IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received November 20, 2019, accepted December 18, 2019, date of publication January 8, 2020, date of current version January 15, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2964885

An Ensemble Detection Model Using Multinomial Classification of Stochastic Gas Smart Meter Data to Improve Wellbeing Monitoring in Smart Cities

WILLIAM HURST, CASIMIRO ADAY CURBELO MONTAÑEZ, NATHAN SHONE, AND DHIYA AL-JUMEILY, (Senior Member, IEEE)
Department of Computer Science, Liverpool John Moores University, Liverpool L3 3AF, U.K.

Corresponding author: William Hurst (w.hurst@ljmu.ac.uk)

This work was supported by EPRSC-EP/R020922/1 Grant.

ABSTRACT Fuel poverty has a negative impact on the wellbeing of individuals within a household; affecting not only comfort levels but also increased levels of seasonal mortality. Wellbeing solutions within this sector are moving towards identifying how the needs of people in vulnerable situations can be improved or monitored by means of existing supply networks and public institutions. Therefore, the focus of this research is towards wellbeing monitoring solution, through the analysis of gas smart meter data. Gas smart meters replace the traditional analogue electro-mechanical and diaphragm-based meters that required regular reading. They have received widespread popularity over the last 10 years. This is primarily due to the fact that by using this technology, customers are able to adapt their consumption behaviours based on real-time information provided by In-Home Devices. Yet, the granular nature of the datasets generated has also meant that this technology is ideal for further scalable wellbeing monitoring applications. For example, the autonomous detection of households at risk of energy poverty is possible and of growing importance in order to face up to the impacts of fuel poverty, quality of life and wellbeing of low-income housing. However, despite their popularity (smart meters), the analysis of gas smart meter data has been neglected. In this paper, an ensemble model is proposed to achieve autonomous detection, supported by four key measures from gas usage patterns, consisting of i) a tariff detection, ii) a temporally-aware tariff detection, iii) a routine consumption detection and iv) an age-group detection. Using a cloud-based machine learning platform, the proposed approach yielded promising classification results of up to 84.1% Area Under Curve (AUC), when the Synthetic Minority Over-sampling Technique (SMOTE) was utilised.

INDEX TERMS Energy and fuel poverty, gas, machine learning, smart meter, smart cities, wellbeing.

I. INTRODUCTION

Fuel poverty remains a prevalent concern [1], [2]; where consumers with long-term health conditions, or individuals living on a low income, can find themselves in the position of whether to keep their homes at a comfortable temperature or pay their energy bills [3]. Yet, with technology improvements in the energy sector, new opportunities have arisen [4]. Smart city technologies can now play a key role in improving the wellbeing of such vulnerable households through use of existing digital technologies [1], [5].

Particularly, this industry has witnessed important technological developments in the real-time data analytics surrounding the generation, transmission, and consumption of water, gas and electricity [5]. An example is the smart meter, a technology that provides real-time consumption information and automates the billing process for the customer and supplier. It is well-documented that the smart meter can play a key role in the reduction in energy poverty. For example, the EU-funded SMART-UP[1] project works with vulnerable customers, who have smart meters installed, to achieve energy savings through small changes to their habits that result in an improvement their living conditions and help to reduce their energy bills. However, is a 'hands-on' approach, reliant on a network of staff working with vulnerable households across member states.

The associate editor coordinating the review of this manuscript and approving it for publication was Miltiadis Lytras.

[1] https://www.smartup-project.eu/about/

Yet, the data generated from smart meters has shown to be technically reliable for the remote and autonomous profiling of individuals at home [6], the detection their age grouping [7] and their monitoring general health [8]. All of which are applications that offer demonstrate how smart gas meter data is technically reliable to measure consumer demographics and support vulnerable households [9]. Such insights have been proven to be possible using both the default 30-minute data samples but also through use of the high-resolution data gathered from the smart meters [8], [10].

Smart meters consist of three layers of technology; i) the physical meter, ii) the communication layer network management and iii) the computer systems that manage the data applications and services [11]. This technology has revolutionised the process for monitoring end-users' consumption of gas and electricity as is a core part of the smart city infrastructure. It is technologies such as this that enable researchers to identify and exploit diverse data in meaningful ways to assist the development of new policies [12] but also establish practical and scalable solutions to modern-day wellbeing challenges. These layers of technology have transformed the dynamics of the power and gas distribution system. Energy distribution, which was once so predictable, is now dynamic and has a random probability distribution pattern. Yet, it has increased the personalised element, enabling the utility provider to have a better understanding of their customers' consumption behaviours.

Typically, smart meters record the consumption at 30-minute intervals. This information is shared mutually with the user, if they have an In-Home Device (IHD), and the operator. Up-to-date information is then provided regarding the consumption amounts, with high levels of accuracy. The readings are then used by the company for purposes such as load balancing, forecasting and accurate billing.

In 2014, a report issued by the European Commission outlined that there was an intention for 45 million gas smart meters to be rolled out within the European Union by the end of 2020. This is the equivalent of around 40% of existing customers owning a gas smart meter. The ambition behind this project is two-fold. Firstly, to provide a more cost-efficient system to the end user, as on average, smart meters provide savings of €160 for gas and €309 for electricity. Secondly, to reduce energy consumption, as on average the energy saving is around 3%.[2]

Smart meters generate a gold-mine of data. Therefore, in addition to the aforementioned benefits of the smart metering infrastructure, an increasing number of projects have emerged offering potential beneficial applications to both the end-user and utility company. Particularly, within this area, a significant number or researchers investigate applications relating to forecasting customer demand [13]. This area of research is particularly challenging, given the high variability of end users' behaviour.

Furthermore, it is also a considerable task to process the data within a smart cities setting, given then volume of data generated. For example, each smart meter generates in the region of 400MB of data on a yearly basis. Consequently, this results in an estimated 4.8 petabytes' worth of data annually. Analysing this dataset is a considerable big data challenge for any utility provider; and just like the work in this paper, a data analytics process will require the use of a cloud-based data processing platform.

Other studies focus on profiling within the smart grid to discern user behaviours, to support demand-side management systems [14]. However, research within this area often requires direct user input provided through survey questions to produce sample representative load profiles. This type of approach may also involve the use of either sub-second sampling to detect appliance usage around the home, or the use of 10-second data samples to detect appliances that are classed as used within the area of activities of daily living to produce effective results [15], [16]. Low samples are required for i) device detection but also ii) for maintaining the uniqueness of consumer patterns. That said, the level of detail within large-scale 30-minute sample data is intrinsically valuable and has been reflected in numerous research investigations [17], [18].

In this paper, we propose a novel approach of using gas smart meter data to improve the wellbeing of occupants in residential properties. Four key measures are observed from gas usage patterns as part of this approach, which are:

Tariff detection – Identifying whether a home is on the expected tariff based on their overall usage profile. For example, often consumers may not be aware that another tariff would be beneficial. A migration to a different tariff would help towards the reduction in energy poverty.

Temporally-aware tariff detection – Identifying whether a home is on the expected tariff, based on the timing of their energy usage. Such information may be used to inform occupants of cheaper alternative tariffs based on their time-of-day consumption habits. Unlike the above tariff-detection process, that is concerned with the full 24-hourt consumption pattern, this experiment factors in time of day in which the consumption took place. This provides a more granular analysis of the consumption patterns at different times of day.

Routine consumption detection – Identify routine patterns of energy usage. This may allow for the occupants to be advised on what changes to their energy usage behaviour could enable cost savings on a given tariff.

Age group detection - Establish the age categories of the occupants, in order to identify those who are at risk of energy poverty. As documented by Robinson et al., energy poverty is of highest risk amongst the elderly community. For that reason, a focus is on the detection of customers aged 65 and over.

To the best of our knowledge, this is the first time this has been attempted on the provided dataset and, thus, the first study of its kind. Other research projects in this area focus

[2]ec.europa.eu/energy/sites/ener/files/documents/1_EN_ACT_part1_v8.pdf

predominantly on the use of electricity data, as opposed to the gas usage dataset applied in this research. Many of such works are outlined in [4]. Additionally, an understanding of consumer load profiling of gas is fundamental for improving energy efficiency and working towards lower carbon emissions [19].

The remainder of this paper is organised as follows. Section II provides a background discussion on related work and the data used in this research. Section III outlines the methodology behind the research. Results are presented in Section IV and the paper is concluded in Section V.

II. BACKGROUND

The challenges vulnerable households face in the U.K. alone, results in upwards of 20,000 deaths each year due to household heating bills [1]. Yet, a growing amount of technology is available to help vulnerable households better manage their costs and their energy consumption at home and as an enabler to improve the wellbeing of vulnerable households. However, the technical challenge surrounding the solutions, means that many potential users are reluctant, unable or, in some circumstances, scared to make use of the technologies available [1]. Smart meters are part of the smart cities concept, are predominantly offer a more efficient ways to heat and light buildings [20]. To analyse smart meter data trends, there is an array of data classification techniques available.

