T. Pavithra
1817137

KNN Algorithm:

K-Nearest Neighbour is one of the simplest ML algorithms based on supervised Learning technique.

KNN algorithm assumes the similarity b/w the new case/data & available cases & put the new case into the category tat is most similar to available categories.

KNN algorithm stores all the available data & classifies a new data point based on similarity. This means when new data appears then it can be easily classified into a well suite category by using KNN algorithm

KNN is a non parametric algorithm, which means it does not make any assumption on underlying data.

It is called a lazy learner algorithm because it does not learn from the training set immediately instead it stores the dataset & at the time of Classification it performs an action on the dataset

KNN algorithm at the training phase just stores the dataset and when it gets new data, then it classifies that data into a category tat is much similar to new data before implementing the KNN algorithm These steps are ~~fotted~~. followed.

→ Read ~~and~~ and explore the train data,

→ Understand the Data.

→ Target variable

→ Read the test data.

→ Understanding the features & the problems.

→ Clean the Data

→ Model the Data using KNN

Read & explore the train Data:-

Load the needed libraries for the dataset. Read the dataset of csv file. Display the dataset with few rows of the data. Get the shape of the data just to know how many rows & coloumns it going to be train.

Understanding the data:

Its important to understand the columns before we move further. Train data has the following columns.

Case number, Arrest, Domestic, Beat, crime updated date. Latitude, Longitude, District, police district, time block Day of the crime (closest station) occured, Location.

Target variable: (primary type)

shows the categories of crimes. There are multiple categorical crimes are displays. It leads to multi class classificatio

Here we classify the crime type crime category is the target variable.

Read the test data:-

Now read the test data, it does not have the target variable and the resolution Replacing. The day of the week with a individual integer value.

```
data-week-dict = {
   "Monday" : 1,
   "Tuesday": 2,
   "Wednesday": 3,
   "Thursday" : 4,
   "Friday" : 5,
   "Saturday": 6,
   "Sunday": 7.
}
```

Replacing the date of the week in training data and testing data as well,

~~I~~ Using similar method we also provide indicater value for police district..(closest station)

Let's try to find some correlation b/w the target variable and numelic variables but before that describe the numeric columns to look for missing values in the data.

Model the data using KNN:-

Lets use KNN algorithm on numaric columns. We take case num, closest station, longtitude, and lattitude for main feature. x-train assign to features, Y-train assign to categories. KNN is fit for (x-train, Y-train) Then it will predict the category type. finally the result is printed.