MACHINE LEARNING ASSIGNMENT 1

Name: Jonathan Sankeerth Kumar Pasupulety

ID: 1001772010

**Introduction:**

In [pattern recognition](https://en.wikipedia.org/wiki/Pattern_recognition), the ***k*-nearest neighbors algorithm** (***k*-NN**) is a [non-parametric](https://en.wikipedia.org/wiki/Non-parametric_statistics) method.

The k-Nearest Neighbors algorithm or KNN for short is a very simple technique.

The entire training dataset is stored. When a prediction is required, the k-most similar records to a new record from the training dataset are then located. From these neighbors, a summarized prediction is made.

Similarity between records can be measured many different ways. A problem or data-specific method can be used. Generally, with tabular data, a good starting point is the [Euclidean distance](https://en.wikipedia.org/wiki/Euclidean_distance).

Once the neighbors are discovered, the summary prediction can be made by returning the most common outcome or taking the average. As such, KNN can be used for classification or regression problems.

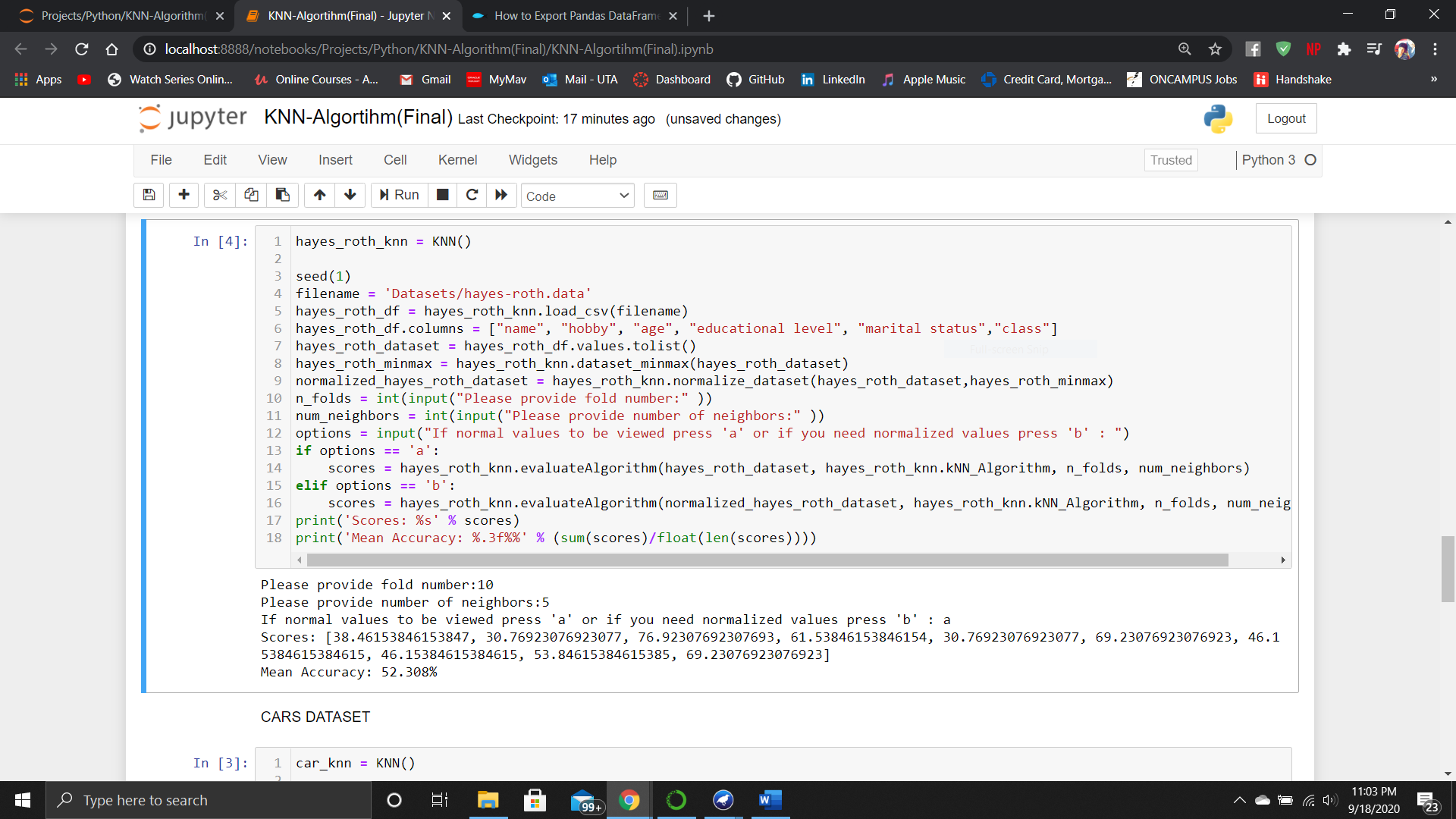
There is no model to speak of other than holding the entire training dataset. Because no work is done until a prediction is required, KNN is often referred to as a lazy learning method.

