**IOSR Journal of Engineering (IOSRJEN)** **www.iosrjen.org**
ISSN (e): 2250-3021, ISSN (p): 2278-8719
Vol. 08, Issue 10 (October. 2018), ||V (I) || PP 36-39

# Comparison of Naïve Bayes and Least-Square Support Vector Machine classifier for prediction of Swine Flu

Dr. Gurmanik Kaur[1], Er. Arwinder Kaur[2]
[1](A.P., Deptt. of EE, SBBSU, Jalandhar, Punjab, INDIA
[2](M. Tech. Scholar, Deptt. of ECE, SBBSU, Jalandhar, Punjab, INDIA
*Corresponding Author: Dr. Gurmanik Kaur*

**Abstract: -** Disease prediction has long been regarded as a critical topic. Artificial intelligence and machine learning techniques have already been developed to solve this type of medical care problem. Our research focuses on this aspect of medical diagnosis by learning pattern through the collected data for swine flu. This research has developed naïve bayes and least-square support vector machine (LS-SVM) classifier for predicting the presence or absence of swine flu. We have generated 96 symptoms sets after consulting with medical practitioners from various hospitals of Punjab, INDIA. Using LS-SVM, we have achieved better prediction accuracy (100%) as compared to naïve bayes model . This assessment presents the importance and advantages posed by LS-SVM model for prediction of biological variables.

**Keywords: -** Swine flu, prediction, naïve bayes, least-square support vector machine

---

Date of Submission: 29-09-2018 Date of acceptance: 14-10-2018

---

## I. Introduction

In April 2009, a novel strain of H1N1 influenza (swine flu) jumped from swines into humans and infected over 200 million people globally, resulting in the first influenza pandemic of the 21st century [1]. WHO declares on June 11, 2009, that swine flu pandemic. There have been nearly 30,000 confirmed swine flu cases across 74 countries. The reports have shown a sharp increase in the number of infections reported in recent days from Chile, Japan, United Kingdom (UK) and other parts of the world, with the most dramatic increase recorded in Australia where more than 1200 cases were reported in a very short duration. The mortality and morbidity due to swine flu continues to remain high in India. In mid-March 2015, the total number of laboratory confirmed cases of pandemic influenza A (H1N1) 2009 virus in India was 29,978, with death of 1793 people. Rajasthan (6203 cases, 378 deaths) and Gujarat (6150 cases and 387 deaths) were the worst affected States as per the data communicated (personal communication) by the Emergency Medical Relief (EMR), Ministry of Health and Family Welfare, Government of India (MOHFWGOI) [2]. What makes this disease worse is the fact that its symptoms resemble those of regular flu. According to medical practitioners the differentiation with swine flu and normal flu is only possible through pathological tests and these laboratory tests are unnecessary for swine flu [3, 4]. An exhaustive case study was carried out on the detection of swine flu wherein various doctors were interviewed and it was found that out of the 10 cases of suspected swine flu, it was very difficult for the doctors to categorize the various flu only and only on the basis of symptoms, but there are ways by which this can be done. These are the various data mining techniques which have been successfully utilized for a highly accurate analysis and modeling of multifaceted and raw biological data. However, there have been very few studies conducted on prediction of swine flu by means of data mining techniques. In this context, Thakkar *et al.* [5] have developed a prototype intelligence swine flu prediction software (ISWPS). They have used 17 symptoms of swine flu and collected 110 symptoms sets from various hospitals and medical practitioners. Naive bayes classifiers used for classifying the patients of swine-flu had classified the patients into three categories (least possible, probable or most probable). It was reported that the efficiency of the results can be further improved by increasing the number of data set, attributes or by selecting weighted features. Borkar and Deshmukh [6] have also proposed naive bayes classifier algorithm for diagnosis of swine-flu disease from its symptoms. The proposed approach showed promising results which may lead to further attempts to utilize information technology for diagnosing patients for swine flu disease. Shinde and Pawar [7] used clustering algorithm K-means to make a group or cluster of swine flu suspects in a particular area. The decision tree algorithm and naive bayes classifier was applied to the same inputs to find out the actual count of suspects and predict the possible surveillance of swine flu in a nearby area from suspected area. The performances of these techniques when compared, the naive bayes classifier performed better than the decision tree algorithm in finding the accurate count of suspects. Tate *et al.* [8] proposed random forest algorithm for prediction of swine flu from input symptoms taken from the patient or user. It was concluded that the random forest algorithm

maintained best accuracy as compared to other classification systems. The proposed algorithm is extendible to deal with mobile/online solutions to support patients as well for medical diagnostics.

This research work focused on the development of naïve bayes and least square support vector machine (LS-SVM) model for swine flu prediction from its symptoms in human beings. The prediction accuracy of the developed models was assessed using coefficient of determination ($R^2$), root mean square error (RMSE) and time taken to generate model.

