5. CLASSIFICATION OF ELECTRICITY CUSTOMER BEHAVIOUR

5.1 Introduction

In chapter 3.3, several data mining classification techniques have been reviewed, including Decision Tree, Naïve Bayes, Back Propagation, Extreme Learning Machine, Online Sequential Extreme Learning Machine, and Support Vector Machine. Among these existing classification techniques, Extreme Learning Machine [89], Online Sequential-Extreme Learning Machine [90] and Support Vector Machine [91] have been selected as appropriate for predicting customer behaviour classes in the present research.

The Extreme Learning Machine learning algorithm and its variant, Online Sequential-Extreme Learning Machine, were chosen as a consequence of the claim that these two classifiers have superior generalization performance and extremely fast learning speeds when compared to other traditional neural network algorithms, including Back Propagation. Support vector machine was selected because it employs the structural risk minimization principle and has proven to have a high generalization capacity for unseen data in many classification problems, such as those associated with electricity load consumption. Some preliminary work using Decision Tree and Naïve Bayes classifiers has been published. In this work, a comparison between these two algorithms was undertaken and the results indicated that Decision Tree performed faster and more accurately as a means of classification relative to Naïve Bayes [77].

Based on the identification procedure that has been described in chapter 4, each daily load pattern is assigned to a pre-labelled class according to whether it manifests normal behaviour, abnormal behaviour, or suspicious behaviour. Here, this classification process is tested using the three classification algorithms ELM, OS-ELM, and SVM in order to identify the best performing and the most accurate classifier. The performance measures used in this analysis are classification accuracy and time processing duration. The model developed from the chosen algorithm will be used to predict categorical class values for future data. Two sets of data for the experiments have been derived from the original customer database, including training data and test data. Each of the training dataset and testing dataset comprise data for weekdays, Saturdays, Sundays, and public holidays.

5.2 Task Definition

The classification task in the present research aims to compare the accuracy and processing speed of all the established methods with ELM. The ultimate intention is to choose the most efficacious method as our model for predicting electricity customer behaviour status in the proposed analysis framework that has been described in chapter 4.

- ❖ Input: Inputs for classification techniques consist of the following items.
  - ➢ A set of individual electricity customer load patterns with a population of L customers, where each consumer load profile is categorised by a vector

$X^l = \{X_h{}^l, h = 1, \ldots, H\}$ whose H components of time-domain data correspond to one-day interval data. The customer data has been separated into two datasets: training datasets selected from 2013 and testing datasets selected from 2014.

- Both training and testing dataset inputs comprising summer, winter and monsoon datasets are normalised into the range [0, 1] for all learning algorithms.
- Numbers of hidden neurons for Extreme Learning Machine and OS- Extreme Learning Machine are tested from 20 to 200, while cost parameter values are tested from 5 to 50.
- The performance of the ELM, OS ELM, and SVM learning algorithms are compared between two activation functions: a simple sigmoid activation function $g(x) = \frac{1}{1+e^{-x}}$ and a radial basis function Gaussian kernel

$$Ø(y) = \exp\left(-\frac{||x-u||^2}{\sigma^2}\right)$$

- Output: The following outputs are expected from these classification learning algorithms.
  - Classification accuracy rates as percentages for training and testing datasets summer datasets, winter datasets and monsoon datasets.
  - Time processing speeds in seconds for training and testing datasets for summer datasets, winter datasets and monsoon datasets.
- Hardware and Software: All the classification simulations for ELM, OS-ELM, and SVM are carried out in the MATLAB 2014a environment running on 3rd generation Intel core i5, 2.5 GHz CPU with 4 GB of memory. The software for ELM, OS-ELM and SVM were tested using MATLAB R2014a.
- Algorithms: Three learning algorithms were selected for this study as set out below.
  - Extreme Learning Machine [9].
  - Online Sequential ELM [10].
  - Support Vector Machine [13].
- Procedure: The following steps make up the classification process.
  - Train the customer data – a set of individual customers' data is used, with the datasets separated according to types of days.
  - Apply the test data – a different set of customers' data separated by types of days is supplied for the testing procedure.
  - Compare the classification accuracies measured as percentages and the time processing durations measured in seconds.