**Results:**

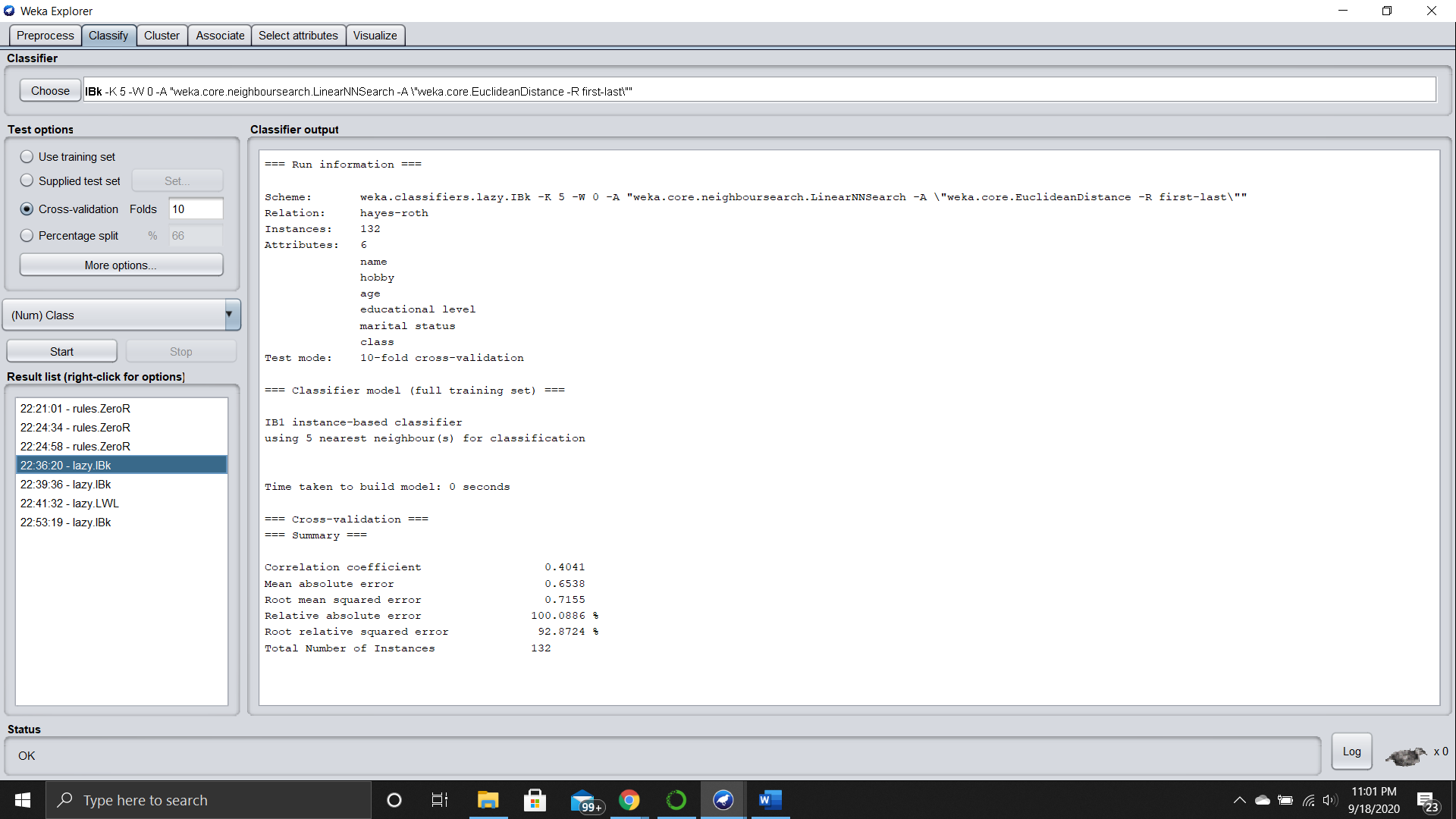
**Hayes-Roth Dataset:**

This is the accuracy achieved by the model created from scratch.

