6. CUSTOMERS' DATA PREDICTION ANALYSIS

6.1 Introduction

The prediction process to be undertaken involves predicting next-day or next-hour load consumptions based on historical data. The data used comprise the identified unmetered power losses profiles flowing from detection and classification procedures. In the present research, Extreme Machine Learning and Online Sequential-Extreme Learning Machine were applied as the means of predicting the unmetered power losses data from the accumulated historical data. The MGVCL consumer datasets were empirically tested and compared as between the selected techniques. Subsequently, the forecasted results were compared with the actual results and any significant deviations were able to be accurately reported as abnormalities in a timely manner. A review of several prediction techniques that have commonly been used for load-forecasting tasks in electric power utilities. Those techniques included statistical analysis, time-series analysis, and artificial intelligence based analysis. However, for the present purpose, Extreme Learning Machine and Online Sequential-Extreme Learning Machine have been selected as the most appropriate prediction techniques for forecasting short-term and medium-term customer electricity load behaviour. These were chosen to deal with the non-stationary properties of customer load demand and neural network fields because they have proven to be the most efficacious techniques to apply in such cases of non-linearity [78].

As outlined in the earlier literature review, the majority of load forecasting objectives are directed at achieving the most accurate estimates of future load values as the means of maximizing potential savings by electricity corporations [79]. However, in the research reported here, the main objective of the prediction module that has been devised is to apply the most accurate forecasted values flowing from normal electricity load behaviour as benchmarks to detect abnormalities in load behaviour in a timely manner. To this end, the prediction results obtained empirically from the Extreme Learning Machine techniques are compared with those from the support vector machine technique to assess their relative accuracy based on Root Mean Squared Error (RMSE) results. The model that delivers its outcomes faster and with greater prediction accuracy will be recommended for inclusion in the proposed framework of analysis.

6.2 Definition of Task

Several case studies [80, 81, and 82] have shown that commercial and industrial customers normally have consumption behaviours that remain consistent throughout the year. In the present research, a case study that focuses on Malaysia is the subject of investigation. Generally, an unmetered power loss flows from abnormal electricity consumption behaviour that deviates from regular behaviour. Such loss is identified according to patterns that apply to days of the week. Here, the customer behaviour is classified as falling into one of the following four categories.

Definition 1: Unmetered power losses Electricity load consumption that registers as zero throughout from 1 to 30 days is considered as indicating an unmetered power loss.

Definition 2: Abnormal unmetered power losses Electricity load consumption that registers any zero-consumption period within from 1 to 30 days is considered as suspect because it indicates abnormal unmetered power losses.

Definition 3: Suspicious unmetered power losses Electricity load consumption that deviates from $P_{Load} = \mu \pm 3\sigma$ within from 1 to 30 days is also considered as suspect because it indicates suspicious unmetered power losses.

Definition 4: Normal Behaviour Electricity load consumption that is consistently within the range of $P_{Load} = \mu \pm 3\sigma$ within from 1 to 30 days is considered as indicating normal behaviour that is designated as power losses.

Zhao [83] suggested several steps that comprise an unmetered power losses forecasting Module. The steps set out below are based on modifications and extensions of the framework proposed by Zhao.

- Determine the unmetered power losses in daily consumption with $P_{Load} = \mu \pm 3\sigma$ as the unmetered power losses threshold
- Training the normal daily consumption, suspicious unmetered power losses consumption, and unmetered power losses value predictors.
- For each future point with its relevant factor Xt, determine whether it represents suspicious unmetered power losses behaviour with the suspicious unmetered power losses occurrence predictor.
- If a suspicious unmetered power loss is predicted at time t, use the suspicious unmetered power losses value predictor to estimate the suspicious value.
- Otherwise, use the normal consumption value model to forecast the price at time t.
- Combine both the normal forecasting results and the suspicious unmetered power losses predictions to form the complete results.

Determining the unmetered power losses status

To describe the relationship between any suspicious unmetered power losses value and its predecessor, one attribute has been created. The value can be unmetered power losses, power losses, or suspicious:

$$\textbf{Unmetered power losses } \mathbf{I_{ex}} = \begin{cases} \text{Unmetered power losses} \\ \text{Power losses} \\ \text{Suspicious Unmetered power losses} \\ \text{Abnormal} \end{cases}$$

Several measures have been used to assess load forecasting accuracy, including those in [84, 85, 86]. The most popular classification performance measure of such accuracy and that which is to be used in this study is defined in the following terms [87].

Input: Inputs for clustering techniques are listed as follows.

