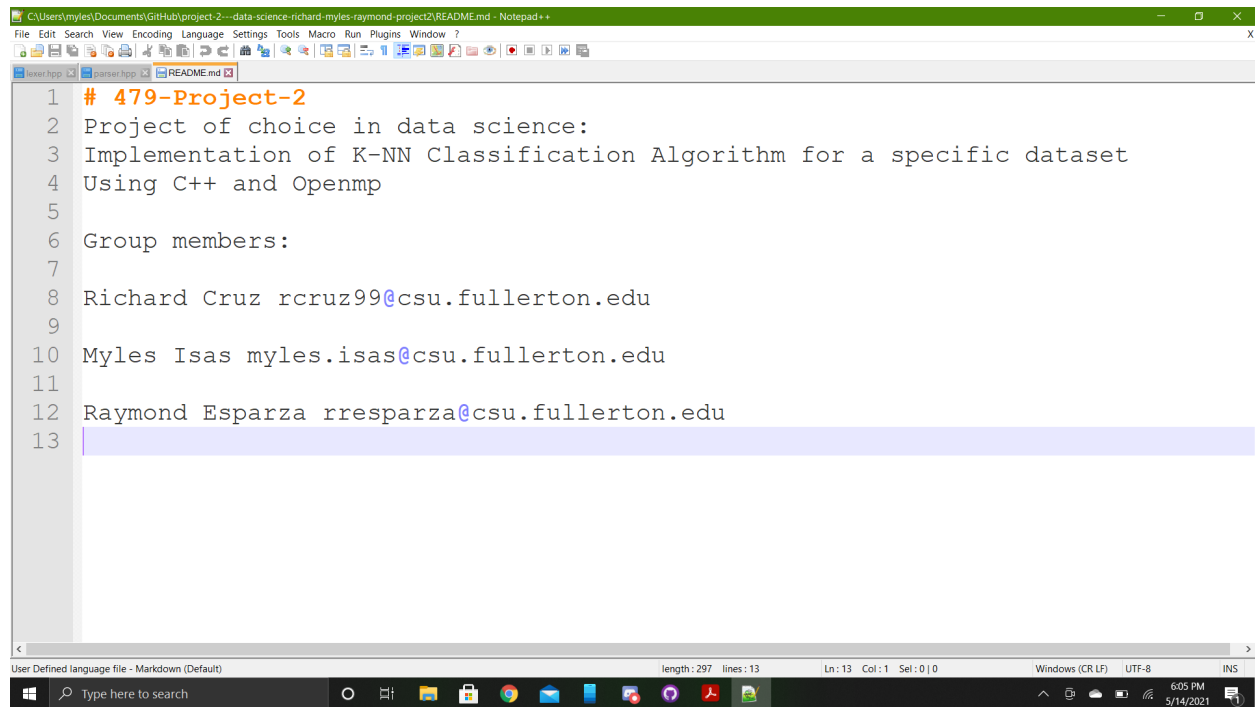


Myles Isas - myles.isas@csu.fullerton.edu  
Richard Cruz - rcruz99@csu.fullerton.edu  
Raymond Esparza - rresparza@csu.fullerton.edu

## **CPSC 479**

### **Project 2 Report:**



```
1 # 479-Project-2
2 Project of choice in data science:
3 Implementation of K-NN Classification Algorithm for a specific dataset
4 Using C++ and Openmp
5
6 Group members:
7
8 Richard Cruz rcruz99@csu.fullerton.edu
9
10 Myles Isas myles.isas@csu.fullerton.edu
11
12 Raymond Esparza rresparza@csu.fullerton.edu
13
```

We chose as our project to implement a K-NN classification algorithm for a specific dataset using C++ and Openmp. We used a dataset on financial information of a list of people. We used the algorithm to determine if our target will default their loan or not by comparing the target's data to the data in the data set. Our data sets tend to favor a majority of people not defaulting. With that said our k value is 5 and it is ideal, too high would show an opposite result.

### **Pseudocode for the Chosen Algorithm:**

euclid\_dist() - shortest distance between two points

main()

    Declare vector<vector> default  
        vector<string> data

    Ifstream ("Default\_Fin.cv")

    while(getline(file, line))

Read in line

```
while(getline(w, cell, ','))
```

Pushback data knows to enter into new column

Create test case

```
vector<double> test_data{}
```

Declare

```
vector<double, double> neighbor;
```

```
tid = omp_get_thread_num();
```

```
#pragma omp parallel private() shared()
```

```
tid = omp_get_thread_num();
```

```
#pragma omp for
```

```
euclid_dist();
```

```
if(tid==0)
```

Neighbor pushback

```
sort (neighbor.begin(), neighbor.end());
```

Output distance and original index

```
if(classification == 0)
```

```
notDefault++
```

Else

```
Defaulted++
```

Print count of notdefault and defaulted

```
if(not defaulted > defaulted)
```

Classify as not defaulted

Else

Classify as defaulted

```
return 0;
```

### **How to compile and run program:**

This program is written in C++. To Run the code, download the file and compile in tuffix using the terminal by typing in g++ knn.cpp -fopenmp

Subsequently type in OMP\_NUM\_THREADS=7 ./a.out to run the program.

## Two snapshots of code executing for some two distinct value of N:

### Snap shot 1:

```
student@tuffix-vm:~/Desktop/High Performance Computing/Project 2$ g++ knn.cpp -fopenmp
student@tuffix-vm:~/Desktop/High Performance Computing/Project 2$ OMP_NUM_THREADS=7 ./a.out
Distance: 548.838 Original index: 5335
Distance: 1698.33 Original index: 9840
Distance: 1815.03 Original index: 6791
Distance: 2166.69 Original index: 7417
Distance: 2253.81 Original index: 7565
Distance: 2682.23 Original index: 9129
Distance: 3160.47 Original index: 9289
Defaulted neighbor count: 0
Not Defaulted neighbor count: 7
The new data is under the classification of 0 or not defaulted
student@tuffix-vm:~/Desktop/High Performance Computing/Project 2$
```

### Snap shot 2:

```
student@tuffix-vm:~/Desktop/High Performance Computing/Project 2$ g++ knn.cpp -fopenmp
student@tuffix-vm:~/Desktop/High Performance Computing/Project 2$ OMP_NUM_THREADS=7 ./a.out
Distance: 548.838 Original index: 5335
Distance: 1698.33 Original index: 9840
Distance: 1815.03 Original index: 6791
Distance: 2166.69 Original index: 7417
Distance: 2253.81 Original index: 7565
Distance: 2682.23 Original index: 9129
Distance: 3160.47 Original index: 9289
Distance: 3217.54 Original index: 3192
Distance: 3326.1 Original index: 614
Distance: 3334.79 Original index: 3889
Distance: 3876.97 Original index: 809
Distance: 3882.57 Original index: 8012
Distance: 4173.21 Original index: 1138
Defaulted neighbor count: 0
Not Defaulted neighbor count: 13
The new data is under the classification of 0 or not defaulted
student@tuffix-vm:~/Desktop/High Performance Computing/Project 2$
```