For 10 folds and 5 neighbors we received an accuracy of 52.308%.

****

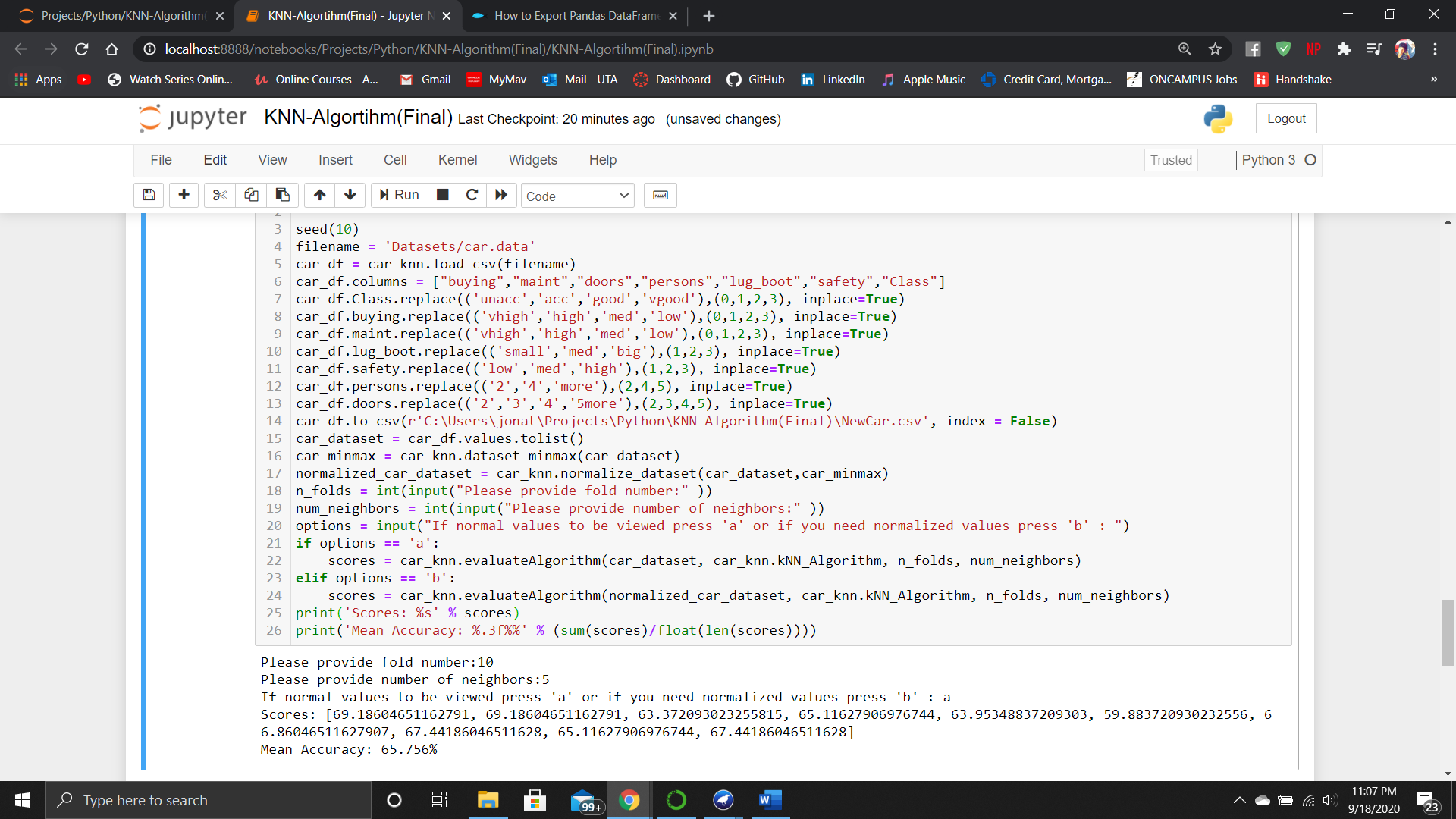
This is the model accuracy got by the model by WEKA.