- A set of individual electricity consumer load patterns. For a population of L consumers, each customer load profile is characterised by a vector, $X^l = \{X_h{}^l, h = 1, \dots, H\}$, whose H consumer of time-domain data correspond to 30-days interval data. The consumer data has been separated into three datasets: training datasets, testing datasets, and validation datasets. These relate to the three years from 2011 until 2014.

- Training data comprises winter datasets, summer datasets, and monsoon datasets.
- Testing data comprises winter datasets, summer datasets, and monsoon datasets.
- Numbers of neurons for Extreme Learning Machine, OS-ELM, and SVM.
- Transfer functions for all algorithms.

Output: The following outputs are expected from applying clustering techniques.

- Classification accuracy rates as percentages for training and testing datasets for winter, summer, and monsoon.
- Time processing speeds in seconds for training and testing datasets for winter, summer, and monsoon.

Hardware and Software: All the classification simulations for Extreme Learning Machine, OS-Extreme Learning Machine, 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 prediction experiments in this research were conducted using MATLAB 14 software with ELM and OS-ELM toolboxes.

Algorithms: Two algorithms, Extreme Learning Machine and OS-Extreme Learning Machine, have been selected to run the data.

Procedure: The following steps will comprise the clustering process.

- Three sets of individual customers are subjects of the experiment based on their differences, including 1) first case study: customers with normal behaviour datasets, 2) second case study: customers with normal and abnormal behaviour datasets, and 3) third customer case study: customers with abnormal behaviour datasets.
- Train the customer data – In each case study, a set of individual customer data is trained. Those datasets are separated according to the day of the week to which they relate, from summer until winter, together with any additional day representing a monsoon.
- Apply the test data – A different set of customer data is supplied for the testing procedure. Another method for this procedure is cross validation [88].
- Compare the classification accuracy as percentages and time processing durations in seconds.

6.3 Prediction Results

The prediction technique was tested with three selected categories of commercial customers in accordance with the definitions specified above. First, a commercial customer identified as having normal behaviour throughout the year. Secondly, a commercial customer identified as having some mixed pattern of normal behaviour, suspicious behaviour, and abnormal behaviour. Thirdly, a commercial customer identified as having abnormal behaviour throughout the year.

6.3.1 Case 1 – Modelling a Customer with Normal Behaviour

In the case of customer 1 identified as exhibiting normal behaviour, comprehensive experiments were conducted to study the effectiveness of the Extreme Learning Machine and OS-Extreme Learning Machine prediction techniques.

Table 6-1: Mean and standard deviation of seasonal consumption for customer 1

Seasons	Area	Mean	Std. Deviation	Upper Limit Threshold	Lower Limit Threshold
Winter	HT Exp	102.29	6.018	120.344	84.236
	Industrial	103.19	6.830	123.68	82.7
	GIDC	105.85	7.606	128.668	83.032
	Urban	107.13	7.435	129.435	84.825
	Jyotigram	104.34	7.249	126.087	82.593
	Agriculture	107.19	8.950	134.04	80.34
Summer	HT Exp	108.66	8.779	134.997	82.323
	Industrial	102.32	8.681	128.363	76.277
	GIDC	101.79	7.583	124.539	79.041
	Urban	109.40	7.596	132.188	86.612
	Jyotigram	104.85	8.735	131.055	78.645
	Agriculture	104.71	8.055	128.875	80.545
Monsoon	HT Exp	107.87	8.709	133.997	81.743
	Industrial	106.93	8.621	132.793	81.067
	GIDC	105.05	9.553	133.709	76.391
	Urban	104.10	8.586	129.858	78.342
	Jyotigram	101.00	8.755	127.265	74.735
	Agriculture	107.31	9.005	134.325	80.295

Table 6-1 shows the mean and standard deviation of electricity consumption for customer 1 from summer until winter. From these values, upper and lower limit thresholds were established. Based on these data, the daily consumption for customer 1 from HT Exp. to Agriculture is shown graphically in Figure 6-1. This customer is said to have normal behaviour because the electricity consumption recorded falls within the threshold limits or has similar curves for seasonal.