In this chapter, the classification performances of four learning algorithms are assessed, including ELM, OS-ELM, and SVM. The evaluations of each algorithm are compared among the numbers of hidden neurons and cost values used. The best performance is chosen and compared among types of days. The average performances are analysed among Extreme Learning Machine sigmoid, Extreme Learning Machine RBF, OS- Extreme Learning Machine sigmoid, OS- Extreme Learning Machine RBF, SVM sigmoid, and SVM RBF. The results obtained are used for the next prediction task, that of predicting normal behaviour, abnormal behaviour, and suspicious behaviour profiles.

5.3 Customer Behaviour Classification

A comprehensive analysis of four datasets separated according to types of days, including summer datasets, winter datasets and monsoon datasets, was conducted using the selected algorithms, ELM, OS-ELM, and LIBSVM. For each algorithm, the sigmoid function and radial basis function nodes activation functions were compared.

5.3.1 Extreme Learning Machine

ELM [89, 92-96] was developed by Huang and is one among the supervised batch learning algorithms in that it uses a finite number of input and output samples for training. For N, the arbitrary distinct samples are $(x_{i,}t_i)$ Є R^n x R^m, Here x_i is a 1×n is an input vector and t_i is a 1×m target vector. The output of an SLFN with Ñ hidden neurons additive nodes can be represented by

$f_{Ñ}(x) = \sum_{i=1}^{Ñ} \beta_i \ G(a_i, b_i, x) \quad x \in R^n \quad a_i \in R^n$ ……………………….5.1

Where a_i and b_i are the learning parameters of hidden nodes and β_i is the weight connecting the ith hidden node to the output node. G (a_i, b_i, x) is the output of the ith hidden node with respect to the input x. In this section, two activation functions are used, the additive hidden node and the RBF hidden node.

For the additive hidden node with activation function g(x):R→R (e.g., sigmoid), G (a_i, b_i, x) is given by

$G(a_i, b_i, x) = g(a_i.x + b_i) \quad b_i \in R^n$ ………………………..5.2

Where a_i is the weight vector connecting the input layer to the ith hidden node and b_i is the bias of the i^{th} hidden node. $a_i.x$ denotes the inner product of vectors a_i and x in R^n

For RBF hidden node with activation function g(x):R→R (e.g., Gaussian), G (a_i, b_i, x) is given by

$G(a_i, b_i, x) = g(b_i||x - a_i||) \quad b_i \in R+$ ………………………..5.3

Where a_i and b_i are the centre and impact factors of i the RBF node. R+ indicates the set of all positive real values. The radial basis function network is a special case of SLFN with radial basis function nodes in its hidden layer. Each radial basis function node has its own centroid and impact factor, with its output given by a radically symmetric function of the distance between the input and the centre.

- ❖ Function Approximation of SLFNs with Additive Neurons - ELM Sigmoid

  Comprehensive testing with the customers' data separated by types of days (summer datasets, winter datasets and monsoon datasets) using the ELM Sigmoid algorithm with different numbers of hidden neurons from 20 to 200 was conducted. The Extreme Learning Machine sigmoid results are separated into two tables. Table 5-1 presents the time processing durations in seconds. Table 5-2 presents classification success rates as percentages. The best classifier for each type of days is chosen on the basis of the highest classification accuracy ascertained from the simulations.

  Table 5-1 shows the results for time processing durations in seconds applying the Extreme Learning Machine with the sigmoid function. It can be seen that datasets with smaller hidden numbers of neurons produced faster time processing speeds compared to datasets with larger hidden numbers of neurons. The larger the numbers of hidden neurons, the longer the time processing duration in seconds. Overall, the number of hidden neurons equal to 20 is the most suitable, with the fastest results being for winter = 0.023 seconds, monsoon = 0.067 seconds, and summer = 0.135 seconds. However, the number of hidden neurons equal to 20 is found to be faster on winter = 0.023 seconds.