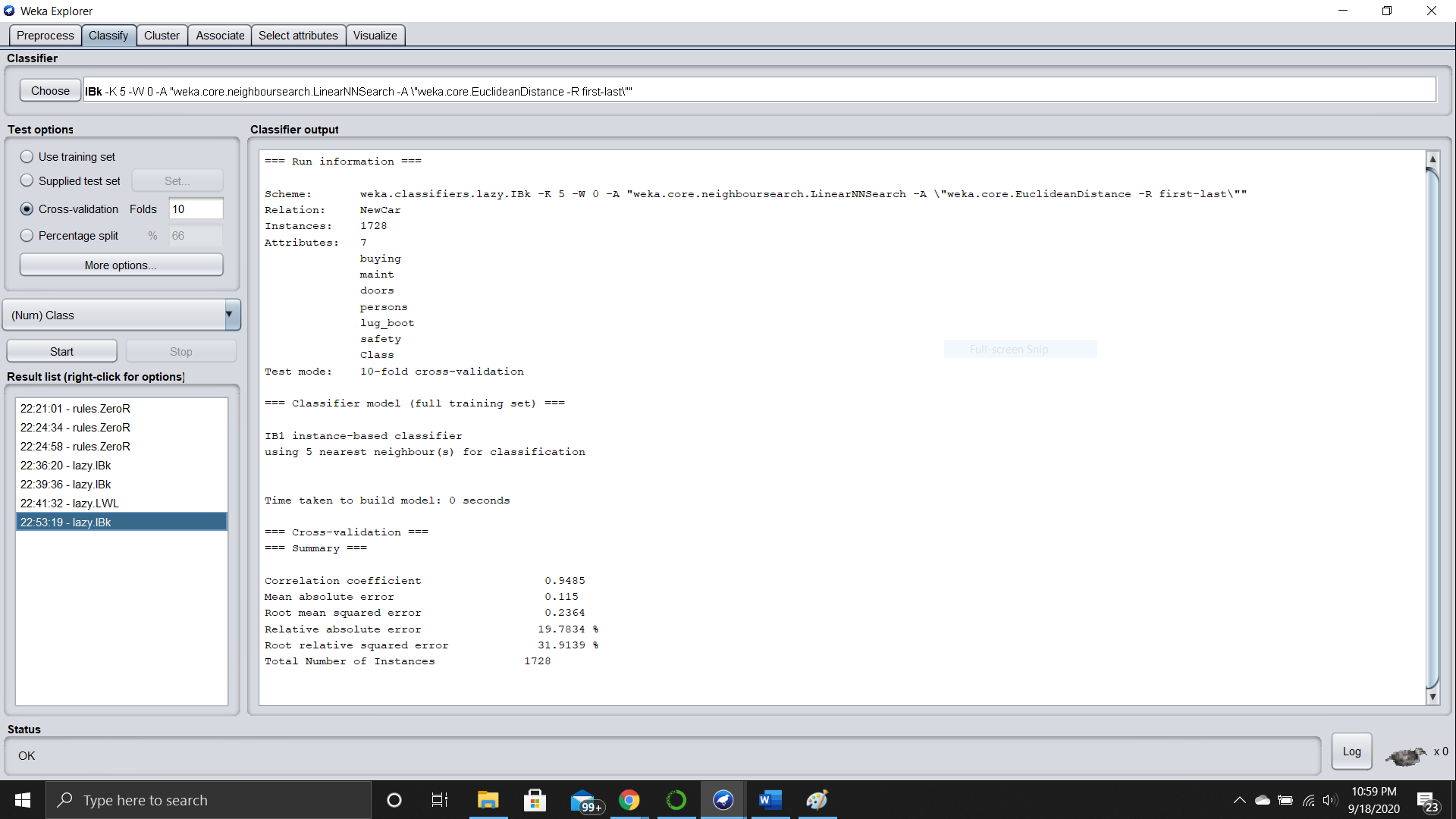
**Car Dataset:**

This is the accuracy achieved by the model created from scratch.

For 10 folds and 5 neighbors we received an accuracy of 65.756 %.