Smart meter data is a time-series dataset, and as such the majority of investigations focus on techniques that are appropriate for time-series data analytics. Given that time-series data is comprised of discrete values, regression analysis is the preferred choice for data analysis processes. However, clustering techniques have also been used to generate notable results. In this section, related research works are presented.

A. RELATED WORK

Traditionally, knowledge of individual consumer behaviour patterns was not essential when planning load forecasting, as discussed by Gros [21]. This is the case particularly within the electricity management network. However, because of the increase in the use of decentralised power, through the introduction of the smart grid, load flow is now increasingly multi-directional. The traditional load curve models, which are comprised of a graph of energy/gas usage over time, are no longer appropriate methods for representing the load profiles from the data generated by smart meters. For that reason, Groß et al. adopt a linear regression approach for the parameterisation of stochastically-generated synthetic load profiles constructed using Markov chains. However, their approach focuses on the technique's application within the wider smart grid in order to compensate for deficiencies within the grid, rather than offering a wellbeing augmentation for the end-user.

Other research projects, including the study conducted by Robinson et al., outline the design of a system which demonstrates how intelligent technologies can be used for unobtrusive energy consumption management to support the elderly in particular [1]. The aim of their research is to alter the behaviour of the consumer to adopt more energy conscious behavioural patterns around the home, and in-turn, reduce their bills. This type of research is having an increasingly positive impact on household energy bills, and has resulted in many technological solutions available in the market place.

Whilst there is a significant amount of research within the electricity profiling and forecasting domain, there are relatively few projects concerning gas smart meter data. Focusing on gas meters specifically, Gupta et al. propose their own mathematical models for constructing gas load profiles from residential gas meters [19]. The aim of their research is to study average load levels for residential units, construct cost effective methods for monitoring systems and compare the electricity consumption against gas consumption. Their approach is based on data collected from a testbed. While the data is validated using a statistical method, the load profile data is based on estimates ascertained from the testbed experiments. Their approach also does not adopt a machine learning analysis of the data, but rather takes a statistical modelling approach to construct the user profiles.

Other approaches for load profiling adopt either a direct-clustering based or indirect-clustering approach. Within this area, research shows the 30-minute data sampling rate of smart meter data is reliable for most clustering approaches for load profiling [22]. Direct clustering refers to a clustering process, such as k-means, where the raw data is clustered without any prior data preparation. Whereas, indirect clustering applies other techniques prior to clustering, such as principal component analysis. For example, Benitez et al. apply a k-means clustering algorithm to generate a dynamic segmentation of daily load profiles as a representative sample of Spanish residential customers [23]. Their approach is able to detect seasonal effects on consumption patterns and their algorithm tends to group higher energy-consuming users into the same cluster. The benefit of their research is that it allows the observer to identify trends of user groups at a glance from a significant dataset. The approach successfully identified a change in consumer behaviour, resulting from a law change affecting the Spanish energy market.

The k-means clustering approach is also adopted by Khan [13] (whose research uses the same data source as that utilised in this paper).[3] In their research, the authors focus on forecasting rather than profiling and adopt an ensemble classification approach, using both the k-means clustering and a linear regression neural network. By converting the non-linear energy meter profiles into linear profiles, the authors are able to forecast consumer load. Their work differs to the research presented in this paper, in that the data used is electricity data rather than gas data. Also, the technique in this paper does not employ a k-means approach.

[3] C. for E. R. (CER), "ER Smart Metering Project - Gas Customer Behaviour Trial, 2009-2010."

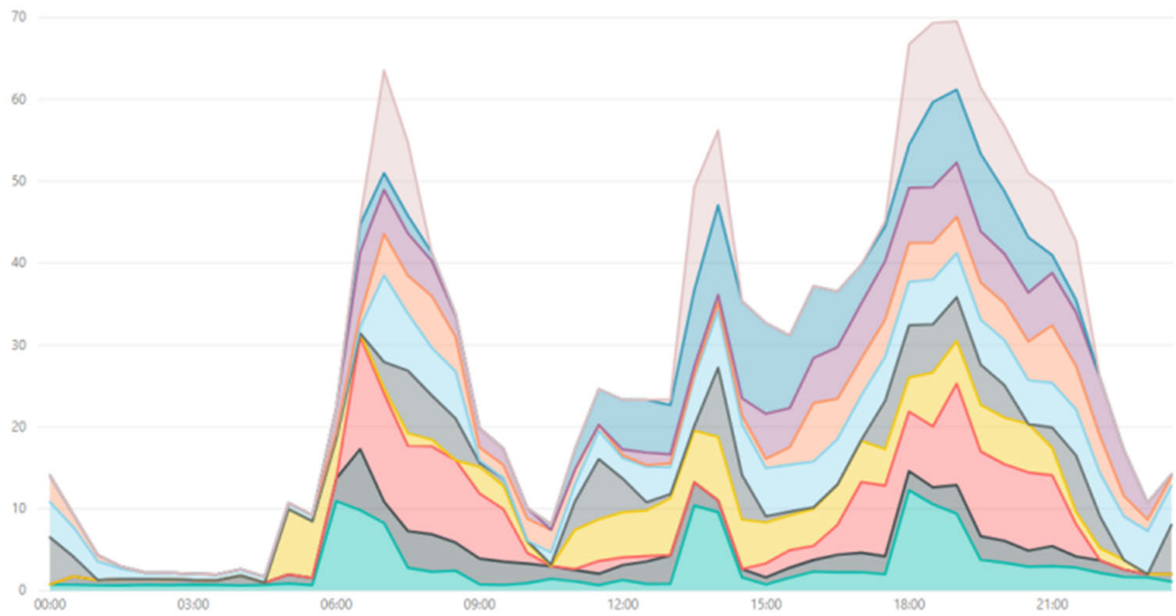

FIGURE 1. Stacked line plot of 10 random users.

B. GAS METER DATA: CASE STUDY

The data used in this research is comprised of 1,033 anonymised residential properties over an 18-month period between 2009 and 2011. The data is gas meter readings collected at a 30-minute sample rate.

Within the dataset, the users are divided into 4 different tariff groups, as detailed in Table 1.

TABLE 1. Tariff allocation.

Allocation Code	Allocation Interpretation	Count
1	Bi-monthly bill	257
2	Monthly bill	248
3	Bi-monthly bill + IHD	263
4	Bi-monthly bill + IHD + variable tariff	265

A sample of this data is presented in Figure 1, which displays a stacked line plot of 24-hours' worth of gas consumption for 10 users randomly selected from the data set. Clear trends in behaviour are reflected in the three peaks of high consumption periods in the morning, lunch time and evening. Each colour represents a single user.

A sample of the raw data is presented in Table 2. The date and timestamp (DT) is displayed in Julian's Day format, with 01 January 2009 as the starting point. As gas bills display usage in kilowatt hours (kWh), the usage is displayed as kW despite gas meters measuring cubic metres. One of the main differences between gas and electricity readings is that the gas data will have prolonged readings of zero values where no gas is used. For example, between time 33504 and 33508, the customer on tariff 2 has no gas consumption for a period of 2.5 hours but may well be active within the home.

TABLE 2. Data sample.

	Tariff 1		Tariff 2		Tariff 3		Tariff 4	
DT	ID	Usage	ID	Usage	ID	Usage	ID	Usage
33501	1000	0.894	1015	0.485	1016	0.000	1024	7.049
33502	1000	0.608	1015	0.453	1016	0.000.	1024	5.674
33503	1000	0.685	1015	0.111	1016	0.000	1024	0.000
33504	1000	0.817	1015	0.000	1016	1.801	1024	0.000
33505	1000	0.608	1015	0.000	1016	5.511	1024	0.000
33506	1000	0.850	1015	0.000	1016	2.397	1024	0.000
33507	1000	0.663	1015	0.000	1016	0.000	1024	0.000
33508	1000	0.607	1015	0.000	1016	4.185	1024	0.000
33509	1000	0.906	1015	3.344	1016	3.170	1024	0.000
33510	1000	0.607	1015	10.91	1016	5.721	1024	1.966

In the case of electricity data, the consumption may peak and drop but there is always a level of energy usage, due to electrical appliances in the house being on standby for example. Additionally, the smart meter itself requires energy consumption to function, so by default an energy reading will always be produced.

Within the dataset, in relation to Figure 1, it is possible to arrange a 24-hour time block into 4 separate periods of activity a) Morning, b) Afternoon, c) Evening and d) Night.

The visualisations shown in Figure 2 serve as a premise to hypothesise four periods of daily activity: morning, afternoon, evening and night. The graphs, which are based on 501,648 rows of data, show the full values for all customers as a sum, to show the overall trend for the time of day for a seven-day period.