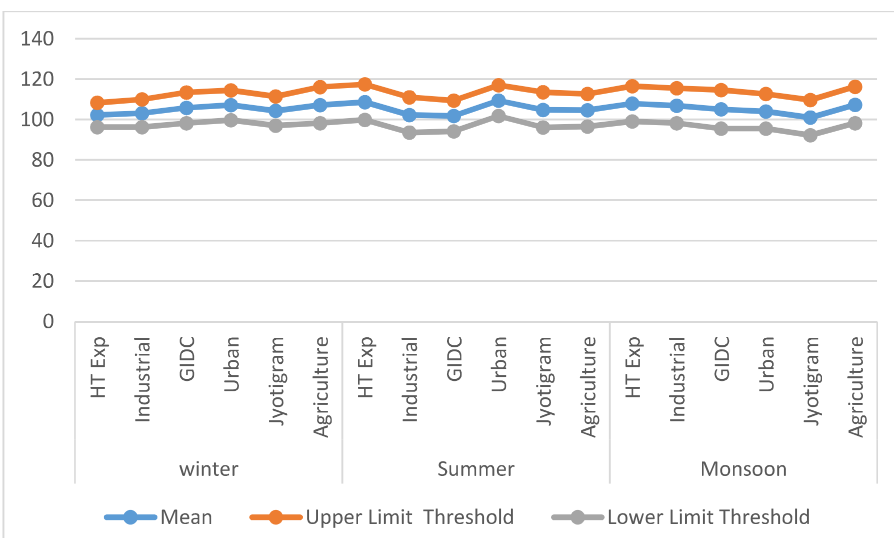

Figure 6.1 Average kWh for customer 1 with upper limit and lower limit

Table 6-2 Root Mean Squared Error Results with the Prediction Algorithms based on different season

Seasons	Area	ELM Sigmoid		ELM Radial basis function		OSELM Sigmoid		OSELM Radial basis function	
		Training	Testing	Training	Testing	Training	Testing	Training	Testing
Winter	HT Exp	0.0456	0.0823	0.0281	0.0571	0.0244	0.0482	0.0384	0.0831
	Industrial	0.0323	0.0434	0.0122	0.0258	0.0325	0.0432	0.0456	0.0993
	GIDC	0.0424	0.0556	0.0174	0.0393	0.0496	0.0989	0.0377	0.0776
	Urban	0.0268	0.0456	0.0235	0.0478	0.0283	0.0359	0.0438	0.0455
	Jyotigram	0.0367	0.0722	0.0176	0.0258	0.0274	0.0571	0.0284	0.0280
	Agriculture	0.0342	0.0623	0.0286	0.0381	0.066	0.1303	0.0267	0.0574
Summer	HT Exp	0.0424	0.0868	0.0177	0.0380	0.0367	0.0775	0.0368	0.0737
	Industrial	0.0321	0.0719	0.0168	0.0341	0.0226	0.0445	0.044	0.059
	GIDC	0.0398	0.0799	0.0263	0.0535	0.0233	0.0460	0.0268	0.0284
	Urban	0.0306	0.0454	0.0233	0.0374	0.0476	0.0960	0.0489	0.0983
	Jyotigram	0.0382	0.0445	0.0266	0.0591	0.0246	0.0486	0.0333	0.0773
	Agriculture	0.0383	0.0511	0.0282	0.0373	0.0482	0.0952	0.0485	0.0965
Monsoon	HT Exp	0.0265	0.0344	0.0171	0.0247	0.0279	0.0531	0.0383	0.0859
	Industrial	0.0287	0.0316	0.0262	0.0583	0.0360	0.0711	0.0458	0.0927
	GIDC	0.0379	0.0867	0.0141	0.0186	0.0462	0.0512	0.0215	0.0474
	Urban	0.0482	0.0923	0.0228	0.0463	0.0347	0.0685	0.0307	0.0668
	Jyotigram	0.0439	0.0894	0.0178	0.0382	0.0278	0.0379	0.0203	0.0459
	Agriculture	0.0328	0.0745	0.0265	0.0569	0.0343	0.0677	0.0445	0.0930

It is apparent from Table 6-2 that different prediction algorithms generate different error rates. For summer (Urban and Jyotigram), winter (Industrial and GIDC) and monsoon (HT Exp. and Industrial), Extreme Learning Machine with the sigmoid function gives the lowest error rates. However, for winter (Industrials and Urban) and monsoon (GIDC and Jyotigram), Online Sequential- Extreme Learning Machine with the sigmoid function produces the lowest error rates. With respect to the higher error rates, ELM radial basis function produced the highest error rates for summer (Urban and Agriculture), winter (Jyotigram and Agriculture) and monsoon (HT Exp. and GIDC) while Online Sequential-Extreme Learning Machine with radial basis function nodes produced the highest error rates for winter (Urban and Jyotigram) and summer (GIDC and Industrial). From these observations, it can be concluded that the sigmoid activation function produced lower error rates when compared to the radial basis function nodes function.