****

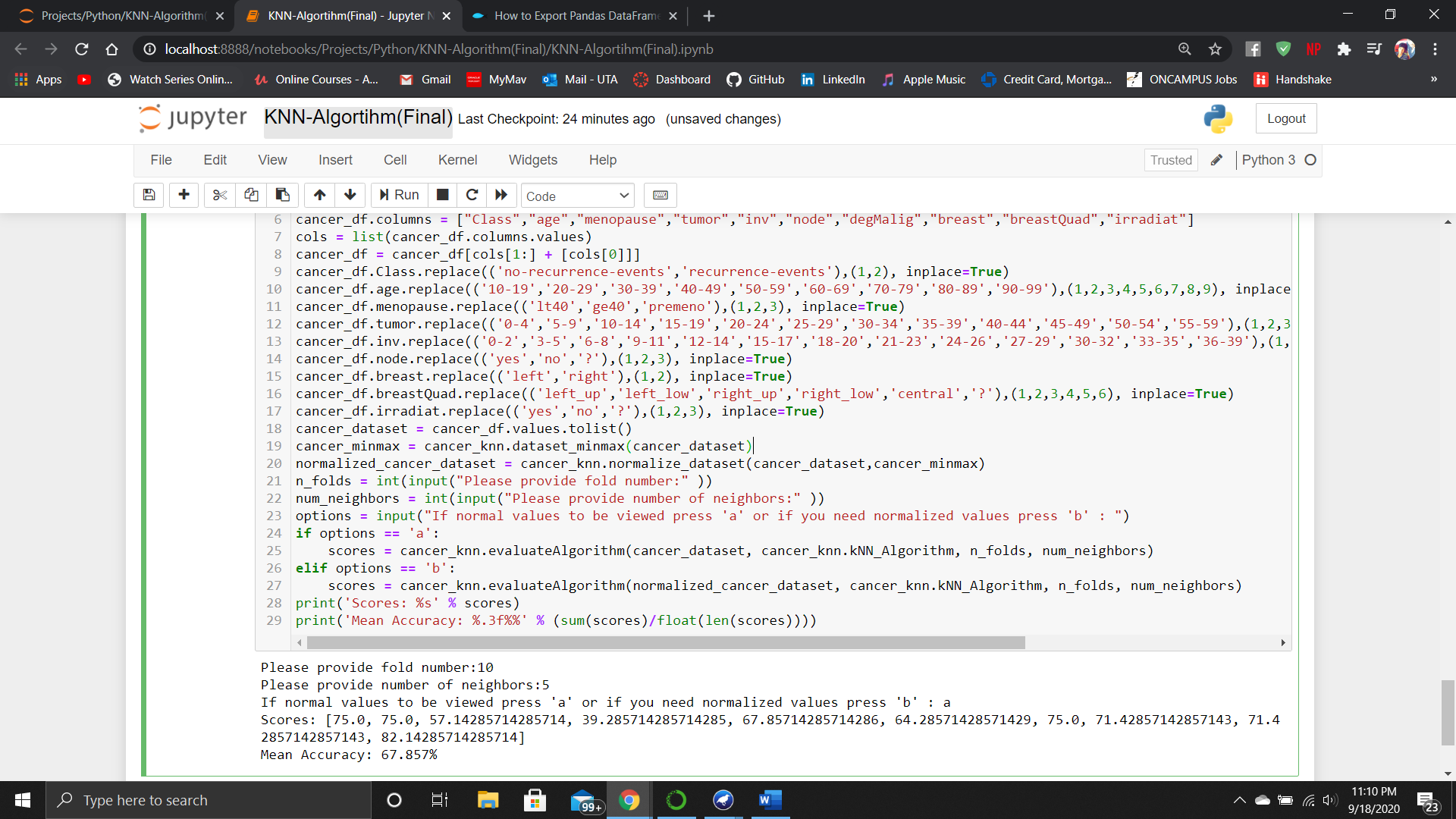
This is the model accuracy got by the model by WEKA