## II. Data Source

As per guidelines of MOHFWGOI (Revised on 11.02.2015), on swine flu, patients with mild fever and cough /sore throat with or without body ache, headache, diarrhea and vomiting have been considered as swine flu positive [9]. A number of medical practitioners from various hospitals of Punjab were consulted for collection of the most weighted symptoms of swine flu and on the basis of collected symptoms 96 cases have been generated, and authenticated the same with the practitioners.

## III. Experimental Methods

### 3.1 Naïve Bayes

Naïve bayes classifier is a probabilistic classifier based on the Bayes theorem. Rather than predictions, the Naïve Bayes classifier produces probability estimates. For each class value they estimate the probability that a given instance belongs to that class. It assumes that the effect of an attribute value on a given class is independent of the values of the other attributes. This assumption is called class conditional independence.

Let $X_{ij}$ be a dataset sample containing records of $i$ number of risk factors alongside their respective diagnosis of a disease, $C$ (target class) collected for $j$ number of records and $H_k$= *{$H_1$ = Yes, $H_2$ = No}* be a hypothesis that $X_{ij}$ belongs to class $C$. For the classification of the risk of disease given the values of the risk factor of the $j^{th}$ record, Naive Bayes classification required the determination of the following:

$P(H_k | X_{ij})$ - Posteriori probability: is the probability that the hypothesis, $H_k$ holds given the observed data sample $X_{ij}$ for $1 \leq k \leq 2$.

$P(H_k)$ - Prior probability: is the initial probability of the target class $1 \leq k \leq 2$

$P(X_{ij})$ is the probability that the sample data is observed for each risk factor $i$

$P(|X_{ij}|H_k)$ is the probability of observing the sample's attribute,

Therefore, the posteriori probability of an hypothesis $H_k$ is defined according to Bayes' theorem as follows:

$$P(H_k | X_{ij}) = \frac{\prod_{i=1}^{n} P(H_k | X_{ij}) P(X_i)}{P(H_k)} \text{ for } k=1,2 \quad (1)$$

Hence, the risk of disease for a record is thus [10-12]:

$$max.[P(H_1|X_{ij}), P(H_2|X_{ij})] \quad (2)$$

3.2 LS-SVM:

LS-SVM models are based on an alternate formulation of SVM regression [13], proposed by Suykens *et al.*[14]. It considers equality type of constraints instead of inequalities as in the standard SVM approach. This leads to solving a set of linear equations instead of a quadratic programming (QP) problem. Thus, LS-SVM reduces the computational complexity.

Consider a set of $N$ data points $\{x_1, y_1, x_2, y_2, \ldots, x_N, y_N\}$, where $x_i \in R^N$ is the $i^{th}$ input vector and $y_i \in R$ is the corresponding output. In the feature space, LS-SVM model thus takes the form:

$$y(x) = w^T \Phi(x) + b \quad (3)$$

where,

$\Phi(x)$= non-linear function that maps the input data into a higher dimensional feature space

$w$=an adjustable weight vector

$b$= the bias term

In LS-SVM, for function estimation, the following optimization problem is formulated:

$$minimize \ \frac{1}{2} w^T w + \gamma \frac{1}{2} \sum_{i=1}^{N} e_i^2 \quad (4)$$

subject to the following equality constraint:

$$y=w^T \Phi(x_i)+b+e_i \quad (5)$$

where,
$e_{i=}$ the error vector *(i=1, 2, …, N)*
$\gamma$= the regularization parameter.
Solving the above optimization problem in dual space leads to finding the coefficients $\alpha_i$ and $b$ in the following solution:

$$y(x) = \sum_{i=1}^{N} \alpha_i K(x, x_i) + b \quad (6)$$

where,
$K(x, x_i)$= the kernel function defined as the dot product between $\Phi(x)$ and $\Phi(x_i)$.
Commonly used kernel functions are polynomial, gaussian, linear, sigmoid, and radial basis functions (RBF) [15].

## IV. Results and Discussion

Out of the total available cases, 50% were used for training and 50% for testing of developed algorithms.

A naive bayes classifier with 2 classes for 7 dimensions was developed using 'kernel' distribution. For each symptom modeled with kernel distribution, the naïve bayes classifier computed a separate kernel density for each class based on the training data for that class.

To develop LS-SVM prediction model, the most important steps are kernel and parameter selection, because they can significantly affect the model performance; consequently, we used the RBF kernel and applied the grid search optimization algorithm with 2-fold cross-validation to obtain the optimal parameter combination or minimum mean square error (MSE). The optimal values of regularization parameter ($\gamma$) and squared bandwidth ($\sigma^2$) were 0.7528 and 5.6768 respectively.