6.3.2 Case 2 – Modelling a Customer with mixed normal and abnormal behaviour

Commercial customer 2 who exhibits a combination of normal, abnormal, and suspicious behaviours selected for the prediction experiments in this case. The comparisons of means and standard deviations on season for consumption identified as normal behaviour, suspicious behaviour, and abnormal behaviour are shown in Tables 6-3, 6-4 and 6-5, respectively.

Table 6-3: Mean and standard deviation of seasonal consumption for normal behaviour

Seasons	Area	Mean	Std. Deviation	Upper Limit Threshold	Lower Limit Threshold
Winter	HT Exp	107.19	4.210	119.82	94.56
	Industrial	106.09	5.438	122.404	89.776
	GIDC	104.95	6.604	124.762	85.138
	Urban	105.23	4.734	119.432	91.028
	Jyotigram	106.44	4.846	120.978	91.902
	Agriculture	106.59	4.053	118.749	94.431
Summer	HT Exp	107.76	4.472	121.176	94.344
	Industrial	106.22	5.687	123.281	89.159
	GIDC	106.99	6.881	127.633	86.347
	Urban	105.70	5.396	121.888	89.512
	Jyotigram	106.35	5.934	124.152	88.548
	Agriculture	107.01	6.358	126.084	87.936
Monsoon	HT Exp	105.27	4.700	119.37	91.17
	Industrial	107.53	5.820	124.99	90.07
	GIDC	106.85	5.551	123.503	90.197
	Urban	106.60	6.089	124.867	88.333
	Jyotigram	105.20	6.153	123.659	86.741
	Agriculture	104.71	3.202	114.316	95.104

Table 6-4: Mean and standard deviation of seasonal consumption for abnormal behaviour

Seasons	Area	Mean	Std. Deviation	Upper Limit Threshold	Lower Limit Threshold
Winter	HT Exp	101.39	5.547	118.031	84.749
	Industrial	100.60	4.842	115.126	86.074
	GIDC	102.63	5.294	118.512	86.748
	Urban	103.23	5.926	121.008	85.452
	Jyotigram	105.85	4.281	118.693	93.007
	Agriculture	104.20	4.303	117.109	91.291
Summer	HT Exp	105.02	5.416	121.268	88.772
	Industrial	101.22	6.587	120.981	81.459
	GIDC	103.69	4.627	117.571	89.809
	Urban	104.71	5.789	122.077	87.343
	Jyotigram	103.35	4.782	117.696	89.004
	Agriculture	103.61	3.326	113.588	93.632
Monsoon	HT Exp	104.47	2.482	111.916	97.024
	Industrial	105.23	3.670	116.24	94.22
	GIDC	106.24	3.436	116.548	95.932
	Urban	102.02	4.596	115.808	88.232
	Jyotigram	102.91	6.069	121.117	84.703
	Agriculture	104.42	7.051	125.573	83.267

Table 6-5: Mean and standard deviation of seasonal consumption for suspicious behaviour

Seasons	Area	Mean	Std. Deviation	Upper Limit Threshold	Lower Limit Threshold
Winter	HT Exp	109.16	8.165	133.655	84.665
	Industrial	109.74	9.130	137.13	82.35
	GIDC	108.31	8.418	133.564	83.056
	Urban	108.60	9.635	137.505	79.695
	Jyotigram	107.77	8.028	131.854	83.686
	Agriculture	107.36	9.250	135.11	79.61
Summer	HT Exp	108.25	8.869	134.857	81.643
	Industrial	108.59	9.585	137.345	79.835
	GIDC	108.05	8.553	133.709	82.391
	Urban	108.35	8.498	133.844	82.856
	Jyotigram	107.53	9.295	135.415	79.645
	Agriculture	107.27	9.554	135.932	78.608
Monsoon	HT Exp	108.64	8.929	135.427	81.853
	Industrial	109.48	8.626	135.358	83.602
	GIDC	109.22	9.503	137.729	80.711
	Urban	109.85	8.587	135.611	84.089
	Jyotigram	108.32	8.485	133.775	82.865
	Agriculture	108.50	9.100	135.8	81.2

