05	9.85E-06	1.02E-05	8.1E-06	0.0026	0.0015	0.0027	0.0024	3.95E-05	3.51E-05	3.39E-05	3.5E-05
3	1.01E-05	1.12E-05	9.25E-06	7.62E-06	0.0025	0.0014	0.0025	0.0022	3.8E-05	3.3E-05	3.66E-05	3.78E-05
4	1.13E-05	1.02E-05	9.93E-06	8.06E-06	0.002	0.0015	0.0026	0.0023	3.76E-05	3.48E-05	3.39E-05	3.61E-05
5	1.1E-05	1.12E-05	1.06E-05	7.21E-06	0.0026	0.0014	0.0028	0.0022	3.98E-05	3.28E-05	3.21E-05	3.42E-05
6	1.01E-05	1.08E-05	8.56E-06	7.62E-06	0.0025	0.0016	0.0024	0.0022	3.96E-05	3.5E-05	3.62E-05	3.78E-05
7	1.15E-05	1.19E-05	9.93E-06	8.1E-06	0.0028	0.0013	0.0026	0.0024	3.54E-05	3.2E-05	3.41E-05	3.49E-05
8	8.76E-06	1.12E-05	9.23E-06	7.21E-06	0.0022	0.0014	0.0025	0.0022	4.12E-05	3.31E-05	3.41E-05	3.46E-05
9	8.76E-06	1.04E-05	9.25E-06	7.23E-06	0.0022	0.0015	0.0025	0.0021	4.12E-05	3.53E-05	3.66E-05	3.78E-05
10	1.09E-05	1.12E-05	9.25E-06	8.07E-06	0.0025	0.0014	0.0025	0.0023	4.05E-05	3.27E-05	3.66E-05	3.6E-05
average	1.04E-05	1.08E-05	9.56E-06	7.73E-06	0.0025	0.0014	0.0026	0.0023	3.91E-05	3.39E-05	3.5E-05	3.6E-05

Low values of mean square error indicate that support vector machine for regression (SVR) is able to predict probiotic bacterial total count of concentrated yoghurt during storage in the refrigerator. These numbers defined model has capacity of estimation

Regular Paper

accuracy for predicting survival capability of probiotic bacteria

$$NMSE = \frac{1}{n}\sum_{i=1}^{n}\frac{(y_i - d_i)^2}{\bar{y}\bar{d}}, \quad \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i, \quad \bar{d} = \frac{1}{n}\sum_{i=1}^{n} d_i$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - d_i|$$

$$SMAPE = \frac{\sum_{i=1}^{n}|d_i - y_i|}{n\sum_{i=1}^{n}(d_i + y_i)}$$

Where y_i and d_i are actual total count (obtained from testing) and predicted total count (by algorithm for i^{th} sample) and n is the number of samples (in training or evaluation stage). $\bar{y}$ and $\bar{d}$ are mean rate of real stock and mean rate of predicted stock, respectively. Among above errors, MSE is used more because it minimizes the mean error and the error variance from statistical standpoint. For this reason, it is interested for researchers in almost all researches.

RESULTS AND DISCUSSION

In this section we report the results of a set of experiments designed to evaluate the approach proposed in this paper. The main objective was to check the accuracy of the predictions of survival of probiotic bacteria (*Lactobacillus Acidophilus*) using a SVR algorithm.

TABLE 2-3 reported the errors estimated with 10-fold cross validation using different options for the regressor for training and testing phase. As it be noted MSE is used more because it minimizes the mean error and the error variance from statistical standpoint.

TABLE 3 : Estimated error of testing for predicting total count of probiotic bacteria at storage time

	TEST											
	MSE				MAE				NMSE			
	1	7	14	21	1	7	14	21	1	7	14	21
1	1.24E-05	1.7E-05	1.11E-05	4.79E-06	0.0026	0.0041	0.0031	0.0019	4.35E-05	2.5E-05	2.79E-05	3.16E-05
2	3.86E-06	1.93E-05	3.88E-06	4.36E-06	0.0018	0.0043	0.0014	0.0014	2.67E-05	2.49E-05	4.5E-05	4.91E-05
3	1.24E-05	7.32E-06	1.23E-05	8.72E-06	0.0024	0.0020	0.0035	0.0029	4.98E-05	4.44E-05	2.49E-05	2.49E-05
4	2.32E-06	1.59E-05	6.17E-06	4.7E-06	0.0011	0.0038	0.0017	0.0018	4.75E-05	2.72E-05	4.78E-05	3.27E-05
5	4.25E-06	7.3E-06	1.71E-07	1.24E-05	0.0020	0.0019	0.0002	0.0024	2.58E-05	4.69E-05	4.98E-05	4.94E-05
6	1.3E-05	1.08E-05	1.85E-05	8.72E-06	0.0030	0.003	0.0042	0.0029	3.48E-05	2.57E-05	2.56E-05	2.49E-05
7	6.58E-09	1.14E-06	6.19E-06	4.36E-06	7.69E-05	0.000	0.0018	0.0014	2.76E-05	4.32E-05	4.42E-05	4.98E-05
8	2.48E-05	7.35E-06	1.25E-05	1.24E-05	0.0049	0.0020	0.0027	0.0025	2.49E-05	4.22E-05	4.13E-05	4.57E-05
9	2.48E-05	1.46E-05	1.23E-05	1.22E-05	0.0049	0.0038	0.0035	0.0034	2.49E-05	2.49E-05	2.49E-05	2.54E-05
10	5.81E-06	7.29E-06	1.23E-05	4.62E-06	0.0024	0.0019	0.0035	0.001	2.49E-05	4.98E-05	2.49E-05	3.4E-05
average	1.04E-05	1.08E-05	9.56E-06	7.73E-06	0.00256	0.0028	0.0026	0.002	3.3E-05	3.54E-05	3.56E-05	3.67E-05

