Exercise 2 Task 2

Table 1 accuracies of power consumption classification using random forest classifier.

House	Accuracy (%)	Confusion matrix
1	92	$\begin{bmatrix} 109 & 5 & 1 \\ 1 & 22 & 0 \\ 4 & 1 & 1 \end{bmatrix}$
2	100	$\begin{bmatrix} 97 & 0 \\ 0 & 47 \end{bmatrix}$
3	81	$\begin{bmatrix} 67 & 7 & 3 \\ 2 & 32 & 7 \\ 0 & 9 & 17 \end{bmatrix}$

The table illustrates the accuracies of power consumption classification by using random forest classifier, which was set all parameter as default, with the confusion matrix of each classification.

From the table above, it showed that Random Forest classifier was one of classifiers, which was suitable for dealing with this problem. The results showed that there were 92%, 100%, and 81% of accuracies for 1st, 2nd, and 3rd household data, respectively. Even though, the first household data had an accuracy at 92%, its data could be considered as imbalanced data. As you see in confusion matrix, most data were placed in one group. So, this might be the reason why the accuracy of first household data was pretty high. The second household data, since its data were divided quite obviously, there was an accuracy at 100%. There were 97 values and 47 values for True Positive and True Negative, respectively.

However, there was decreased accuracy for the third household data since its data were quite continuous and distributed. The confusion matrix showed that there was a problem with 2nd and 3rd classification, which had been predicted incorrectly for a half.

In conclusion, result of power consumption classification by using Random Forest classifier had been suitable for dealing with this problem. The highest of classification was at 100% of accuracy for clear-separated data like 2nd household data and the lowest at 81% of accuracy for continuous data like 3rd household.

Chayun Kongtongvattana