

CHAPTER 5

IMPLEMENTATION AND TESTING

5.1 Implementation

This project is implemented in PHP language and used laravel framework. In this sub-chapter will explain how to use the program and how the program work. There are two step in this program, which is import, select variable and calculation process.

5.1.1 Import

In this process, the data must be imported to database because it will be used in classification process. The data must be as csv format or excel format.

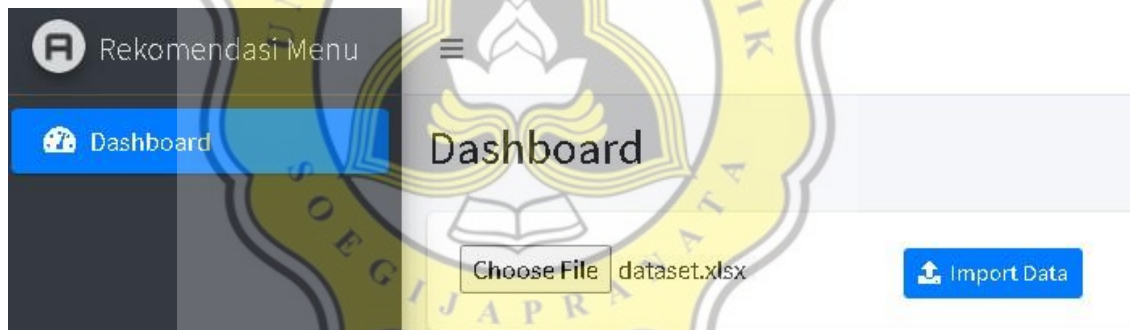


Figure 5.1 : Import Menu

Figure 5.1 The picture above is the very first step to carry out the calculation process. In this step, the data will be saved to the database. The data to be saved must be in csv or excel format.

```
1. $data = [];  
2.  
3.     foreach($collection as $index => $dataset) {  
4.         $data[$index]['gender']           = $dataset[0]  
   ?? null;  
5.         $data[$index]['age']             = $dataset[1]  
   ?? null;
```

```

6.          $data[$index]['fav_drink']           = $dataset[2]
   ?? null;
7.          $data[$index]['fav_drink_flavour'] = $dataset[3]
   ?? null;
8.          $data[$index]['recommend_drink']    = $dataset[6]
   ?? null;
9.          $data[$index]['fav_food']          = $dataset[4]
   ?? null;
10.         $data[$index]['fav_food_flavour'] = $dataset[5]
   ?? null;
11.         $data[$index]['recommend_food']    = $dataset[7]
   ?? null;
12.         $data[$index]['created_at']        = now();
13.         $data[$index]['updated_at']        = now();
14.
15.     }
16.
17.     Dataset::insert($data);
18.     \DB::commit();

```

The illustration and codes above shown there are many steps that needed to do before classification process. First step is the training data will be imported and saved into database. In line 3-14 is declaration process to import value in database. While line 17-18 is query command to do insert data.

5.1.2 Dataset

In this process, the data must be imported to database because it will be used in classification process. The data must be as csv format or excel format. The imported data will be displayed. The data displayed is data that has not undergone the calculation process. The data consists of eight columns

Dataset

Show entries Search:

No	Jenis Kelamin	Usia	Menu Minuman	Rasa Minuman	Merekomendasikan Minuman	Menu Makanan	Rasa Makanan	Merekomendasikan Makanan
1.	Laki - laki	Dewasa (Usia 26 - 40 Tahun)	Classic Coffee	Pait	Tidak	Ricebowl	Gurih	Tidak
2.	Laki - laki	Remaja (Usia 16 - 25 Tahun)	Manual Brewing	Asam	Tidak	Ricebowl	Gurih	Ya
3.	Laki - laki	Remaja (Usia 16 - 25 Tahun)	Manual Brewing	Asam	Ya	Ricebowl	Gurih	Tidak
4.	Perempuan	Remaja (Usia 16 - 25 Tahun)	Classic Coffee	Asam	Tidak	Savoury	Gurih	Tidak
5.	Perempuan	Remaja (Usia 16 - 25 Tahun)	Non Coffee	Manis	Tidak	Pasta	Pedas	Ya
6.	Laki - laki	Remaja (Usia 16 - 25 Tahun)	Squash & Shake	Pait	Ya	Sweet	Pedas	Ya
7.	Laki - laki	Dewasa (Usia 26 - 40 Tahun)	Es kopi susu	Asam	Ya	Ricebowl	Pedas	Tidak
8.	Laki - laki	Dewasa (Usia 26 - 40 Tahun)	Es kopi susu	Asam	Tidak	Ricebowl	Pedas	Ya