Regular Paper

CONCLUSIONS

We have presented a method to estimate the survival of probiotic bacteria (*Lactobacillus Acidophilus*) from empirical experiments. For this purpose we used an Artificial Intelligence tool, the Support Vector Machines for Regression (SVR). We discussed the use of different options for configuring these learning algorithms. The results show that using a nonlinear function it is possible to achieve accurate predictions for predicting the survival of probiotic bacteria. Calculated error are so small and it means that support vector machine for regression (SVR) is able to predict probiotic bacterial total count of concentrated yoghurt during storage in the refrigerator.

REFERENCES

[1] A.Rakotomamonjy; Variable selection using SVM-based criteria. Journal of Machine Learning Research, **82(3)**, 1357ñ1370 **(2003)**.

[2] M.Alborzi; Introduction to Neural Networks, 2nd Edition. Sharif University Press. AOAC. Official methods of analysis of the AOAC, 15thed. (S.Williams (Ed.)), Arlington, USA: Association of Official Analytical Chemists, **(2002)**.

[3] M.Bari, R.Ashrafi, M.Alizadeh, L.Rofehgarineghad; Effects of different of yogurt starter or probiotic bacteria, storage time & different concentration of cysteine on the micro flora characteristics of BioñYogurt. Research Journal of Biological Sciences, **4(2)**, 137-142 **(2009)**.

[4] C.J.Lin; On the convergence of the decomposition method for support vector machines. IEEE Transactions on Neural Networks, **12(6)**, 1288ñ 1298 **(2001)**.

[5] C.C.Chang, C.J.Lin; LIBSVM: A library for support vector machines.ACM Transactions on Intelligent Systems and Technology, **2(27)**, 27-37 **(2011)**.

[6] J.J.Del Coz, G.F.BayÛn, J.DÌez, O.Luaces, A.Bahamonde, C.SaÒudo; Trait selection for assessing beef meat quality using nonlinear SVM. Proceedings of the 18th Annual Conference on Neural Information Processing Systems (NIPS). Vancouver, British Columbia, Canada, 321-328 **(2004)**.

[7] O.N.Donkor, S.L.I.Nilmini, P.Stolic, T.Vasilgevic, N.P.Shah; Survival & activity of selected probiotic organism in set-type yoghurt during cold storage. International Dairy Journal, **17(4)**, 92-101 **(2007)**.

[8] E.Boser, I.M.Guyon, V.N.Vapnik; A training algorithm for optimal margin classifiers. In D. Haussler, editor, *Proceedings of the Annual Conference on Computational Learning Theory*, Pittsburgh, PA, ACM Press, 144-152 July **(1992)**.

[9] S.Ganga, Meena, Suneel Gupta, C.Gautam, Majumdar, Rintu Banerjee; Growth Characteristics Modeling of *Bifidobacterium bifidum* using RSM and ANN, Gautam C., **54(6)**, 1357-1366 **(2011)**.

[10] Y.Huang, L.J.Kangas, B.A.Rasco; Applications of artificial neural networks (ANNs) in food science. Critical Reviews in Food Science and Nutrition, **47(2)**, 113-126 **(2007)**.

[11] J.C.Burges; A tutorial on support vector machines for pattern recognition. Data Mining and Knowledge Discovery, **2(2)**, 121ñ167 **(1998)**.

[12] T.Joachims; Making large-scale SVM learning practical. Advances in Kernel Methods Support Vector Learning, MIT Press, Cambridge MA. Chapter 11, **(1998)**.

[13] A.Sofu, F.V.Ekinci; Estimation of storage time of yogurt with artificial neural network modeling. Journal of Dairy Science., **90(7)**, 18-25 **(2007)**.

[14] Sumit Goyal, G.K.Goyal; Cascade and feed forward back propagation artificial neural networks models for prediction of sensory quality of instant coffee flavored sterilized drink, Advances in Computer Science and its Applications, **93(1)**, 23-32 **(2012)**.

[15] S.Yeganehzad, M.Mazaheri-Tehrani, F.Shahidi; Studying microbial, physicochemical & sensory properties of directly concentration probiotic yogurt. African Journal of Agricultural Research, **2(8)**, 366-369 **(2007)**.

[16] Y.J.Lee, O.L.Mangasarian; SSVM: A smooth support vector machine for classification. Computational optimization and Applications, **20(1)**, 5ñ22 **(2001)**.

[17] V.N.Vapnik; Statistical Learning Theory. Wiley-Inter science, New York, **(1998)**.

[18] V.Vapnik; The Nature of Statistical Learning Theory, 2nd Edition. Springer, New York, **(2001)**.