# Predicting Maps using the K-Nearest Neighbors Algorithm

Faaiz Ashiq  
L1F22BSCS0579  
Student  
FOIT&CS, University of Central Punjab  
Lahore, Pakistan   
l1f22bscs0579@ucp.edu.pk

## Abstract

This paper investigates the application of the K-Nearest Neighbors (KNN) algorithm for completing partially filled geospatial maps. Using maps of Italy with varying levels of known data (10% to 50%), this study implements and evaluates the performance of KNN in predicting unknown pixels as either 'Land' or 'Water'. The algorithm is tested with both Euclidean and Manhattan distance metrics for K values of 1, 3, 5, 7, and 9. Performance is measured using a confusion matrix to derive accuracy, sensitivity, and specificity, as well as the Jaccard Similarity Index. The results are analyzed to determine the optimal combination of parameters for this image completion task.

## I. INTRODUCTION

The challenge of dealing with incomplete data, specifically in the context of satellite or geographical maps, is significant. The K-Nearest Neighbors (KNN) algorithm is a simple, non-parametric machine learning method suitable for classification tasks. The primary goal of this project is to implement the KNN algorithm to predict the classification of unknown pixels (Land or Water) on a map of Italy. This study systematically evaluates the impact of two distance metrics (Euclidean and Manhattan) and five different values of K on the algorithm's predictive performance across datasets with varying levels of completeness.

## II. DATASET DESCRIPTION

The dataset consists of partial maps of Italy where a certain percentage of pixels (10%, 20%, 30%, 40%, 50%) are pre-filled. The pixels are classified into two categories: 'Land' and 'Water'. A complete map of Italy is provided as the ground truth for training the neighbors and for evaluating the final predictions.

## III. METHODOLOGY & IMPLEMENTATION DETAILS

A. Pixel Representation

Each pixel is treated as a data point, defined by its coordinates (x, y) and its label (Land or Water).

B. K-Nearest Neighbors (KNN) Algorithm

For each unknown pixel, the algorithm calculates the distance to all known pixels. It then identifies the 'K' nearest neighbors and assigns the class of the unknown pixel based on the majority vote among these neighbors.

C. Distance Metrics

Manhattan Distance: D\_M = |x1 - x2| + |y1 - y2|

Euclidean Distance: D\_E = sqrt((x1 - x2)^2 + (y1 - y2)^2)

D. Evaluation Metrics

Performance is evaluated using a confusion matrix, accuracy, sensitivity, specificity, and the Jaccard Similarity Index.

Accuracy = (TP + TN) / (TP + TN + FP + FN)

Sensitivity = TP / (TP + FN)

Specificity = TN / (TN + FP)

J(A, B) = |A ∩ B| / |A ∪ B|

## IV. RESULTS

This section presents the results of the KNN algorithm for varying K values, distance metrics, and fill percentages. Each table corresponds to a specific fill percentage and shows performance metrics.

## V. DISCUSSION

The discussion analyzes the effect of K, the choice of distance metric, and the initial fill percentage. Generally, smaller odd values of K (such as 3 or 5) provide the best results. Euclidean distance tends to outperform Manhattan distance, likely due to its more natural geometric interpretation. Higher initial fill percentages improve prediction accuracy.  
  
While calculating distance I have made a box around every current pixel, difference occurs when I made a box of 5x5 prediction, accuracy is good, despite 3x3.

## VI. CONCLUSION

This study successfully demonstrated the use of the KNN algorithm for a geospatial image completion task. The results indicate that performance is highly dependent on the amount of available data, the chosen K value, and the distance metric. It was found that Euclidean distance with a K value of 5 yielded the highest accuracy, particularly on maps with at least 40% known pixels. This confirms that KNN can be an effective, albeit computationally simple, method for such problems.

## REFERENCES

[1] T. M. Cover and P. E. Hart, 'Nearest neighbor pattern classification,' IEEE Transactions on Information Theory, vol. 13, no. 1, pp. 21-27, Jan. 1967.