Table 6-6: Root Mean Squared Error Results for normal behaviour

Seasons	Area	ELM Sigmoid		ELM Radial basis function		OSELM Sigmoid		OSELM Radial basis function	
		Training	Testing	Training	Testing	Training	Testing	Training	Testing
Winter	HT Exp	0.0854	0.0955	0.0563	0.1182	0.0388	0.0824	0.0497	0.0532
	Industrial	0.0824	0.874	0.0619	0.0633	0.0449	0.0535	0.0388	0.0815
	GIDC	0.0806	0.1637	0.0570	0.0645	0.0480	0.0565	0.0455	0.1006
	Urban	0.0727	0.1479	0.0322	0.0656	0.0368	0.0776	0.0334	0.0768
	Jyotigram	0.0868	0.1703	0.0474	0.0960	0.0347	0.0789	0.0404	0.0852
	Agriculture	0.0792	0.1612	0.0597	0.1172	0.0396	0.0833	0.0423	0.0432
Summer	HT Exp	0.0821	0.1667	0.0678	0.1388	0.0465	0.0952	0.0302	0.0667
	Industrial	0.0750	0.0823	0.0562	0.0598	0.0434	0.0525	0.0211	0.0268
	GIDC	0.0656	0.1394	0.0444	0.0837	0.0323	0.0618	0.0308	0.0659
	Urban	0.0745	0.1518	0.0388	0.0449	0.0318	0.0608	0.0348	0.0439
	Jyotigram	0.0713	0.0741	0.0464	0.0980	0.0459	0.0532	0.0495	0.1092
	Agriculture	0.0829	0.1686	0.0558	0.1171	0.0432	0.0945	0.0376	0.0871
Monsoon	HT Exp	0.0805	0.1643	0.0672	0.1372	0.0399	0.0864	0.0478	0.1051
	Industrial	0.0783	0.1595	0.0668	0.1331	0.0384	0.0836	0.0387	0.0396
	GIDC	0.0760	0.1588	0.0532	0.1072	0.0465	0.0968	0.0362	0.0793
	Urban	0.0768	0.0805	0.0527	0.0561	0.0457	0.0977	0.0334	0.0762
	Jyotigram	0.0852	0.1728	0.0548	0.0551	0.0443	0.0467	0.0416	0.0423
	Agriculture	0.0705	0.0720	0.0562	0.1130	0.0327	0.0375	0.0233	0.0548

In Table 6-6, it is apparent that for customer 2 with normal behaviour, Extreme Learning Machine with the sigmoid function produced the lowest error rates for winter (HT Exp and

Industrial), summer (Jyotigram and Industrial) and monsoon (Urban and Agriculture), while Online Sequential-Extreme Learning Machine with the sigmoid function produced the lowest error rates for summer (Industrial and Jyotigram) and monsoon (Jyotigram and Agriculture). However, in the winter (Industrial and GIDC) dataset, Online Sequential-Extreme Learning Machine with radial basis function nodes showed the lowest error rate. Meanwhile, most of the highest error rates were produced by Extreme Learning Machine radial basis function on winter (Industrial and GIDC), summer (Industrial and Urban) and monsoon (Urban and Jyotigram) and by Online Sequential-Extreme Learning Machine radial basis function on winter (HT Exp and Agriculture), summer (Industrial and Urban) and monsoon (Industrial and Jyotigram). From these observations it can be confirmed that Extreme Learning Machine with the sigmoid function produced the lowest error rates overall, while Extreme Learning Machine radial basis function produced the highest error rates. Online Sequential-Extreme Learning Machine with the sigmoid function and with the radial basis function nodes function produced either the lowest or the highest error rates for normal behaviour for customer 2.

Table 6-7: Root Mean Squared Error Results for abnormal behaviour

Seasons	Area	ELM Sigmoid		ELM Radial basis function		OSELM Sigmoid		OSELM Radial basis function	
		Training	Testing	Training	Testing	Training	Testing	Training	Testing
Winter	HT Exp	0.0385	0.0742	0.0363	0.0733	0.0683	0.1224	0.0141	0.0273
	Industrial	0.0457	0.0974	0.0448	0.0495	0.0547	0.0566	0.0364	0.0767
	GIDC	0.0443	0.0887	0.0472	0.0947	0.0564	0.1068	0.0387	0.0768
	Urban	0.0328	0.0669	0.0477	0.0512	0.0638	0.1259	0.0383	0.0399
	Jyotigram	0.0399	0.0510	0.0432	0.0885	0.0688	0.0732	0.0498	0.0971
	Agriculture	0.0480	0.0492	0.0428	0.0828	0.0587	0.1111	0.0402	0.0420
Summer	HT Exp	0.0382	0.0743	0.0407	0.0882	0.0538	0.1050	0.0218	0.0423
	Industrial	0.0374	0.0745	0.0306	0.0668	0.0557	0.1032	0.0227	0.0447
	GIDC	0.0265	0.0527	0.0432	0.0822	0.0576	0.1073	0.0367	0.0727
	Urban	0.0357	0.0731	0.0226	0.0478	0.0585	0.0667	0.0435	0.0855
	Jyotigram	0.0448	0.0898	0.0424	0.0463	0.0664	0.1249	0.0483	0.0489
	Agriculture	0.0439	0.0867	0.0373	0.0428	0.0652	0.0667	0.0436	0.0454
Monsoon	HT Exp	0.0221	0.0286	0.0385	0.0760	0.0689	0.0796	0.0382	0.0397
	Industrial	0.0391	0.0777	0.0491	0.0962	0.0568	0.1088	0.0396	0.0413
	GIDC	0.0480	0.0488	0.0450	0.0473	0.0647	0.1252	0.0333	0.0357
	Urban	0.0472	0.0945	0.0459	0.0434	0.0605	0.0637	0.0442	0.0862
	Jyotigram	0.0352	0.0705	0.0476	0.0986	0.0503	0.0956	0.0458	0.0901
	Agriculture	0.0340	0.0668	0.0338	0.0675	0.0622	0.1172	0.0476	0.0922