C. DISCUSSION

Evidence of the benefits of a smart meter are documented in the findings of the report published by the ISSDA CER Smart Metering Project [24]. Over the period of 18 months,

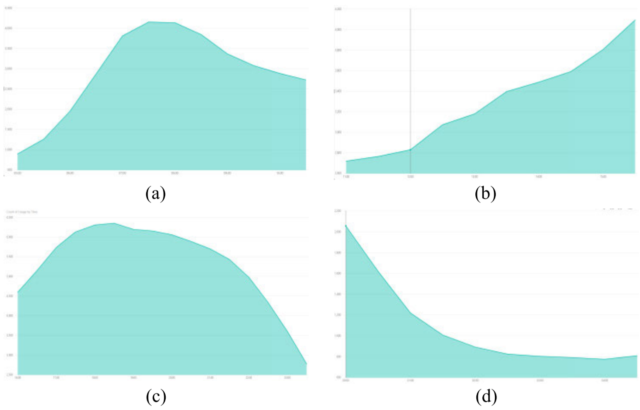

FIGURE 2. (a) All customers for morning period over 7 days. (b) All customers for afternoon period over 7 days. (c) All customers for evening period over 7 days. (d) All customers for night period over 7 days.

the consumption of gas drops on each of the four tariffs. This change is reflected in Table 3, which shows the difference in the consumption levels during the smart meter trial period.[4] Based on the statistics presented, it should be possible to detect a variation in the tariff types as the variation in consumption changes for each. None of the tariff options have the same change in consumption, however, tariffs 2 and 3 are the closest, but overall the standard deviation is 0.57373 between the four tariffs.

TABLE 3. Change in consumption over 18 months [25].

Tariff	1	2	3	4
Change over the trial	-2.2	-2.8	-2.9	-3.6

The saving produced by the tariffs displays a variability between the different user groups. The following section presents a methodology that can be applied to detect this subtle variation in consumption patterns between the customers on different tariffs.

III. METHODOLOGY

This research is timely due to i) an underlying switch in the technologies being used to monitor home gas and energy consumption; ii) the need for advanced data analytics to process, analyse and interpret the vast datasets generated by the smart metering infrastructure; iii) the growing need for remote profiling, for bespoke applications, such as health care monitoring [26], bad data detection [27], anomaly detection or load forecasting [28]; and iv) The growing trend for uncovering general information about a consumer using only their home energy readings [29]–[32]. Most research in this area makes use of electrical energy readings from smart meters. However, the focus of this paper is on gas meter data; making this research stand out from other related projects. Gas data analysis is often neglected from a machine learning point of view.

The contribution of this research involves four key observations: i) tariff detection; ii) temporally-aware tariff detection; iii) routine consumption detection and iv) age group detection, which are combined to produce an ensemble detection model.

A. ENSEMBLE DETECTION METHODOLOGY

Not all citizens have the capacity to make use of smart city services [33]. As outlined in [3], a typical use-case example would be an individual is living alone with arthritis (or other long term health condition) and on a low income. Often the support provided involves an enhanced installer visiting and providing the user with an IHD to support their energy management. However, no intelligent services are supplied with the device, and the ownness is still on the user to modify their home behaviour and fuel consumption. Therefore, an autonomous detection process is advantageous to support the wellbeing of vulnerable groups. The ensemble detection model to facilitate this is presented in Figure 3.

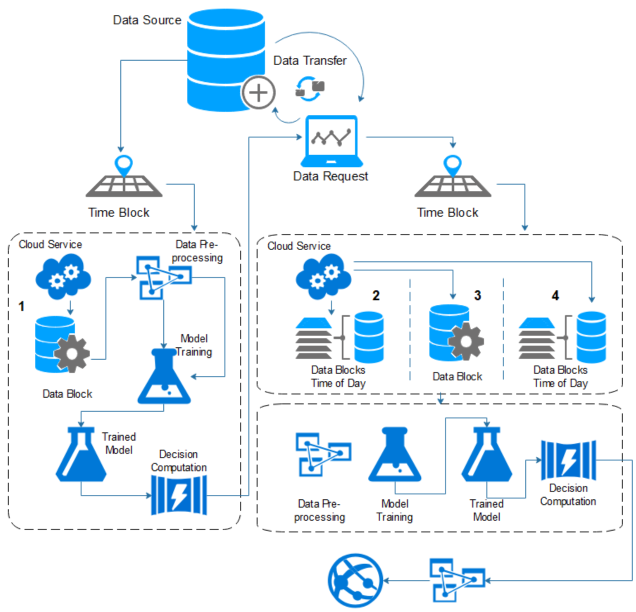

FIGURE 3. Ensemble methodology.

The model is a multi-stage process. 1) An age group detection process is conducted to detect whether the individual is in a 65+ age grouping. If a 70% AUC accuracy is achieved for a single classifier, 2) the next stage involves the detection of the routines of high home-activity. 3) The detection of the tariff band of the user is conducted both without factoring time of day and then also 4) with factoring time blocks (morning, afternoon, evening and night).

The multi-stage process is outlined as follows (in reverse order). The full 18-month dataset is used in the experiments; however, the entire dataset is not used in one go. Rather samples are selected from the overall dataset to make the data

[4]C. for E. Regulation, "Report On Smart Metering Technology Trials for Commission for Energy Regulation," 2011.

pre-processing requirements less intensive. This is done for three reasons, i) our initial experiments when using the entire dataset showed no improvement in the classification accuracy when more than 1-months' worth of data was employed; ii) using smaller samples of the dataset makes the experiments reproducible for other researchers without access to cloud analytics, and iii) the experiments are more realistic, that is, in a real-world setting there would not be access to such a large dataset but samples would be available in a real-time setting.

In each experiment, different classification algorithms are tested to find the optimal approach. The algorithms selected for the experiments include a boosted decision tree, decision forest, decision jungle, neural network, Support Vector Machine (SVM) and Bayes point machine. Each are outlined as follows.

1. Boosted decision tree is ideal for an accurate prediction as it employs an ensemble learning method. By using this approach, each newly formed tree corrects for the errors of the first tree. Decision trees are able to capture non-linear data.
2. Decision forest, which is an ensemble learning approach with bootstrap aggregating applied, where each new tree is grown from a new random sample from the dataset. Outputs from the classification are achieved by voting, where outputs of the models are aggregated.
3. Decision jungles build on the decision forest approach; however, they integrate an ensemble of decision Directed Acyclic Graphs (DAGs) which allows tree branches to merge.
4. Neural networks function by employing a set of inter-connected layers. When there is an input into the first layer, a connection to an output layer is facilitated through use of an acyclic graph. This graph is typically comprised of weighted edges and nodes to form a decision [34].
5. SVMs are commonly used as a benchmark in machine learning experiments [35] due to their flexibility, simplicity and tendency to perform well under simple classification tasks. Its prediction is based on two possible outcomes where it recognises patterns in a multi-dimensional feature space called the hyperplane.
6. Bayes Point Machine uses a Bayesian method. However, it is based on a linear classification approach. One advantage of this technique is that it is not prone to overfitting to the training data. In our experiments, training iterations are set to 30, which is the recommended value for accuracy [36].

B. EXPERIMENT 1–TARIFF DETECTION

One month's worth of gas meter readings is analysed, which totals to 1,302,336 rows of raw data with the class labels. This experiment serves as a benchmark test of the machine learning approach to see if the detection of variability in the dataset is, in-fact, possible. The data used for the experiment is taken from the latter part of the dataset, as the variation should be stronger due to the customers adapting to their tariff. The process employs a direct classification approach. In other words, only the raw data is used for the classification and no features or data transformation are applied to the dataset.

Given the nature of a cloud processing platform, the classifiers can be run simultaneously.

The first stage of the experiment employs a direct classification approach, where the raw dataset is classified using a one tariff vs all approach. This serves as a standard experiment for comparison with more advanced techniques later in the research. The second phase involves extracting features from the dataset to adopt an in-direct classification. Statistical features including maximum and minimum values, mean, median and standard deviation of the d-dimensions, variance, skewness and kurtosis of the d-dimensions.

The features are calculated at two-hour time blocks. This is due to the selection of skewness and kurtosis as features, as both require minimum three values as input. This approach is further outlined in our previous work [7]. Variance is calculated using (1) where $\bar{X}$ is the sample mean, and n is the sample size [37].

$$\sigma^2 = \frac{\sum (x-\bar{x})^2}{(n-1)} \tag{1}$$

Similarly, the standard deviation calculation takes x for the sample mean and n is the sample size, as displayed in (2) [37].

$$\sigma = \sqrt{\frac{\sum (x-\bar{x})^2}{(n-1)}} \tag{2}$$

The calculation for skewness (S) is outlined in (3), where s is the sample standard deviation and x is the mean value [38].

$$S = \frac{n}{(n-1)(n-2)} \sum_{j=1}^{n} (\frac{x_j - \bar{x}}{s})^3 \tag{3}$$

Likewise, kurtosis, which is a measure of outliers [39], also uses the standard deviation (s) and is calculated in (4) [38].

$$\left\{ \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum_{j=1}^{n} (\frac{x_j-\bar{x}}{s})^4 \right\} - \frac{3(n-1)}{(n-2)(n-3)} \tag{4}$$

The inclusion of the features adds a cleaning stage to the methodology to account for any missing values. Rows with missing values are removed prior to the normalisation. Synthetic Minority Oversampling Technique (SMOTE) is then used to compensate for the missing values and the imbalance in the dataset. SMOTE employs a statistical approach for ensuring a balance in a dataset, by generating new instances from existing minority cases [40]. The advantage of SMOTE is that new instances are not duplicated from existing minority cases. Rather, the algorithm is able to take samples of the feature-space for each target class. It also calculates the nearest neighbours in the feature-space and uses this information to generate new examples that combine features of the target case with features of its neighbours.