****

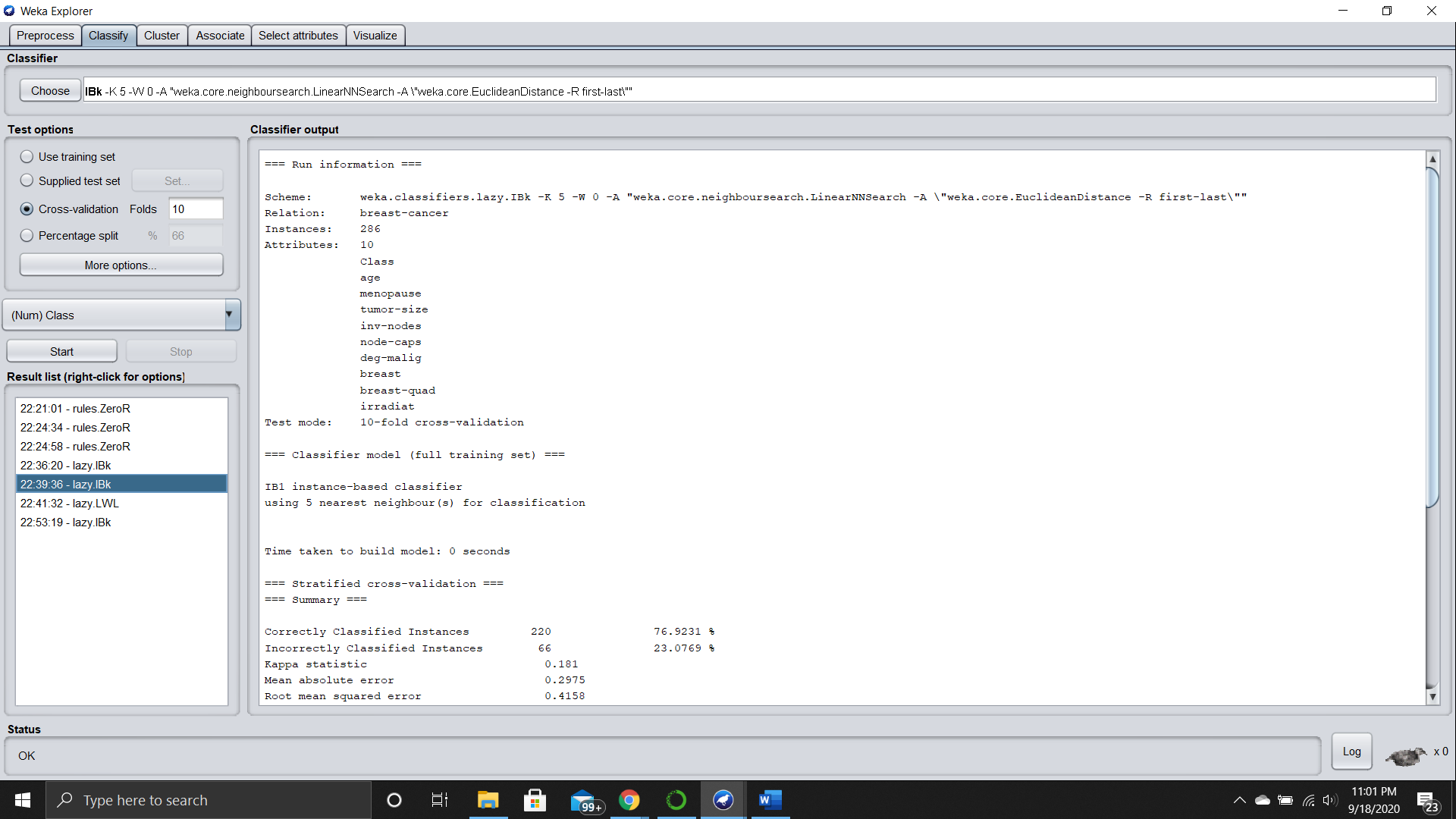
**Breast Cancer Dataset:**

This is the accuracy achieved by the model created from scratch.

For 10 folds and 5 neighbors we received an accuracy of 67.857%.

****

This is the model accuracy got by the model by WEKA

****

**References:**

<https://machinelearningmastery.com/tutorial-to-implement-k-nearest-neighbors-in-python-from-scratch/>

<https://machinelearningmastery.com/k-fold-cross-validation/>

<https://machinelearningmastery.com/implement-resampling-methods-scratch-python/>

<https://archive.ics.uci.edu/ml/datasets/Hayes-Roth>

<https://archive.ics.uci.edu/ml/datasets/Car+Evaluation>

<https://archive.ics.uci.edu/ml/datasets/Breast+Cancer>

<https://www.kaggle.com/elikplim/knn-and-ann-mlp-for-car-evaluation>

<https://techvidvan.com/tutorials/python-lists/>