Comparison of the performance indices of the developed models, as shown in Table 1, revealed that the LS-SVM model has the higher value of $R^2$ and lower value of RMSE for prediction of swine flu's presence/ absence. Moreover, the time taken to build model (in seconds) is also minimum in case of LS-SVM.

**Table 1- Comparison of the models using various performance indices**

| Performance indices | Naïve bayes | LS-SVM |
|---|---|---|
| $R^2$(%) | 66.67 | 100 |
| RMSE | 0.5774 | 0 |
| Time taken to generate model (in seconds) | 1.3084 | 0.0712 |

## V. Conclusion

A performance comparison of naïve bayes and LS-SVM model revealed potential capability of the latter in predicting the presence or absence of swine flu. The results of this study are highly encouraging and may provide valuable reference for researchers and engineers who are interested in applying LS-SVM model for the prediction of biological variables. The results may also be helpful in physician's diagnosis, in clinical medicine. Our future research is targeted at studying a hybrid model for prediction of swine flu.

## References

[1]. W. A. Fischer, M. Gong, S. Bhagwanjee, and J. Sevransky, Global burden of Influenza: Contributions from Resource Limited and Low-Income Settings, *Global Heart, 9(3),* 2014, 325-336.
[2]. Emergency Medical Relief (EMR), Ministry of Health and Family Welfare, Government of India. Available from: http://mohfw.nic.in/index3.php?lang=1&level=0&deptid=115.
[3]. G. L. Dandagi, S. M. Byahatti, An insight into the swine-influenza A (H1N1) virus infection in humans, *Lung India*, *28(1)*, 2011, 34-38.
[4]. T. N. Jilani, A. H. Siddiqui, *H1N1 Influenza (Swine Flu)* (Treasure Island, FL: StatPearls Publishing, 2018).
[5]. B. A. Thakkar, M. I. Hasan, M. A. Desai, Health care decision support system for swine flu prediction using naïve bayes classifier, *Proc. IEEE International Conference on Advances in Recent Technologies in Communication and Computing*, Kottayam, India, 2010, 101-105.

[6]. Ms. A. R. Borkar, Dr. P. R. Deshmukh, Naïve Bayes Classifier for Prediction of Swine Flu Disease, *International Journal of Advanced Research in Computer Science and Software Engineering*, *5(4)*, 2015, 120-123.
[7]. M. J. Shinde, S. S. Pawar, Comparative study of decision tree algorithm and naïve bayes classifier for swine flu prediction, *International Journal of Research in Engineering and Technology*, *4(6)*, 2015, 45-50.
[8]. A. Tate, U. Gavhane, J. Pawar, B. Rajpurohit, G. B. Deshmukh, Prediction of Dengue, Diabetes and Swine Flu Using Random Forest Classification Algorithm, *International Research Journal of Engineering and Technology*, *4(6)*, 2017, 685-690.
[9]. Ministry of Health and Family Welfare, Government of India. Guidelines on categorization of Influenza A H1N1 cases during screening for home isolation, testing treatment, and hospitalization (Revised on 11.02.2015). Available from: http://mohfw.gov.in/showfile.php?lid=3071.
[10]. D. Khanna, A. Sharma, Kernel-Based Naive Bayes Classifier for Medical Predictions, in Bhateja V., Coello Coello C., Satapathy S., Pattnaik P. (Ed.) Intelligent Engineering Informatics. *Advances in Intelligent Systems and Computing*, 695 (Singapore: Springer, 2018).
[11]. E. Frank, L. Trigg, G. Holmes, I. H. Witten, Technical Note: Naive Bayes for Regression, *Machine Learning*, *41*, 2000, 5-25.
[12]. K. M. Al-Aidaroos, A. A. Bakar, and Z. Othman, Medical data classification with Naive Bayes approach, *Information Technology Journal*, *11(9)*, 2012, 1166-1174.
[13]. V. Vapnik and A. Lerner, Pattern Recognition Using Generalized Portrait method, Automation and Remote Control, 24(6), 1963, 774-780.
[14]. J. A. K. Suykens, J. D. Brabanter, L. Lukas, and J. Vandewalle, Weighted Least Squares Support Vector Machines: Robustness and Sparse Approximation, *Neurocomputing*, 48(1-4), 2002, pp. 85-105.
[15]. J. A. K. Suykens, J. Vandewalle, Least squares support vector machine classifiers, *Neural Processing Letters*, *9(3)*, 1999, 293-300.

IOSR Journal of Engineering (IOSRJEN) is UGC approved Journal with Sl. No. 3240, Journal no. 48995.

Dr. Gurmanik Kaur1" Comparison of Naïve Bayes and Least-Square Support Vector Machine classifier for prediction of Swine Flu" IOSR Journal of Engineering (IOSRJEN), vol. 08, no. 10, 2018, pp. 36-39.