FIGURE 4. **Experiment 1 (a) Min-max scaling, (b) Z-score normalisation.**

Prior to splitting the data for classification, the values in the dataset are normalised using sliding Z-score, as displayed in Figure 4b. This is calculated using (5):

$$Z = \frac{x - \bar{x}}{Std(x)} \tag{5}$$

Z-score normalisation is appropriate in this case, as it ensures that the raw data conforms to a common scale for the classification.

Sliding Z-score is used in each of the experiments. Min-max scaler, displayed in Figure 5a, is also considered as a normalisation approach and is calculated as outlined in (6). However, the values generated by the min-max scaling results in a lower standard deviation, which supresses the effect of outliers [41] and produced a lower classification accuracy during the initial experimentation.

$$MM\left(x_{ij}\right) = \frac{x_{ij} - x_{min}}{x_{max} - x_{min}} \tag{6}$$

C. EXPERIMENT 2–TEMPORALLY-AWARE TARIFF DETECTION

The second factor detects behavioural differences in the four different time blocks mentioned previously (morning, afternoon, evening and night), in order to detect the tariff. Figure 5 presents the positive and negative correlation between the features, within the time blocks.

The experiments are conducted with a reduced dataset of 501,648 rows (7 days' worth of data) but with the division of the data into their corresponding time blocks. As before, statistical features are extracted from the dataset for the classification. Given the mean values in the features, the scatter matrix is the m-by-m positive semi-definite matrix. Where T denotes matrix transpose, μ is the sample mean and multiplication is with regards to the outer product [40], as expressed in (7).

$$\begin{aligned} S_m &= \sum_{i=1}^{m}(x_i-\mu)(x_i-\mu)^T = \sum_{i=1}^{m}(x_i-\mu)\otimes(x_i-\mu)^T \\ &= \left(\sum_{i=1}^{m} x_i x_i^T\right) - m\mu\mu^T \end{aligned} \tag{7}$$

In this case, from the visual inspection, the features predominantly have a positive correlation. A positive correlation is denoted by a progressive incline in the data points; for example, when the general pattern of the data points within a square is from bottom left to top right. An example of this would be Min to Max or Min to Mean. A negative correlation is a slope in the data points from top left to bottom right; for example, Min to Variance and Min to Standard deviation. Further to this, Figure 6 displays a correlation between skewness (x-axis) and kurtosis (y-axis).

The data presented is over a 7-day period for the afternoon period only, for all tariffs.

Figure 7 displays a stacked line plot of all the features over a 24-hour period. The difference between variance and skewness demonstrates why the choices of features are ideal for supporting the classification.

Theoretically, based on Figure 7, for some of the time blocks, it should be easier for the classifiers to separate the tariffs from each other. For example, in the afternoon and evening periods where there is a high gas consumption, there is also high variation in the consumption patterns.

D. EXPERIMENT 3–ROUTINE CONSUMPTION DETECTION

As Figures 1 and 2 displayed in the background section, there are clear trends and differences in the consumption patterns at certain times of day. In this measure, these four time periods are added in as class labels. A random sample of the dataset is displayed in Table 4.

TABLE 4. **Tariff allocation.**

ID	Date	Usage	Time	Tariff	Time Block
1000	02/12/2009	0.606969	04:00:00	1	Night
1000	02/12/2009	9.398459	06:30:00	1	Morning
1000	02/12/2009	10.30438	14:00:00	1	Afternoon
1000	02/12/2009	1.777624	23:00:00	1	Evening

The aim of the experiment is to demonstrate that, it is possible to identify different times of day based on consumption

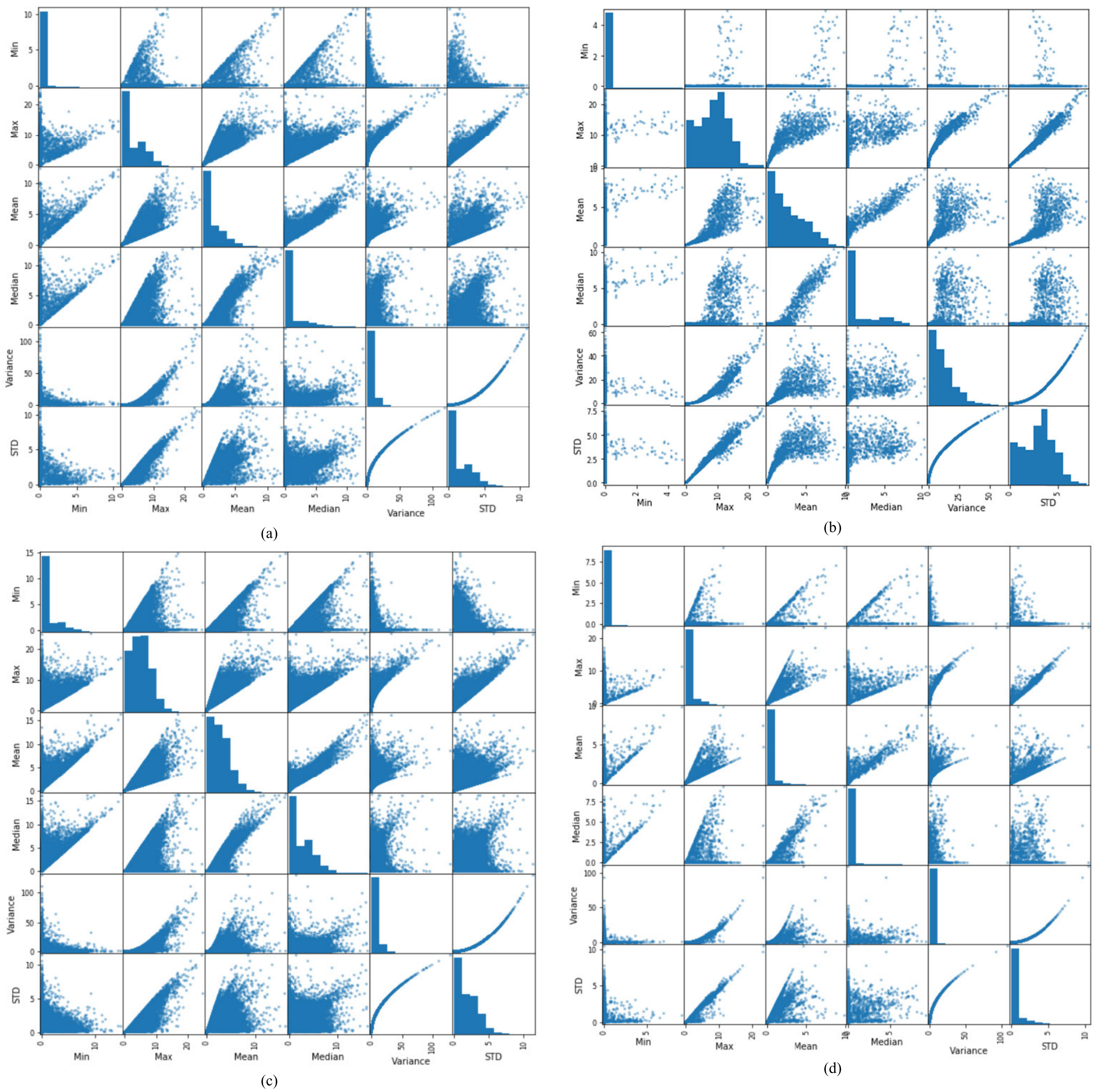

FIGURE 5. Experiment 2 scatter matrix (a) Morning, (b) Afternoon, (c) Evening and (d) Night.

patterns. In total, this section is comprised of multiple smaller experiments. Initially, a benchmark experiment is conducted using a multiclass approach. A multiclass decision forest and multiclass decision jungle allow for the classification of all four time-periods in the same experiment. Next, a detection of the individual time blocks is conducted. This process is comprised of four experiments, 1) Morning vs Afternoon, Evening and Night; 2) Afternoon vs Morning, Evening and Night; 3) Evening vs Morning, Afternoon and Night and 4) Night vs Morning, Afternoon and Evening. The results from this experiment are presented in Section IV-B.

E. EXPERIMENT 4–AGE GROUP DETECTION

For this final observation factor, the focus is on the identification of the over 65's grouping. The premise and benefits of this work is outlined in our previous research [7]. For future applications of this research, this process will identify social clusters for health care cluster mapping.