Table 5-2 shows the results for classification accuracies as percentages applying the Extreme Learning Machine with the sigmoid function. It is apparent that datasets with larger hidden numbers of neurons produced higher classification accuracies except for Sunday datasets. Overall, the number of hidden neurons are different between different types of days. The best number of hidden neurons for winter is found to be 160 with 92.66% accuracy, while for monsoon, it is 20 with 94.21% accuracy and summer 60 with 90.13% accuracy.

Table 5-1 shows the time processing durations in seconds. It can be seen that summer training datasets took longer times than did testing datasets to complete the classification task due to the larger datasets compared to other types of days. The other datasets, including winter and monsoon have similar time processing speeds and improved significantly in testing data. Table 5-2 shows the for-classification accuracies as percentages. It can be seen that as the number of hidden neurons increases, the classification success rate increases accordingly. The classification accuracy results for testing datasets are slightly lower than for training datasets for all types of days, although the patterns remain the same consistently.

Table 5-1: Results for time processing duration in seconds with ELM

No. of hidden neurons	Summer		Winter		Monsoon	
	Training (secs)	Testing (secs)	Training (secs)	Testing (secs)	Training (secs)	Testing (secs)
20	0.135	0.131	0.023	0.023	0.067	0.062
40	0.265	0.182	0.143	0.112	0.124	0.112
60	0.342	0.210	0.198	0.209	0.187	0.132
80	0.587	0.226	0.388	0.237	0.190	0.140
100	0.732	0.253	0.390	0.239	0.213	0.195
120	0.987	0.268	0.679	0.289	0.233	0.200
140	1.232	0.292	0.937	0.308	0.329	0.265
160	1.356	0.323	1.072	0.312	0.672	0.542
180	1.490	0.345	1.328	0.321	0.882	0.589
200	1.643	0.376	1.398	0.336	0.999	0.623

Table 5-2: Results for classification accuracy as percentages with ELM

No. of hidden neurons	Summer		Winter		Monsoon	
	Training (%)	Testing (%)	Training (%)	Testing (%)	Training (%)	Testing (%)
20	91.76	89.65	88.87	88.37	94.78	94.21
40	92.24	90.05	89.65	89.08	95.00	94.56
60	92.37	90.13	91.76	90.79	95.21	95.08
80	92.60	90.26	92.00	91.04	95.88	95.65
100	92.81	90.98	92.50	91.33	96.46	95.84
120	92.92	91.08	92.65	91.40	97.76	96.43
140	93.01	91.20	93.28	91.52	98.12	96.71

160	93.25	91.33	94.09	92.66	98.87	97.22
180	93.27	91.52	94.31	92.77	98.91	97.50
200	93.49	91.64	94.97	92.82	98.93	97.89

5.3.2 OS – ELM

A sequential learning algorithm referred to as online sequential extreme learning machine (Online Sequential-Extreme Learning Machine) [10, 90, 97] that can handle both additive neurons and RBF nodes. It was developed to minimise the limitation of Extreme Learning Machine as proposed by Huang. As the Extreme Learning Machine algorithm belongs among the group learning algorithms, this limited its more application. In a practice, the training data may arrive one-by-one and the online sequential learning is required to provide for such variety. In addition, some online industrial applications prefer sequential learning algorithms as they do not need to retrain whenever a new dataset appears. So far, two types of Online Sequential-Extreme Learning Machine have been proposed: i) Online Sequential-Extreme Learning Machine based on recursive least squares (RLS), and ii) improved Online Sequential -Extreme Learning Machine known as (OLS) for Online Sequential-Extreme Learning Machine (RLS).

In this section, two activation functions are applied, namely additive nodes and RBF nodes. In OSELM with additive nodes, the input weights and biases are randomly generated and on this basis, the output weights are analytically determined. Similarly, in OS- Extreme Learning Machine with RBF nodes, the centres and widths of the nodes are randomly generated and fixed and then based on this, the output weights are analytically determined. Unlike any other sequential learning algorithms that require so many more characteristics to be tuned, Online Sequential-Extreme Learning Machine only specified the number of hidden nodes to be required. For the initialization, the number of parameter required should be at least equal to the number of hidden nodes. Following the initialization phase, the learning phase commences either on a one-by-one basis as desired. Once a dataset is used, it is discarded as of no further use.

