

# Representing Paths to Parks using K-Nearest Neighbor

Billena, Artacho, Constantino, Torre

October 3, 2023

## Introduction

**Have you ever wondered how your brain seems to reason out the path that you need to take in order to reach a place?**

Is it by using landmarks as to instruct you as to where you should meet? Is it by using a map to know the nearest location of that landmark?

We are inclined to use the nearest neighboring places in order to reach another place right? Since it is the most "known" location from the place we can meet up. Which is why, I personally give KNN another distinguishing meaning which is "Know the Nearest Neighbor" Classification Algorithm. I think you're thinking that it should be KtNN right? Yes, but hey what's this, there's a popping up animation of how it works. (insert animation for knn - to be animated)

**K-Nearest Neighbor** is a supervised machine learning algorithm that can be used to solve both classification and regression tasks. Classification? Regression? An example would you as a person, if you were to be classified, how would you be classified and related to that group?

**Classification** is simply how does a person or an object belong to a group, it's characteristics such as interests, classes taken, preferences over food, etc. **Regression** on the other hand, is how you are **related** to that particular group based on either your or their influence on you.

## Methodology / Process

### Finding the K-Nearest Neighbors

Thinking back about the process that landmarks are used in order to determine how close we are to a certain place, we need to measure it to create an accurate representation of how far each location is to each other.

## Distance Measurement

### Euclidean Distance

**Euclidean Distance** is an algorithm that we can think of like a triangle ruler that we can use in order to measure the distance between these two points.

$$Euclidean = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

Euclidean Algorithm is commonly found in finding the greatest common divisor between two given numbers which is enclosed by the parenthesis and squared as to why it is squared, well as said earlier it is a triangular ruler where the rules of the **Pythagorean Equation** matters in order to find the hypotenuse of the triangle which is the longest part of triangle which has the formula:

$$x^2 + y^2 = z^2$$

In our case it is used to represent the **displacement** which is the Nearest Shortcut to reach that location. Another point to consider is why do we need to square root it, is there any particular reason why? Square rooting the results will give us two width and height to measure the triangle that we will be using as our ruler to measure the distance which is shown below: (insert triangle diagram)

### Manhattan Algorithm

**Manhattan Distance** measures the distance as the sum of the absolute differences between the coordinates of the points. It is named after the grid-like layout of streets in Manhattan, where the distance between two points is the sum of the absolute differences in their x and y coordinates. (insert grid structures here)

$$Manhattan = |x_2 - x_1| + |y_2 - y_1|$$

### K-Nearest Neighbor Implementation

**Problem:** How can we use this knowledge in order to solve real world problems?

- Solve for the Location of the Parks that are near to each other.
- Solve for the location of the user which is using the parks as the landmark in asking a friend for directions.
- Know where a friend is using the landmarks that he / she is currently seeing.
- Know the path to take when visiting all of the parks and representing the path that is efficient and effective for fares to save money
- A much more endless possibilities for searching for a more applications especially in location searching.

There are a lot of possible implementation in knowing the nearest neighbor or the KNN Classification problems but the problem that we have focused in knowing is **Knowing the path to take when visiting all of the parks and representing the path that is efficient and effective for fares to save money** because we want to make most of our time and money.

## 0.1 Steps in Implementation

### 1. Import the Packages to Represent the Data:

#### (a) Pandas

Used in order to read the data which is a .csv file(comma-seperated values) to the model

#### (b) Networkx

Used to represent the relationship of the data points with the

#### (c) Matplotlib

Used to show the representation made by the Networkx package

## Implementation

## Results

## References

- <https://towardsdatascience.com/k-nearest-neighbors-knn-explained-cbc31849a7e3>
- <https://saturncloud.io/blog/algorithm-for-minimum-manhattan-distance-a-guide-for-data-scientists/>