BSCS 7th SEMESTER

STUDENT DETAILS

Student Name: Syed Ahtsham **Student ID Number** 04071813015

ASSIGNMENT DETAILS

Subject Title: Artificial Intelligence

Submission Date: January 17, 2022

Submitted To: Dr. Ayyaz Hussain

TABLE OF CONTENTS

References:

R. J. Lyon, B. W. Stappers, S. Cooper, J. M. Brooke, J. D. Knowles, Fifty Years of Pulsar Candidate Selection: From simple filters to a new principled real-time classification approach, Monthly Notices of the Royal Astronomical Society 459 (1), 1104-1123, DOI: 10.1093/mnras/stw656

HTRU 2 Dataset Information:

Attributes in each cell of HTRU 2.csv dataset file starting from left-hand side to right:

1. Mean of the integrated profile.
2. Standard deviation of the integrated profile.
3. Excess kurtosis of the integrated profile.
4. Skewness of the integrated profile.
5. Mean of the DM-SNR curve.
6. Standard deviation of the DM-SNR curve.
7. Excess kurtosis of the DM-SNR curve.
8. Skewness of the DM-SNR curve.
9. Class

HTRU 2 Summary:

17,898 total examples

1,639 positive examples

16,259 negative examples

Classification Model using K-Nearest Neighbor (K-NN) Supervised Learning Algorithm:

I split 17,898 total examples into 80:20 training set, and testing set.

Training Set: 14318

Testing Set: 3580

Distance and Similarity Measures:

I used 5 different distance and similarity measures, which are as follows:

1. Euclidean Distance
2. Manhattan Distance
3. Infinity Norm
4. Cosine Similarity Measure
5. Intersection Similarity Measure

Value of K:

The value of K is taken 3, 7, 11, 15, 19, and 27. And the model was given test data and used different distance and similarity measures against each value of K.

K is taken such that: **K < Sqrt(N)** where N is the total number of examples in the

Dataset

Accuracy:

I computed the Percentage Accuracy of the model for different measures using the following formula:

Percentage Accuracy = (Total Correct Predictions / Total Predictions Made) * 100

Graphical Representation of K vs Accuracy:

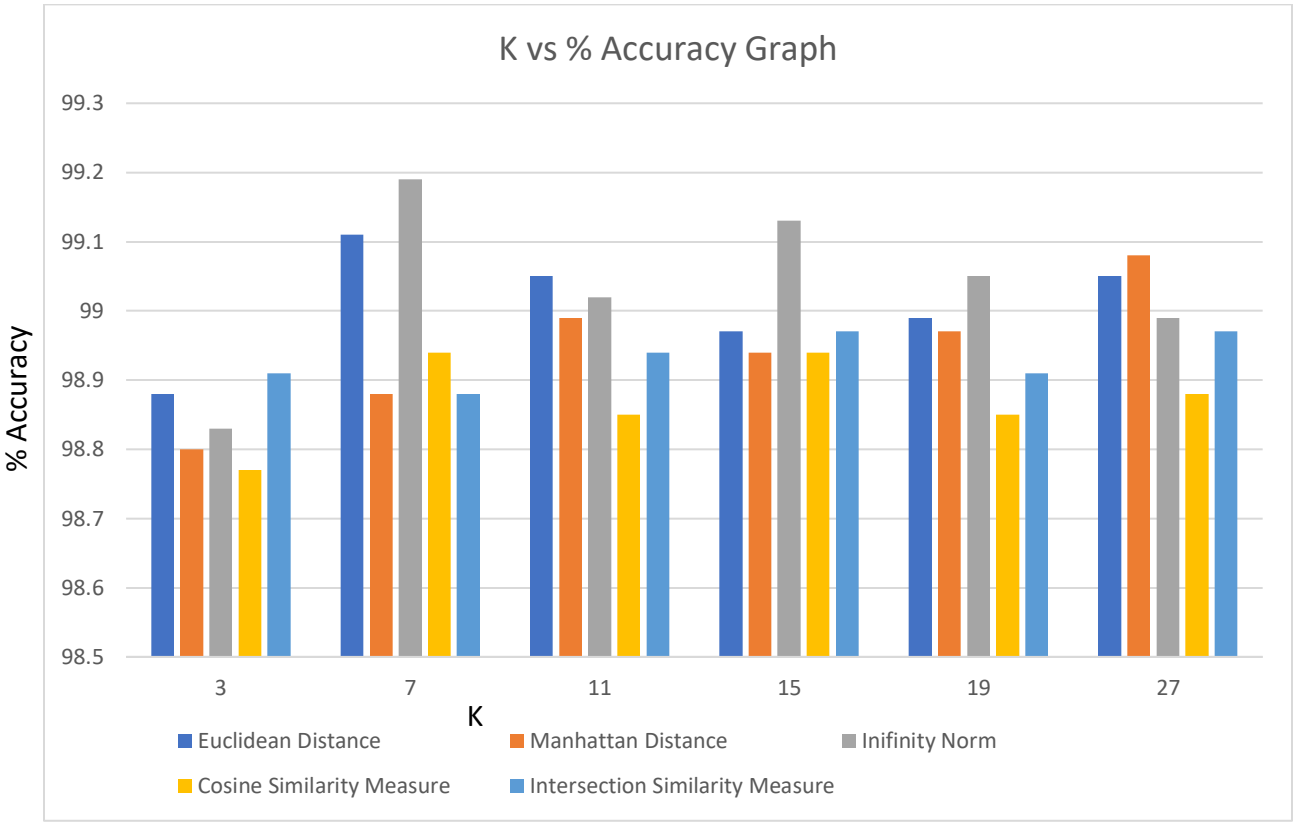

Fig: K versus % Accuracy Graph

K is taken on X-axis and %Accuracy is taken on Y-axis.

From the Graph we can clearly see that, **for K = 7, and by using Infinity Norm Distance Measure**, we get the highest Accuracy.

And for the small value of K, as in case of k = 3, we get lower accuracy and for higher values of k, as well.