❖ OS-ELM SIGMOID

Another comprehensive analysis applying the OS-Extreme Learning Machine with the sigmoid function was conducted using the same datasets as used in the Extreme Learning Machine with different numbers of hidden neurons from 20 to 200. The OS-Extreme Learning Machine with sigmoid function results are separated into two tables. Table 5-3 presents the time processing durations in seconds. Table 5-4 presents classification success rates as percentages. The best classifier for each type of days is chosen on the basis of the highest classification accuracy ascertained from the simulations.

Table 5-3 shows the results for time processing durations in seconds applying the OS-Extreme Learning Machine with the sigmoid function. It can be seen that datasets with smaller hidden numbers of neurons produced faster time processing speeds compared to datasets with larger hidden numbers of neurons. The larger the number of hidden neurons, the longer the time processing duration in seconds. Overall, minimum 20 hidden neurons is the most suitable and gave the fastest results for all types of days, with summer = 0.044 seconds, winter = 0.095 seconds and monsoon = 0.234 seconds.

Table 5-4 shows the results for classification accuracies as percentages applying the OS-Extreme Learning Machine with the sigmoid function. It can be seen that the summer and monsoon datasets have higher classification accuracies with the number of hidden neurons equal to 20. However, winter have higher classification accuracies with the number of hidden neurons equal to 200. The best number of hidden neurons for summer is found to be 20 with 94.73% accuracy, for winter it is 200 with 95.98% accuracy, and for monsoon it is 20 with 94.70% accuracy.

Table 5-3 shows the time processing durations in seconds. It can be seen that weekday training datasets took longer times than testing datasets to complete the classification task due to the larger datasets compared to other types of days. The other datasets, including summer, winter and monsoon have similar time processing speeds and improved significantly in testing data. Table 5-4 shows the classification accuracies as percentages. It can be seen that as the classification success rate fluctuates as the number of hidden neurons increases. The classification accuracy results for testing datasets are slightly lower than for training datasets for all types of days.

Table 5-3: Results for time processing durations in seconds with OS-ELM

No. of hidden neurons	Summer		Winter		Monsoon	
	Training (secs)	Testing (secs)	Training (secs)	Testing (secs)	Training (secs)	Testing (secs)
20	0.045	0.044	0.095	0.095	1.343	0.234
40	0.156	0.097	0.145	0.132	1.832	0.759
60	0.271	0.198	0.158	0.141	2.573	1.234
80	0.297	0.203	0.237	0.178	3.767	1.553
100	0.311	0.257	0.373	0.261	5.986	2.435
120	0.514	0.312	0.463	0.367	10.763	2.980
140	0.672	0.432	0.589	0.422	12.275	3.722
160	0.892	0.478	0.655	0.494	13.684	3.904
180	1.070	0.622	0.687	0.573	24.782	4.742
200	1.179	0.800	0.865	0.733	27.970	5.023

Table 5-4: Results for classification accuracy as percentages with OS-ELM

No. of hidden neurons	Summer		Winter		Monsoon	
	Training (%)	Testing (%)	Training (%)	Testing (%)	Training (%)	Testing (%)
20	94.74	94.73	92.22	92.33	94.71	94.70
40	94.22	94.06	92.64	92.34	95.02	94.50
60	94.34	94.14	93.78	92.87	95.23	95.05
80	95.69	94.27	94.01	92.99	95.50	95.21
100	95.88	94.96	94.40	93.08	95.78	95.53
120	95.93	95.05	94.69	93.27	96.71	95.78
140	96.08	95.05	95.09	93.44	97.13	96.03
160	96.24	95.36	95.34	93.56	97.69	96.21
180	97.26	95.58	95.37	93.70	97.80	96.59
200	97.40	95.66	95.98	94.05	97.92	96.74

5.3.3 SVM

A comprehensive analysis applying support vector machine [13,47,98,99] with the sigmoid function was conducted using the same datasets as in Extreme Learning Machine and OS-Extreme Learning Machine with different cost parameter values from 5 to 50. The support vector machine with sigmoid results are separated into two tables. Table 5-5 presents the time processing durations in seconds. Table 5-6 presents classification success rates as percentages. The best classifier for each type of days is chosen on the basis of the highest classification accuracy ascertained from the simulations.