In Table 6-7, it is apparent that for customer 2 with abnormal behaviour, Extreme Learning Machine with the sigmoid function produced lowest error rates for winter (Jyotigram and Agriculture) and monsoon (HT Exp and GIDC) while Online Sequential-Extreme Learning Machine with the sigmoid function produced lowest error rates for winter (Industrial and Jyotigram), summer (Urban and Agriculture) and monsoon (HT Exp and Urban). It is apparent, too, that the highest error rates are produced mainly by Extreme Learning Machine with radial basis function nodes on winter (Industrial and Urban), summer (Jyotigram and Agriculture) and monsoon (GIDC and Urban). The Online Sequential-Extreme Learning Machine with radial basis function nodes tends to produce both highest error rates on winter (Agriculture and

Urban) and summer (Jyotigram and Agriculture) and lowest error rates on summer (HT Exp, GIDC and Industrial).

Table 6-8: Root Mean Squared Error Results for suspicious behaviour

Seasons	Area	ELM Sigmoid		ELM Radial basis function		OSELM Sigmoid		OSELM Radial basis function	
		Training	Testing	Training	Testing	Training	Testing	Training	Testing
Winter	HT Exp	0.0034	0.0036	0.0023	0.0032	0.0035	0.0051	0.0051	0.0064
	Industrial	0.0055	0.0075	0.0039	0.0047	0.0023	0.0032	0.0049	0.0067
	GIDC	0.0055	0.0074	0.0067	0.0099	0.0046	0.0066	0.0034	0.0043
	Urban	0.0038	0.0053	0.0045	0.0053	0.0037	0.0058	0.0033	0.0046
	Jyotigram	0.0055	0.0059	0.0033	0.0049	0.0058	0.0089	0.0037	0.0047
	Agriculture	0.0034	0.0043	0.0068	0.0093	0.0049	0.0070	0.0038	0.0042
Summer	HT Exp	0.0069	0.0098	0.0058	0.0081	0.0035	0.0053	0.0056	0.0065
	Industrial	0.0043	0.0048	0.0056	0.0077	0.0044	0.0062	0.0044	0.0059
	GIDC	0.0077	0.0084	0.0044	0.0049	0.0055	0.0071	0.0043	0.0052
	Urban	0.0038	0.0053	0.0032	0.0043	0.0046	0.0062	0.0035	0.0045
	Jyotigram	0.0048	0.0065	0.0063	0.0080	0.0047	0.0063	0.0047	0.0059
	Agriculture	0.0034	0.0047	0.0045	0.0052	0.0059	0.0087	0.0044	0.0053
Monsoon	HT Exp	0.0071	0.0094	0.0074	0.0104	0.0043	0.0063	0.0035	0.0048
	Industrial	0.0052	0.0073	0.0050	0.0071	0.0032	0.0041	0.0037	0.0042
	GIDC	0.0048	0.0067	0.0049	0.0067	0.0041	0.0059	0.0052	0.0063
	Urban	0.0067	0.0091	0.0066	0.0096	0.0044	0.0066	0.0058	0.0077
	Jyotigram	0.0059	0.0064	0.0048	0.0065	0.0053	0.0073	0.0034	0.0043
	Agriculture	0.0035	0.0040	0.0067	0.0097	0.0042	0.0060	0.0033	0.0041

