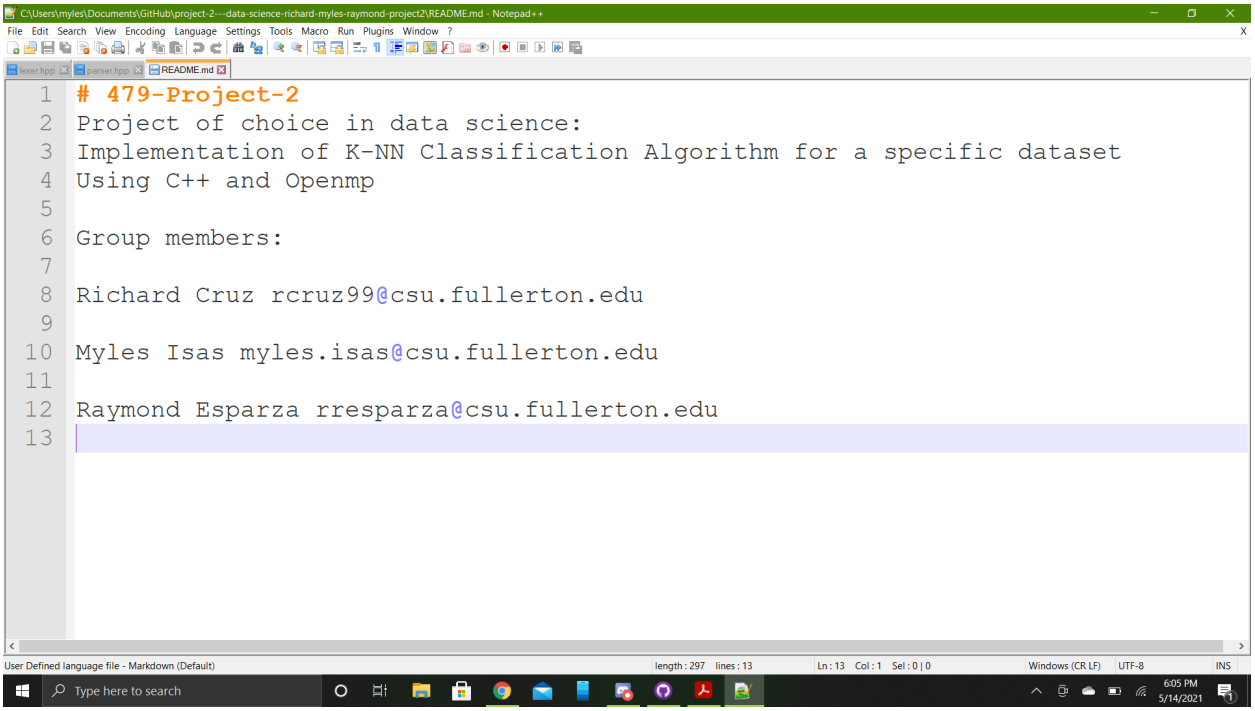
Myles Isas - myles.isas@csu.fullerton.edu
Richard Cruz - rcruz99@csu.fullerton.edu
Raymond Esparza - rresparza@csu.fullerton.edu

CPSC 479
Project 2 Report:



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))

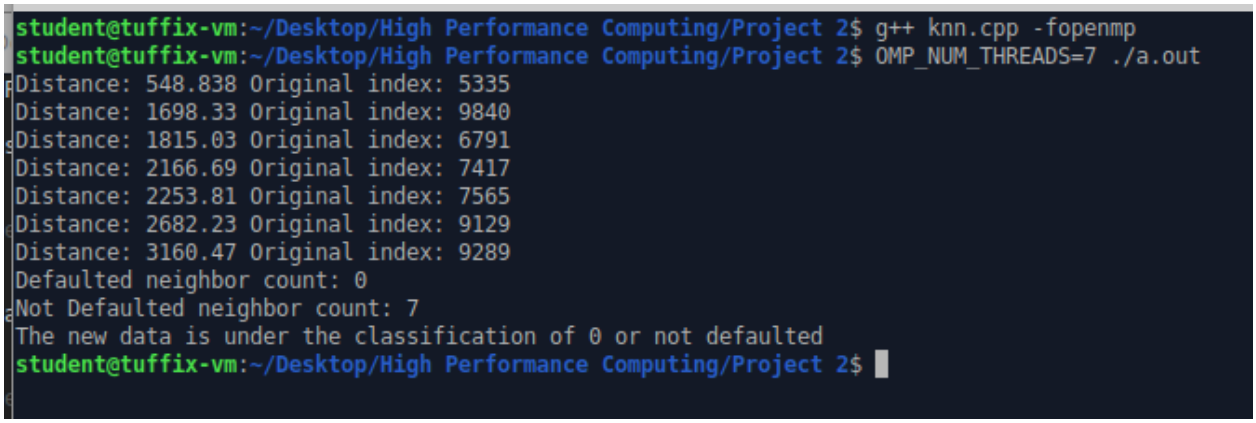
```
            Read in line

        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:



Snap shot 2: