tuple and test tuples are calculated by using equation 3. In case of Euclidian distance, root square differences of training and test tuple are calculated. By using equation 2, Chebychev distances between tuples are calculated. And based on these distances, k closest training samples are found. Based on majority voting criteria, unknown tuples are classified and class labels are assigned. This kNN classifier is evaluated using different measures as given below.

6. Classifier Evaluation

It is very important that how accurately classifier will predict the class label of test examples. Different measures are available for evaluating the performance of classifies.

- True positive vs. True Negative: True positive means positive tuples that are correctly classified as positive. True negative is negative tuples that are correctly predicted as negative. In case of KDD dataset, true positive is normal classes are classified as normal. And attack tuples are predicted as attack class.
- False positive rate (FPR): It is the fraction of negative tuples that are classified as positive. FPR is calculated using following equations.

$$FPR = \frac{False\ Positive}{(True\ Negative\ + False\ Positive\ )} \quad (4)$$

- False Negative Rate (FNR): It is the fraction of positive tuples that are classified as negative. It calculated using:

$$FNR = \frac{False\ Negative}{(True\ Positive\ + False\ Negative\ )} \quad (5)$$

- Accuracy: It is the percentage of test tuples that are correctly classified by classifier. It is defined as:

$$Accuracy = \frac{True\ positive\ + true\ negative}{positive\ + negative} \quad (6)$$

- Sensitivity: It is proportion of positive tuples that are correctly identified. Sensitivity is calculated using:

$$sensitivity = \frac{true\ positive}{Positive} \quad (7)$$

- Specificity: It is a proportion of negative tuples that are correctly identified by classifier. It is calculated using:

$$Sensitivity = \frac{true\ negative}{negative} \quad (8)$$

7. Experimental Results

Result is evaluated after implementation of K nearest neighbor algorithm using Euclidian distance, Manhattan distance and Chebychev Distance in terms of accuracy, sensitivity and specificity. Experiment is performed on KDD dataset.

Comparative graphs of three distance functions are given below. Manhattan distance gives better accuracy than Chebychev Distance and Euclidian distance as shown in figure 3.

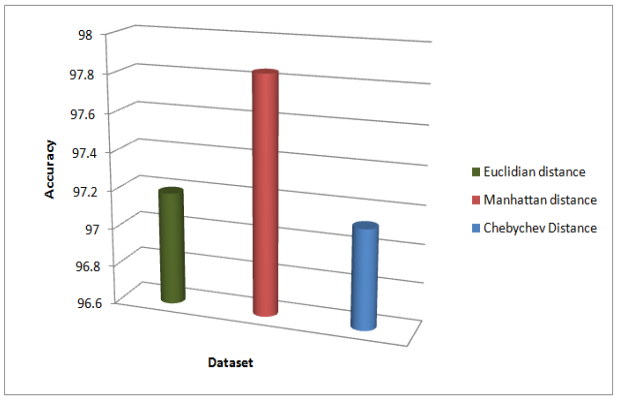

Figure 3: Accuracy Graph

As shown in figure 4, Manhattan distance gives sensitivity rate up to 96.76 %. Chebychev and Euclidian distance gives 95.68% and 95.79% sensitivity rate respectively.

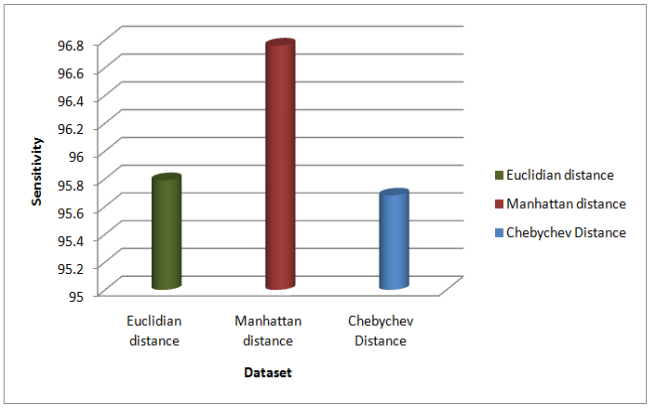

Figure 4: Sensitivity Graph

Following graph describes specificity rate of three distance function. Percentage of negative tuples that are correctly identified by Manhattan distance based K-NN classifier is more as compare to other two distances.