In Table 6-8, it is shown that for customer 2 with suspicious behaviour, Extreme Learning Machine with the sigmoid function produced the lowest error rates on winter (HT Exp and Jyotigram), summer (Industrial and GIDC) and monsoon (Jyotigram and Agriculture), while Extreme Learning Machine with radial basis function nodes produced the lowest error rates on summer (GIDC and Agriculture) and winter (Industrial and Urban), and Online Sequential - Extreme Learning Machine with the sigmoid function produced the lowest error rate on monsoon. From the results in Table 6-8, it can also be seen that Extreme Learning Machine with the sigmoid function produced the lowest error rates and Online Sequential-Extreme Learning Machine with Radial basis function produced the highest error rates. However, both Extreme Learning Machine with Radial basis function nodes and Online Sequential-Extreme Learning Machine with the sigmoid function can produce either higher or lower error rates depending on the days of the week.

6.3.3 Case 3 – Modelling a Customer with abnormal behaviour

Commercial customer 3 has been identified as showing suspicious behaviour because the relevant records reveal instances of zero consumption in the customer profiling. After investigation and validation involving the average load profiles and standard deviations, it is confirmed that this customer does have a load curve that includes zero consumption. The data examined below are applied to develop the load curve based on types of days.

Table 6-9: Mean and standard deviation of seasonal consumption for customer 3

Seasons	Area	Mean	Std. Deviation	Upper Limit Threshold	Lower Limit Threshold
Winter	HT Exp	100.32	4.445	113.655	86.985
	Industrial	101.74	4.660	115.72	87.76
	GIDC	103.06	4.778	117.394	88.726
	Urban	103.57	4.885	118.225	88.915
	Jyotigram	103.14	4.638	117.054	89.226
	Agriculture	102.43	6.334	121.432	83.428
Summer	HT Exp	100.24	5.595	117.025	83.455
	Industrial	101.46	5.825	118.935	83.985
	GIDC	102.03	4.800	116.43	87.63
	Urban	103.15	5.478	119.584	86.716
	Jyotigram	103.04	4.928	117.824	88.256
	Agriculture	104.55	4.345	117.585	91.515
Monsoon	HT Exp	100.78	5.804	118.192	83.368
	Industrial	101.67	5.352	117.726	85.614
	GIDC	102.26	6.508	121.784	82.736
	Urban	103.40	5.236	119.108	87.692
	Jyotigram	104.21	5.623	121.079	87.341
	Agriculture	103.57	5.085	118.825	88.315

In Table 6-9, commercial customer 3 is categorized as exhibiting abnormal behaviour because the data reveal consistent zero consumption recordings during the season. This customer is used as an example to show how the prediction techniques can be applied to forecasting cases of abnormal consumption. Based on the means and standard deviations, upper and lower limit thresholds are set up. However, during the zero consumption periods, those thresholds will become negative, an outcome that does not make sense in a load consumption profile. Therefore, in this study, all the negative values from the lower limit threshold have been converted into zero values. Moreover, after applying the prediction techniques, any forecasted negative values in the profile have also been converted into zero values.

Table 6-10: Root Mean Squared Error Results with the Extreme Learning Machine and Online Sequential-Extreme Learning Machine Prediction Algorithms based on different season

Seasons	Area	ELM Sigmoid		ELM Radial basis function		OSELM Sigmoid		OSELM Radial basis function	
		Training	Testing	Training	Testing	Training	Testing	Training	Testing
Winter	HT Exp	0.0567	0.1524	0.0393	0.0735	0.0567	0.1157	0.1004	0.1986
	Industrial	0.0955	0.1894	0.0569	0.1063	0.0786	0.1599	0.0365	0.0725
	GIDC	0.1043	0.2077	0.0753	0.1915	0.0844	0.1713	0.0557	0.1105
	Urban	0.0156	0.0312	0.0817	0.1532	0.0978	0.1987	0.0736	0.1452
	Jyotigram	0.0392	0.0442	0.1024	0.3025	0.0347	0.0702	0.0884	0.1748
	Agriculture	0.0558	0.1108	0.0262	0.0497	0.0886	0.1809	0.0963	0.1902
	HT Exp	0.0872	0.1732	0.0489	0.0912	0.1038	0.2113	0.0349	0.0686
	Industrial	0.0329	0.0657	0.0653	0.0686	0.0986	0.2006	0.0223	0.0843
	GIDC	0.1022	0.2536	0.0848	0.1523	0.0183	0.0393	0.0595	0.1178