In this experiment, a dimensionality reduction process, using Principal Component Analysis (PCA), reduces the features from eight to four. Again, our previous work has demonstrated an improved detection level when PCA is introduced within this classification methodology. The four newly

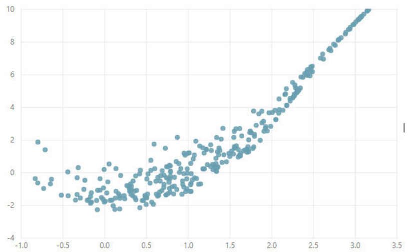

FIGURE 6. **Experiment 2 Skewness vs Kurtosis.**

generated columns contain an approximation of the feature space of the 8 original features. Figure 8 displays scatter plot visualisations of the four newly generated features.

The PCA-generated features are split into a training and a test set. The classification is scored using the split data as a validation.

IV. EXPERIMENTS AND RESULTS

Each of the classifiers' performance is calculated using a confusion matrix to assess the success of the classification or Area Under the Curve (AUC) and error. The AUC measures the entire two-dimensional area underneath a Receiver Operating Characteristic (ROC) curve. The ROC curve displays the true positive against the false positive predictions. AUC has been used instead of another measure (e.g. f1 score) as AUC assesses the whole range of thresholds rather than a specific one as measured by f1 score. Therefore, this produces a more holistic perspective on the classifier performance is. AUC measures the probability that test values from a randomly selected pair of binary class samples are correctly ranked and is thus a convenient global measure for the quantification of classification accuracy.

A. RESULTS OF EXPERIMENT 1

Experiment 1 results are divided in to two parts, i) the raw data (direct) classification and ii) the in-direct classification.

1) RAW DATA WITH DIRECT CLASSIFICATION

In this section, a one tariff vs all tariffs classification is employed. In this case, tariff 4 is selected for the one vs all, as this is the tariff, which displayed the greatest variation compared throughout the dataset. The benchmark experiment serves as a comparison between direct and indirect classification and a justification for the choice of in-direct classification in the subsequent experiments. All six classifiers are evaluated. The results are presented in Table 5.

The Boosted decision tree and the decision jungle achieved the highest AUC accuracy scoring 51.9%; which is a low scoring classification. However, all classifiers produced a relatively low score. Figure 9 displays the precision (y-axis) against the recall score (x-axis) for each of the classification experiments between values 0 to 1.

It is clear from the benchmark experiment that an indirect classification approach is needed to increase the accuracy of the prediction for all classifiers.

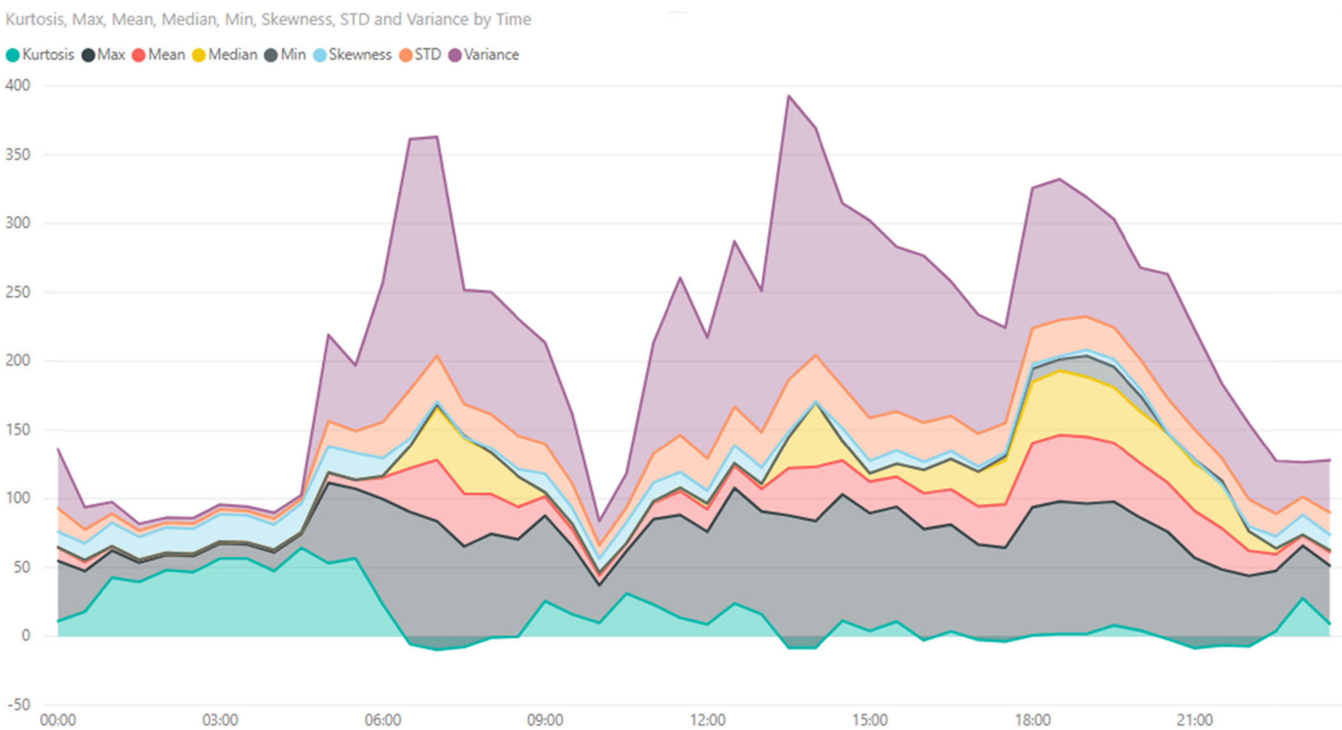

FIGURE 7. **Experiment 2 line plot of all feature values.**

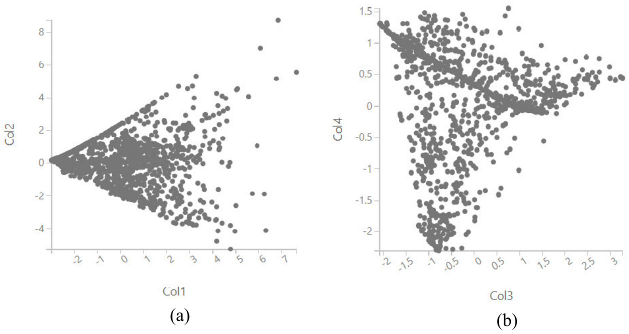

FIGURE 8. PCA features 1 and 2(a), PCA Features 3 and 4(b).

TABLE 5. Tariff benchmark classification results.

Classifier	Accuracy	Precision	Recall	F1 Score	AUC
Decision Tree	0.740	0.558	0.014	0.027	0.519
Decision Forest	0.732	0.364	0.040	0.072	0.516
Decision Jungle	0.740	0.599	0.011	0.021	0.519
Neural Network	0.739	1.000	0.000	0.000	0.500
SVM	0.739	1.000	0.000	0.000	0.507
Bayes PM	0.739	1.000	0.000	0.000	0.507

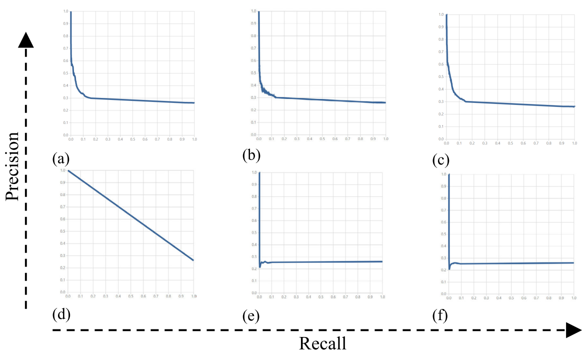

FIGURE 9. Precision vs Recall for benchmark test. (a) Decision tree, (b) Decision forest, (c) Decision jungle, (d) Neural network, (e) SVM, (f) Bayes binary machine.

TABLE 6. Two-class boosted decision tree results.

Statistics	Value	Classification	Score
Mean	0.572	Accuracy	0.635
Median	0.613	Precision	0.626
Min	0.001	Recall	0.783
Max	1.000	F1 Score	0.696
STD	0.249	AUC	0.690

2) INDIRECT CLASSIFICATION

In this section, the results for each classifier are presented individually. Initially, the two-class decision tree demonstrates a remarkable improvement with a 69% AUC success rate, as outlined in Table 6. This is calculated from the ROC curve displayed in Figure 10a.

Similarly, the decision forest achieved a higher accuracy of 77.4%, compared to 51.6% scored using the direct classification approach. The full results for the decision forest are outlined in Table 7.