Figure 5.2 : Dataset

Figure 5.2 the next step the data will be displayed in the form of a table. The displayed data will then undergo preprocessing to simplify the calculation process. The data will then be processed based on gender, age range, food or beverage variables and certain methods

5.1.3 Select Variable and Methode

In this step, user must chose the variable where is needed to do classification process. The program will automatically do filtering and sorting in this step.

Figure 5.3 : Calculated Menu

Figure 5.3. in the picture above there are 4 options to choose what variables will be calculated and by what method. the process will be displayed on the screen when the data will be calculated. In the calculation process itself, it is required to choose the method and variables to be classified.

```

19. private function _calculateEntropy(int $total, int
    $totalTrue, int $totalFalse) {
20.     $entropy = (((($totalTrue / $total) * -1) *
    (log(($totalTrue / $total), 2))) + (((($totalFalse / $total) * -1)
    * (log(($totalFalse / $total), 2)))));
21.     return is_nan($entropy) ? 0 : $entropy;
22. }
23.
24. private function _calculateGain($dataTotal, $dataEntrophies)
    }

```

The code shown in above is for calculated when user chose Decision Tree C45 this code contain a formula to calculated entropy and gain. Entropy is a process that is needed to determine gain based on predetermined parameters. While the gain is the final result that will be used as a decision node in the process of calculating the c45 algorithm.

```

1.     $resultKnn = ResultKnn::orderBy('distance', 'asc')-
>take($kValue)->get();
2.     foreach($resultKnn as $knn) {
3.         $knn->result_k = $knn->label;
4.         $knn->save();
5.     }
6.
7.     $knn = ResultKnn::orderBy('distance', 'asc')->get();
8.     return [
9.         'data' => $knn,
10.        'result' => (clone $knn)->whereNotNull('result_k')-
>mode('result_k')
11.    ];
12.    }
13.
14.    private function _addLabel($totalTrue, $totalFalse) {
15.        if ($totalTrue > $totalFalse) {
16.            return 'Rekomendasi';
17.        }
18.
19.        return 'Tidak Rekomendasi';
20.    }
21.
22.    private function _sumDistance($total, $totalTrue,
    $totalFalse, $sampleTotal, $sampleTotalTrue, $sampleTotalFalse) {
23.        return sqrt(pow(($total - $sampleTotal), 2) +
    pow(($totalTrue - $sampleTotalTrue), 2) + pow(($totalFalse -
    $sampleTotalFalse), 2));
24.    }
25. }

```

The code shown in above is for calculated when user chose K-Nearest Neighbor this code contain a formula to calculated KNN. In line 1-5 is process to create label by nearest neighbor from k value where is has been set. In line 7-20 is a process to create label from all of nearest neighbor and set the result by frequently label come out. In line

22-25 is a method to calculate distance. The method to calculate distance in this project is using Ecludian distance

5.2 Testing

In Classification Algorithm, there are must do a comparison between result and actual data because each result will give different accuracy data depending on supporting data that affect it. The result of comparison between actual data and result will be calculated using Confusion Matrix Method.

Result C45 :

Menu	Rasa	Label Aktual	Label Hasil Hitung
Chocolate	Manis	Tidak Rekomendasi	Rekomendasi
Classic Coffee	Pait	Tidak Rekomendasi	Tidak Rekomendasi
Non Coffee	Manis	Tidak Rekomendasi	Tidak Rekomendasi
Classic Coffee	Asam	Tidak Rekomendasi	Tidak Rekomendasi
Non Coffee	Asam	Tidak Rekomendasi	Tidak Rekomendasi
Es kopi susu	Asam	Tidak Rekomendasi	Tidak Rekomendasi
Chocolate	Asam	Tidak Rekomendasi	Tidak Rekomendasi
Chocolate	Pait	Tidak Rekomendasi	Tidak Rekomendasi
Manual Brewing	Asam	Tidak Rekomendasi	Tidak Rekomendasi
Squash & Shake	Manis	Rekomendasi	Rekomendasi
Es kopi susu	Pait	Rekomendasi	Tidak Rekomendasi
Tea	Manis	Tidak Rekomendasi	Tidak Rekomendasi