Summer	Urban	0.0784	0.1556	0.0922	0.1754	0.0337	0.0385	0.0736	0.0752
	Jyotigram	0.0627	0.0647	0.0378	0.0715	0.0683	0.1384	0.0884	0.1745
	Agriculture	0.0278	0.0552	0.0132	0.0578	0.0895	0.1817	0.0286	0.0565
Monsoon	HT Exp	0.0753	0.1497	0.0388	0.0724	0.0730	0.349	0.0763	0.1506
	Industrial	0.0435	0.0462	0.0623	0.1178	0.0502	0.1025	0.0746	0.1478
	GIDC	0.0989	0.1969	0.0287	0.0538	0.0519	0.2057	0.0820	0.1622
	Urban	0.0273	0.0542	0.0835	0.1960	0.0852	0.1732	0.0971	0.1916
	Jyotigram	0.0168	0.0337	0.0525	0.0582	0.0858	0.1784	0.0951	0.3078
	Agriculture	0.0853	0.1795	0.0884	0.1661	0.1086	0.2276	0.0230	0.0252

In Table 6-10, it is apparent that Extreme Learning Machine with the sigmoid function produced the lowest error rates on winter (HT Exp and Jyotigram), summer (GIDC and Jyotigram) and monsoon (Industrial and Agriculture) while Online Sequential-Extreme Learning Machine with the sigmoid function produced the lowest error rates on winter (Industrial and Urban), summer (GIDC and Urban) and monsoon (HT Exp and GIDC). Moreover, Extreme Learning Machine with radial basis function nodes produced the highest error rates on winter (GIDC and Jyotigram), summer (Agriculture and Industrial) and monsoon (Urban and Jyotigram), while Online Sequential-Extreme Learning Machine with radial basis function nodes produced the highest error rates on summer (Industrial and Urban) and monsoon (Jyotigram and Agriculture).

6.4 Analysis of Customer Behaviour Predictions

In the first case study, a customer profile identified as normal behaviour in accordance with the classification process outlined in chapter 4 was selected. The profile details were separated into summer, winter and monsoon to ensure that behaviour variations across days were captured. In this first case study, the analysis revealed that monsoon datasets shown the lowest error rates, while summer datasets shown the highest error rates.

In the second case study, a customer profile identified as representing a combination of the three behaviours, normal, abnormal, and suspicious, was selected. The profile details were separated into datasets for season summer, winter and monsoon. From this case study, on average, the forecasting for normal behaviour showed that the monsoon dataset has the lowest error rates, while the summer dataset has the highest error rates. By contrast, the forecasting for abnormal behaviour showed that the winter dataset has the lowest error rates. At the same time, it was evident that forecasting suspicious behaviour showed that the winter dataset has the lowest error rates and the summer dataset has the highest error rates.

In the third case study, a customer profile identified as abnormal behaviour was selected. Such behaviour had been detected because of consistent instances of zero consumption. The profile details were separated into datasets for summer, winter and monsoon. From this case study, on average, the forecasting results for abnormal behaviour showed that the monsoon dataset has the lowest error rates, while the summer dataset has the highest error rates.

6.5 Conclusions

This chapter witness, few of the greatly used prediction techniques for forecasting the load demands and the same is categorically represented as winter forecast, summer forecasts and monsoon forecasts. A case study was carried out on different load behaviours of three commercial customers. The base of the data, for conducting the test, was accommodating two years inter-data intervals. For grouping the customer's behaviour, two different prediction algorithms, namely Online Sequential-Extreme Learning Machine and Extreme Learning Machine, has been used to carry out the study. Relative performances of the two said methods has been compared and analyzed. Extreme Learning Machine algorithm has been proven to be faster than that of the support vector machine in batch processing mode, at the same time it has also proven its worth when compared to Online Sequential - Extreme Learning Machine algorithm in online sequential learning mode. Prediction accuracy of Extreme Learning Machine is found better then Online Sequential-Extreme Learning Machine algorithm, and it generated better results in terms of the actual classification rates.

- It has been found that algorithm of Extreme Learning Machine with the sigmoid function produces the lowest error rates in forecasting load results, while Extreme Learning Machine with radial basis function nodes produced the highest error rates in forecasting load results, based on types of seasons, for normal behaviour datasets.

- For abnormal behaviour datasets, Online Sequential-Extreme Learning Machine with the sigmoid function produces lowest error rates in forecasting load results. Online Sequential-Extreme Learning Machine with radial basis function nodes produces the highest error rates in forecasting load results, based on types of seasons.

- While considering suspicious behaviour datasets, Extreme Learning Machine with the sigmoid function produces the lowest error rates in forecasting load results. Extreme Learning Machine, with radial basis function nodes, produces highest error rates in forecasting load results, based on types of seasons.