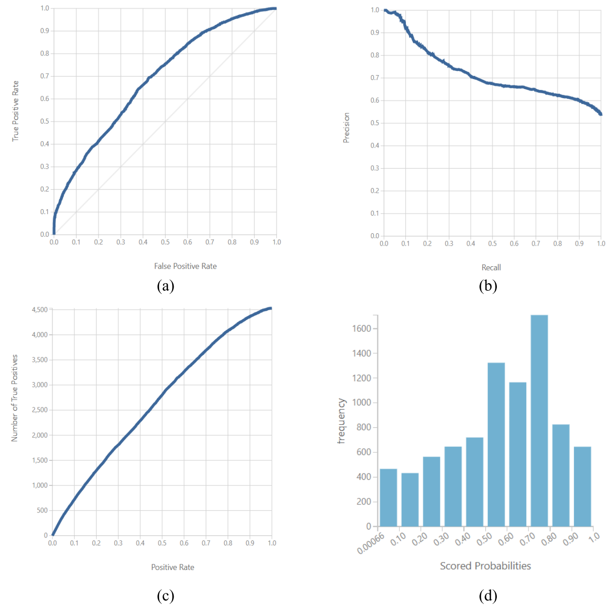

FIGURE 10. Boosted decision tree plots. (a) ROC curve, (b) Precision/Recall, (c) Lift and (d) Scored probabilities.

TABLE 7. Two-class decision forest results.

Statistics	Value	Classification	Score
Mean	0.534	Accuracy	0.705
Median	0.543	Precision	0.726
Min	0.000	Recall	0.717
Max	1.000	F1 Score	0.721
STD	0.297	AUC	0.774

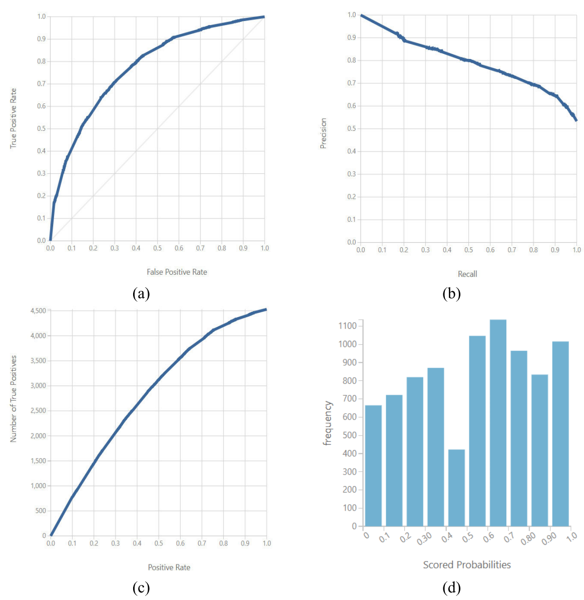

FIGURE 11. Decision forest plots. (a) ROC curve, (b) Precision/Recall, (c) Lift and (d) Scored probabilities.

The decision tree classification is the highest performing classifier for the second experiment. On visual inspection, the scored probabilities displayed in Figure 11(d) are superior, when compared with the other classifiers.

TABLE 8. Two-class decision jungle results.

Statistics	Value	Classification	Score
Mean	0.534	Accuracy	0.608
Median	0.530	Precision	0.608
Min	0.000	Recall	0.743
Max	1.000	F1 Score	0.669
STD	0.124	AUC	0.661

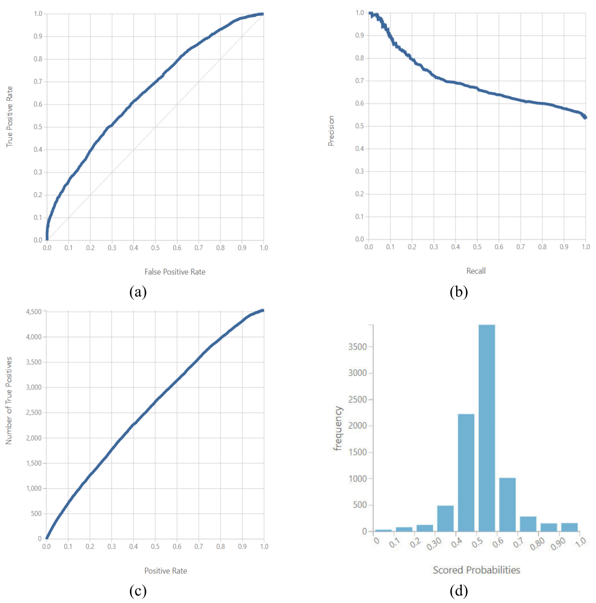

FIGURE 12. Decision jungle plots. (a) ROC curve, (b) Precision/Recall, (c) Lift and (d) Scored probabilities.

TABLE 9. Neural network, SVM and Bayes PM results.

Classifier	Accuracy	Precision	Recall	F1 Score	AUC
Neural Network	0.538	0.538	0.939	0.684	0.525
SVM	0.533	0.533	1.000	0.695	0.498
Bayes PM	0.532	0.535	0.926	0.678	0.516

Despite the advantages offered by the decision jungle, the results are lower than that of both the boosted decision tree and the decision forest. The decision jungle is able to perform with 66.1% accuracy, as documented in Table 8; with the results visualised in the plots displayed in Figure 12.

The two-class neural network and Bayes point machine classifiers again scored similar results to the benchmark test, with the SVM actually declining in accuracy by 9%. Each is considerably less effective than the decision tree approaches. The classification results are detailed in Table 9, which presents the accuracy, precision, recall F1 and AUC scores for each. As before, Figure 13 displays the precision against the recall score for each of the classification experiments and the scored probabilities as histograms.

B. RESULTS OF EXPERIMENT 2

As demonstrated in Experiment 1, an in-direct classification process generates higher results. Therefore, in this

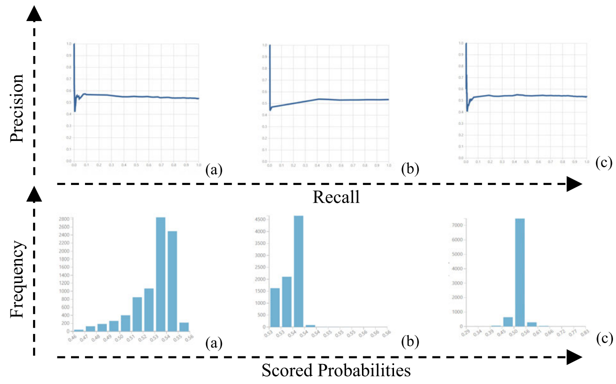

FIGURE 13. Precision vs Recall for In-Direct classification. (a) Neural network, (b) SVM, (c) Bayes point machine.

section only an in-direct approach is used. As previously, a one-vs-all approach is adopted, this makes it a two-class classification process. Again, tariff 4 is selected, for the one vs all test, so that the results can be compared with experiment one.

The full results for experiment two are presented in three tables; Tables 10 and 11 display the classification for the decision tree, decision forest and decision jungle. Table 12 displays the results for the Neural Network, SVM and Bayes PM classifiers. During the morning period, the decision forest is the highest scoring classifier and is able to separate the data with 78.7% accuracy and the boosted decision tree is able to perform with 72.8% accuracy. With an overall classification AUC mean of 72.57% the decision tree approaches, offer a higher success rate than the three other techniques, which score 50.83% as mean average.

Tariff detection in the afternoon, again demonstrates the highest success rate when using a decision tree approach, which have a mean average of 75.2% classification accuracy, with the decision forest scoring the highest with 78.8% accuracy. The evening results are again comparable, scoring a 72.56% mean accuracy. However, the evening mean accuracy drops to 69.16%, yet in this case the boosted decision tree approach is able to detect with a 79.0% accuracy to maintain the high accuracy rate. Throughout the afternoon, evening and night period, the Neural Network, SVM and Bayes PM have a 51.58% mean classification score. However, for the afternoon period, the neural network is able to achieve a high 72.9% successful classification score. The results presented in Tables 10 to 12 are visualised in Figures 14 and 15.

As the visualisation in Figure 14 demonstrates, there is a consistent trend in the classification accuracy of the decision tree, decision forest and decision jungle approaches. The scored probabilities exhibit a similar overall distribution for the morning, afternoon and evening time blocks. However, the night time block exhibits the greatest variation, particularly relating to the decision jungle results. In Figure 15, the three other classifiers are evaluated. As the trend demonstrates in the precision plots, the results are inconsistent and

TABLE 10. Two class boosted decision tree (DT), decision forest (DF) and decision jungle (DJ) statistics.

	Morning			Afternoon			Evening			Night		
Statistics	DT	DF	DJ	DT	DF	DJ	DT	DF	DJ	DT	DF	DJ
Mean	0.4386	0.4692	0.4822	0.4669	0.4595	0.4755	0.4386	0.4692	0.4822	0.4555	0.4572	0.4550
Median	0.4347	0.5000	0.4996	0.4419	0.6650	0.4794	0.4347	0.5000	0.5008	0.3916	0.4454	0.4392
Min	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1677
Max	0.9999	1.0000	0.8226	1.0000	1.000	0.9891	0.9999	1.0000	0.8320	1.0000	0.9076	0.6588
STD	0.3029	0.2811	0.1055	0.335	0.3151	0.1402	0.3029	0.2811	0.1057	0.3549	0.1202	0.0506

TABLE 11. Two class boosted decision tree (DT), decision forest (DF) and decision jungle (DJ) classification.