Akurasi : 85.71%

Figure 5.4 : Decision Tree c45 Result

Figure 5.4. the following table shows the results of the calculations using the c45 decision tree algorithm. in the table image there are menu parameters, taste, actual labels, and calculated results labels. while at the bottom left is the result of the calculation of the confusion matrix which is a conclusion from the final result. the calculation results show an accuracy of 85.71%

Data KNN :

Menu	Rasa	Label Aktual	K = 1	K = 5
Chocolate	Manis	Tidak Rekomendasi	Tidak Rekomendasi	Tidak Rekomendasi
Classic Coffee	Pait	Tidak Rekomendasi	Tidak Rekomendasi	Tidak Rekomendasi
Non Coffee	Manis	Tidak Rekomendasi	Tidak Rekomendasi	Tidak Rekomendasi
Classic Coffee	Asam	Tidak Rekomendasi	Tidak Rekomendasi	Tidak Rekomendasi
Non Coffee	Asam	Tidak Rekomendasi	Tidak Rekomendasi	Tidak Rekomendasi
Es kopi susu	Asam	Tidak Rekomendasi	Tidak Rekomendasi	Tidak Rekomendasi
Chocolate	Asam	Tidak Rekomendasi	Tidak Rekomendasi	Tidak Rekomendasi
Chocolate	Pait	Tidak Rekomendasi	Tidak Rekomendasi	Tidak Rekomendasi
Manual Brewing	Asam	Tidak Rekomendasi	Tidak Rekomendasi	Tidak Rekomendasi
Squash & Shake	Manis	Rekomendasi	Rekomendasi	Tidak Rekomendasi
Es kopi susu	Pait	Rekomendasi	Rekomendasi	Tidak Rekomendasi
Tea	Manis	Tidak Rekomendasi	Tidak Rekomendasi	Tidak Rekomendasi

Akurasi:
 K=1:100.00%
 K=5:78.00%

Figure 5.5 : K-Nearest Neighbor Result

Figure 5.5 the following table shows the results of the calculations using the KNN algorithm. In the table image there are menu parameters, taste, actual labels, and calculated results labels. While at the bottom left is the result of the calculation of the confusion matrix which is a conclusion from the final result. The calculation results show an accuracy of 100% if using one k parameter and 78% when using 5 K parameter.

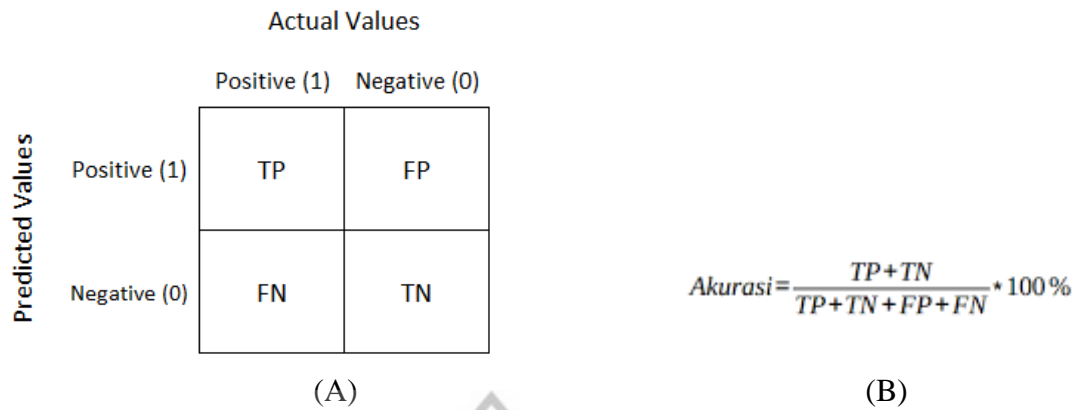


Figure 5.6 : K-Nearest Neighbor Result

Value of accuracy in **Figure 5.4** and **Figure 5.5** is a result from confusion matrix. The formula of confusion matrix is show in **Figure 5.6** from this process can be compared where is the best algorithm between decision tree c45 and K-Nearest Neighbor