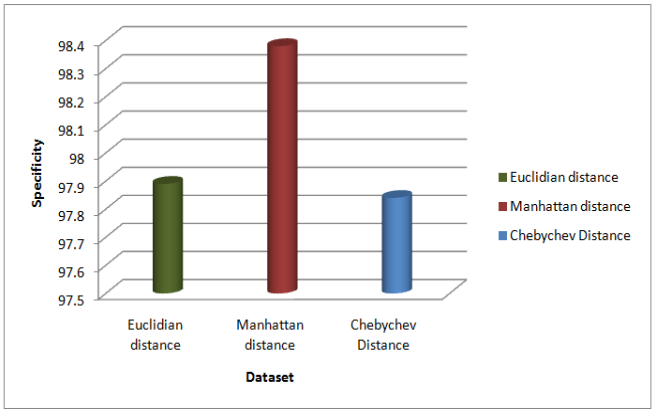

Figure 5: Specificity Graph

Following table shows the false positive rate calculated using distance measures. In case of KDD dataset, false positive rate is attack class that is incorrectly classified as normal class. Manhattan distance based KNN gives lower false positive rate as compare to Chebychev Distance and Euclidian Distance.

Table 2: False positive rate

Chebychev Distance	Euclidian Distance	Manhattan Distance
0. 0215	0.0210	0.016

Number of normal class tuples are predicted as attack class is given in following table 3. FNR is given using these three distance measure on KDD dataset. In case of Manhattan distance, normal classes are classified as attack classes is less as compare to other distance function.

Table 3: False Negative Rate

Euclidian distance	Chebychev Distance	Manhattan distance
0.042	0.043	0.032

8. Conclusion

K-nearest neighbor is implemented using Euclidian distance, Manhattan distance and Chebychev distance on KDD dataset. KNN classifier is evaluated in terms accuracy, specificity, sensitivity and false positive rate. These distance measures are compared. It is observed that Manhattan distance gives better results than other distance measures. Performance of Euclidian distance is low as compare to Chebychev distance.

References

[1] Jiawei Han, Micheline Kamber, Jian Pei, \Data Mining concepts and Technologies", Third Edition Elsevier.
[2] Pang-Nang Tan, Michael Steinbach, Vipin Kumar, "Data Mining".
[3] Nitin Bhatia , Vandana, "Survey of Nearest Neighbor Techniques" (IJCSIS) International Journal of Computer Science and Information Security, Vol. 8, No. 2, 2010
[4] T. M. Cover and P. E. Hart, "Nearest Neighbor Pattern Classification", IEEE Trans. Inform. Theory, Vol. IT-13, pp 21-27, Jan 1967.
[5] T. Bailey and A. K. Jain, "A note on Distance weighted k-nearest neighbor rules", IEEE Trans. Systems, Man Cybernatics, Vol.8, pp 311-313, 1978.
[6] Asha Gowda Karegowda , M.A. Jayaram, A.S. Manjunath "Cascading K-means Clustering and K-Nearest Neighbor Classifier for Categorization of Diabetic Patients", International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-1, Issue-3, February 2012.
[7] Dasarathy, B. V., "Nearest Neighbor (NN) Norms, NN Pattern Classification Techniques". IEEE Computer Society Press, 1990.
[8] "K-means with Three different Distance Metrics", International Journal of Computer Applications (0975 – 8887) Volume 67– No.10, April 2013.
[9] Mahbod Tavallaee, Ebrahim Bagheri, Wei Lu, and Ali A. Ghorbani," A Detailed Analysis of the KDD CUP 99 Data Set", Proceedings of the 2009 IEEE Symposium on Computational Intelligence in Security and Defense Applications (CISDA 2009).
[10] M.Vidhya, "Efficient Classification of Portscan Attacks using Support Vector Machine", Proceedings of 2013 International Conference on Green High Performance Computing, March 14-15, 2013, India