	Morning			Afternoon			Evening			Night		
Classification	DT	DF	DJ	DT	DF	DJ	DT	DF	DJ	DT	DF	DJ
Accuracy	0.670	0.714	0.612	0.690	0.732	0.641	0.670	0.714	0.610	0.708	0.641	0.571
Precision	0.667	0.734	0.591	0.662	0.738	0.617	0.667	0.734	0.588	0.694	0.703	0.616
Recall	0.622	0.633	0.615	0.656	0.642	0.566	0.622	0.633	0.621	0.649	0.371	0.168
F1 Score	0.643	0.679	0.603	0.659	0.687	0.590	0.643	0.679	0.604	0.670	0.486	0.263
AUC	0.728	0.787	0.662	0.765	0.788	0.703	0.728	0.787	0.662	0.790	0.709	0.576

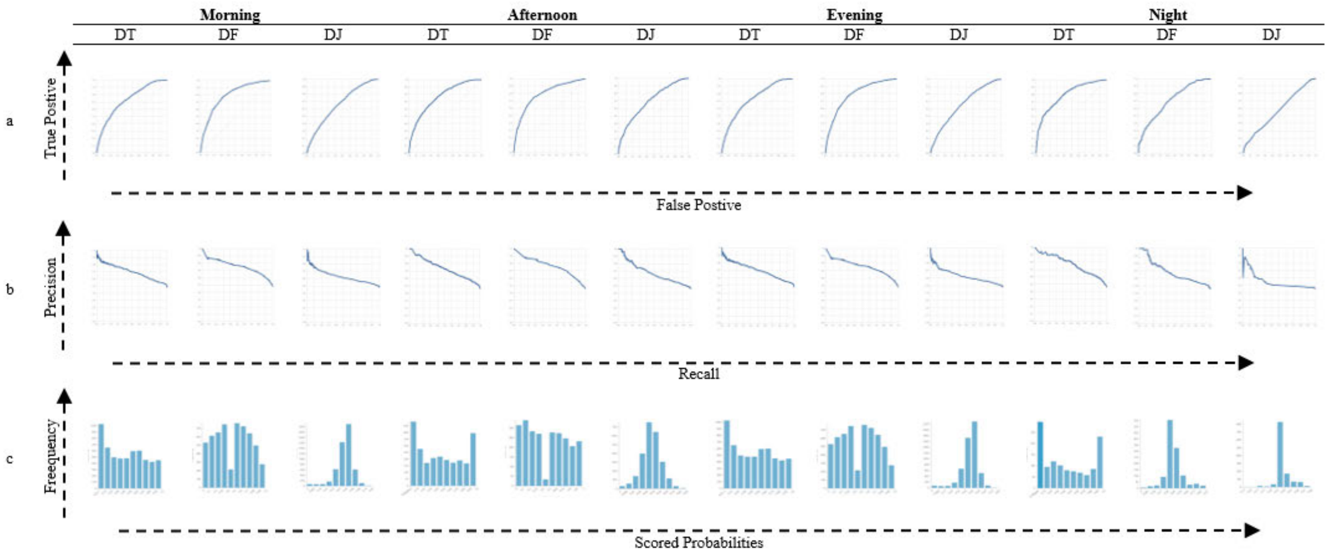

FIGURE 14. Data Trends for (a) ROC Curve, (b) Precision and (c) Scored probabilities for Decision Tree (DT, Decision Forest (DF) and Decision Jungle (DJ).

TABLE 12. Classification neural network, SVM and Bayes PM results.

Statistics	Morning			Afternoon			Evening			Night		
	NN	SVM	BPM	NN	SVM	BPM	NN	SVM	BPM	NN	SVM	BPM
Accuracy	0.521	0.522	0.523	0.546	0.539	0.543	0.522	0.522	0.523	0.667	0.564	0.586
Precision	0.469	1.000	0.504	0.581	0.485	0.501	0.588	1.000	0.504	0.796	0.531	0.640
Recall	0.006	0.000	0.210	0.022	0.157	0.214	0.004	0.000	0.210	0.366	0.400	0.218
F1 Score	0.011	0.000	0.297	0.042	0.237	0.300	0.007	0.000	0.297	0.501	0.456	0.325
AUC	0.512	0.503	0.510	0.538	0.525	0.529	0.517	0.503	0.510	0.729	0.560	0.691

do not register highly over 50% for the morning, afternoon and evening periods. However, for the night time block, the results are higher, with the neural network achieving a 72.9% accuracy, the SVM achieving a 56% AUC success rate and the Bayes point machine able to classify with a 69.1% accuracy score.

C. RESULTS OF EXPERIMENT 3

The third experiment begins with a benchmark multiclass classification. The aim is to detect the different times of day based solely on the gas consumptions readings. Figure 16 displays the scored probabilities of the multiclass experiment, with the trend line. Figure 17 displays the confusion matrix of

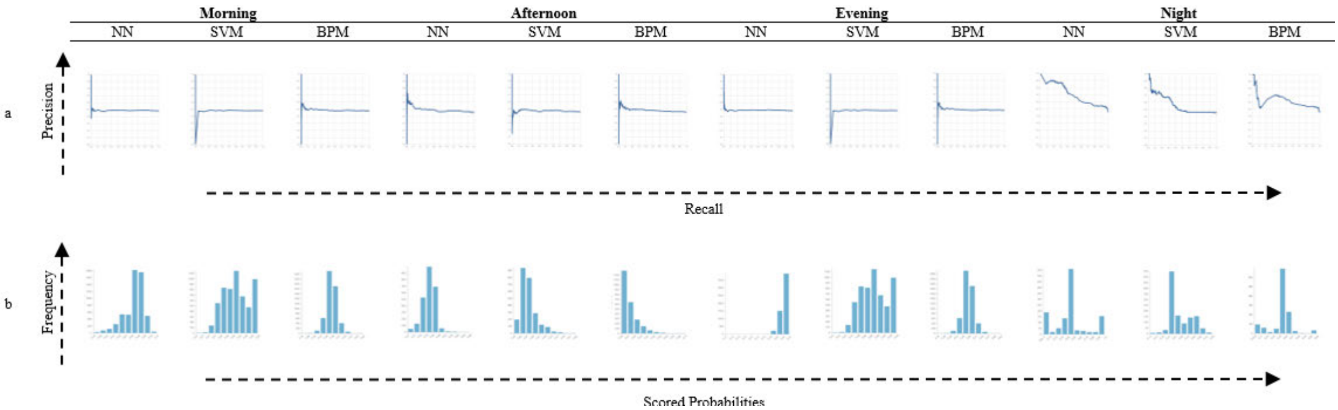

FIGURE 15. Data trends for neural network, SVM and Bayes point machine results visualisation, (a) Precision and (b) Scored probabilities.

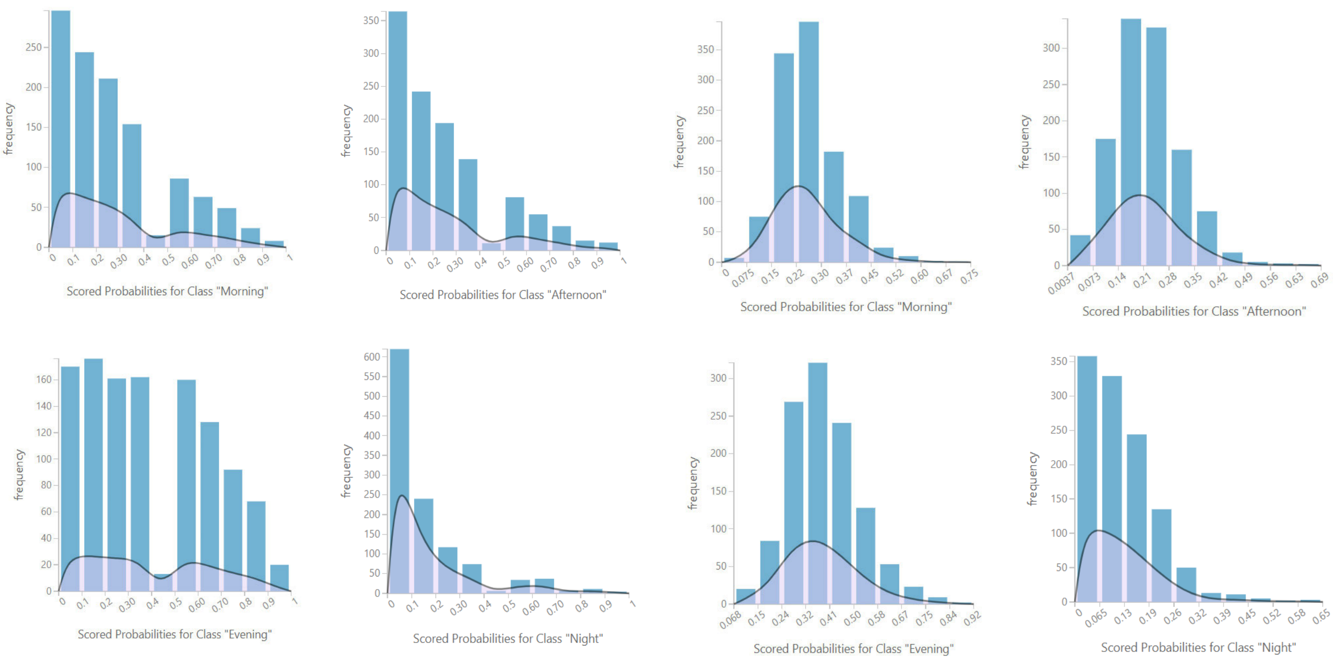

FIGURE 16. Multiclass forest vs multiclass jungle scored probabilities.