Table 5-5 shows the results for time processing durations in seconds applying SVM with the sigmoid function. It can be seen that datasets with different cost parameters yield different time processing speeds. The cost parameters suitable for each type of days are different, including summer with 40 and 7.58 seconds, winter with 20 and 10.61 seconds and monsoon with 30 and 1.78 seconds.

Table 5-6 shows the results for classification accuracies as percentages applying SVM with the sigmoid function. It can be seen that different types of days have different numbers of cost parameters for the highest classification accuracy. The best cost parameter for summer is found to be 50 with 96.8119% accuracy, for winter it is 50 with 96.25% accuracy and for monsoon it is 5 with 94.6069% accuracy.

Table 5-5 shows the time processing durations in seconds. It can be seen that weekday training datasets took longer (and fluctuating) times than did testing datasets to complete the classification task due to the large datasets compared to other types of days. The other datasets, including summer, winter and monsoon, have similar time processing speeds, however the time processing results are slightly slower in the case of the testing data compared to the training data. Table 5-6 shows the classification accuracies as percentages. It can be seen that the weekday dataset classification accuracy for testing data is higher than that for training data. The classification accuracy results for testing datasets on other types of days are slightly lower than for training datasets for all types of days.

Table 5-5: Results for time processing durations in seconds using SVM sigmoid

C	Summer		Winter		Monsoon	
Cost Parameters	Training (secs)	Testing (secs)	Training (secs)	Testing (secs)	Training (secs)	Testing (secs)
5	2.780	2.694	0.023	0.023	0.067	0.062
10	2.561	2.637	0.143	0.112	0.124	0.112
15	2.335	2.452	0.198	0.209	0.187	0.132
20	2.870	2.730	0.388	0.237	0.190	0.140
25	2.347	2.237	0.390	0.239	0.213	0.195
30	2.114	2.087	0.679	0.289	0.233	0.200
35	1.006	1.006	0.937	0.308	0.329	0.265
40	1.239	1.213	1.072	0.312	0.672	0.542
45	1.652	1.493	1.328	0.321	0.882	0.589
50	1.679	1.582	1.398	0.336	0.999	0.623

Table 5-6: Results for classification success rate as percentages using SVM sigmoid

C Cost Parameters	Summer		Winter		Monsoon	
	Training (%)	Testing (%)	Training (%)	Testing (%)	Training (%)	Testing (%)
5	91.76	89.65	88.87	88.37	94.78	94.21
10	92.24	90.05	89.65	89.08	95.00	94.56
15	92.37	90.13	91.76	90.79	95.21	95.08
20	92.60	90.26	92.00	91.04	95.88	95.65
25	92.81	90.98	92.50	91.33	96.46	95.84
30	92.92	91.08	92.65	91.40	97.76	96.43
35	93.01	91.20	93.28	91.52	98.12	96.71
40	93.25	91.33	94.09	92.66	98.87	97.22
45	93.27	91.52	94.31	92.77	98.91	97.50
50	93.49	91.64	94.97	92.82	98.93	97.89

5.4 Conclusion

On an average, to complete the classification required time duration, it is proved that the smaller the numbers of hidden neurons required for the Extreme Learning Machine and Online Sequential-Extreme Learning Machine algorithms, it will speed up the classification process. Also, from the above observed that the algorithm the support vector machine has the largest time duration required when compared to Extreme Learning Machine or Online Sequential-Extreme Learning Machine. However, for the the classification algorithms, it has been proved that Online Sequential-Extreme Learning Machine with the radial basis function nodes generate the largest classification accuracy, while Extreme Learning Machine with the radial basis function nodes generate the smallest classification accuracy. On the other hand, it has also been proved that the time duration required, support vector machine with radial basis nodes has the slowest speed while Extreme Learning Machine with the sigmoid function has the fastest speed.