TABLE 13. Multiclass forest vs multiclass jungle results.

Statistics	Multiclass Forest Score	Multiclass Jungle Score
Overall Accuracy	0.324	0.355
Average Accuracy	0.662	0.677
Micro-Averaged Precision	0.324	0.355
Macro-Averaged Precision	0.289	0.292
Micro-Averaged Recall	0.324	0.355
Macro-Averaged Recall	0.283	0.276

the results. For the most part, the overall classification process results in low scores and the multiclass approach struggles to separate the different times of day from each other when the information is provided all at once. The results are outlined in Table 13.

As demonstrated in the confusion matrix plots in Figure 17, the majority of the time blocks are challenging to separate. However, for both the multiclass forest and multiclass jungle, the separation of the evening time block scored the highest accuracy. This may be due to the fact that, as discussed

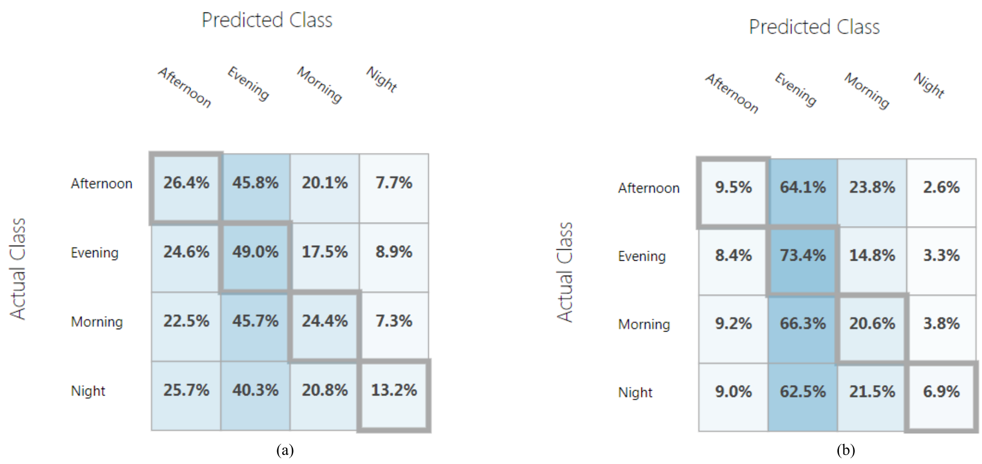

FIGURE 17. Multiclass forest (a) vs multiclass jungle (b) confusion matrices.

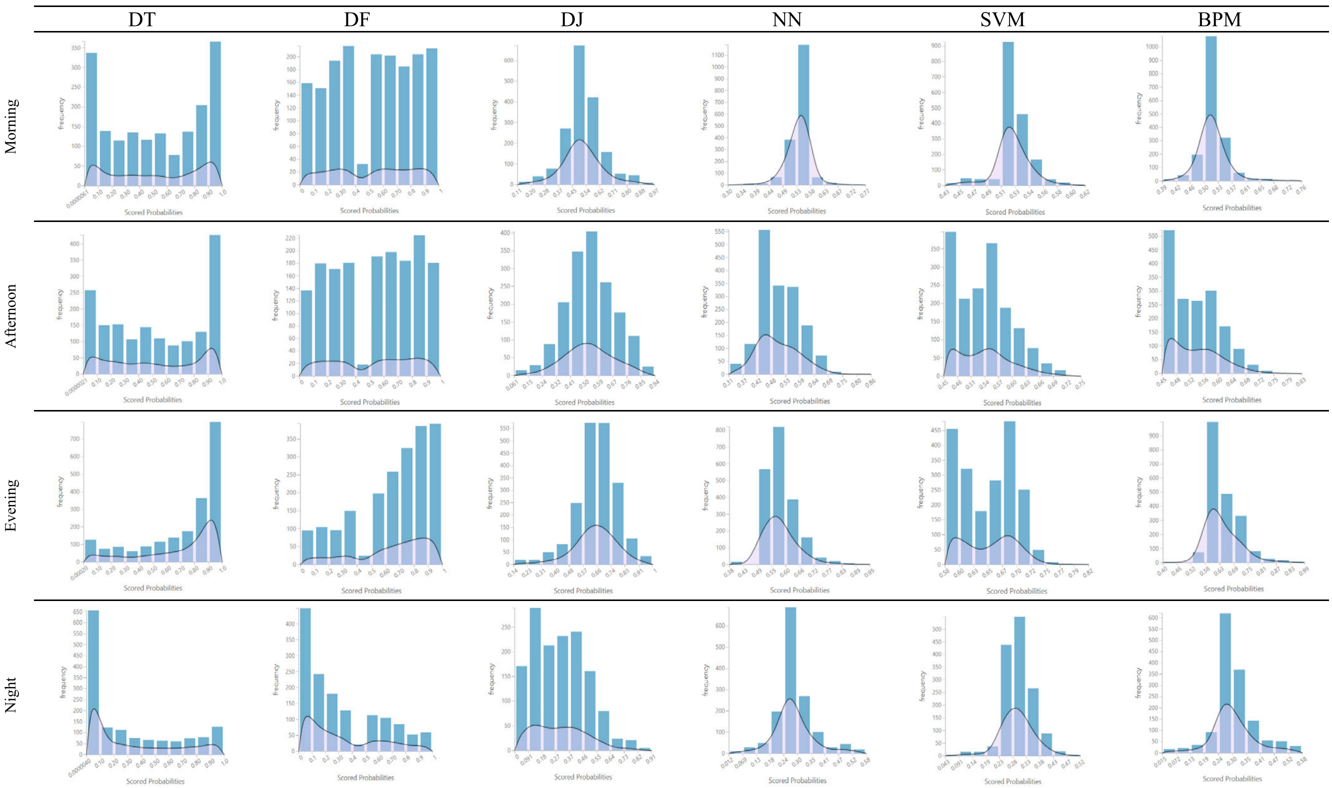

FIGURE 18. Data trends for the results of the scored probabilities for boosted decision tree, decision forest and decision jungle.

in the background section, this time of day contains the highest use of gas consumption and may therefore produce the highest variation in consumption compared to the other times. The multiclass jungle is able to outperform the multiclass forest and is able to detect the evening time block with a 73.4% accuracy however, each of the others produce a low score. For example, the night time block achieves a 2.6% accuracy which is a lower probability than randomly guessing the time block the data value belongs to. Therefore, to improve the quality of the classification process, a one vs all classification is once again adopted for the remainder of the experiment. Individual time blocks are extracted and compared with all of the others for the detection process. By employing a two-class approach, the results are improved significantly. The boosted decision tree, decision forest and decision jungle perform with the highest accuracy for each of the time periods. The trends in the data results are presented in Figures 18, 19 and 20.

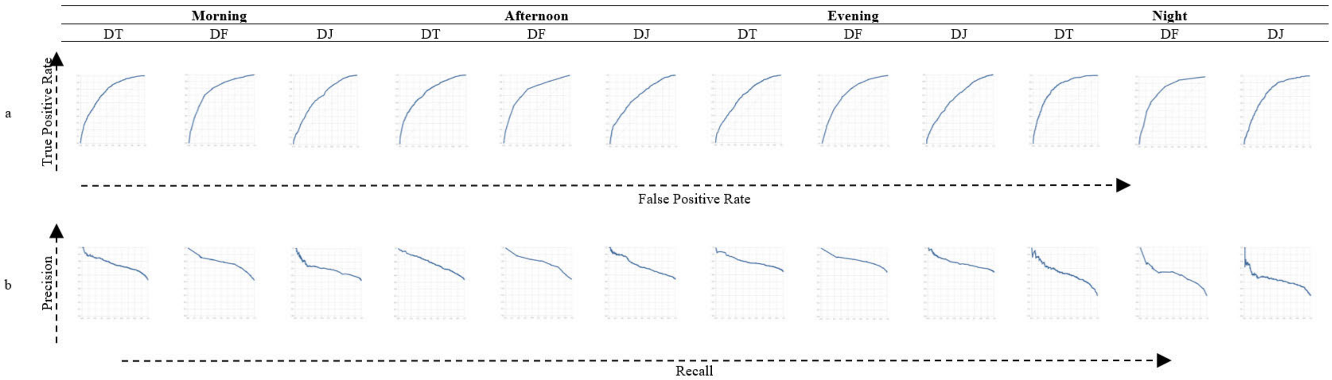

FIGURE 19. Data Trends for ROC Curve (a) and Precision plot (b) for boosted decision tree, decision forest and decision jungle.

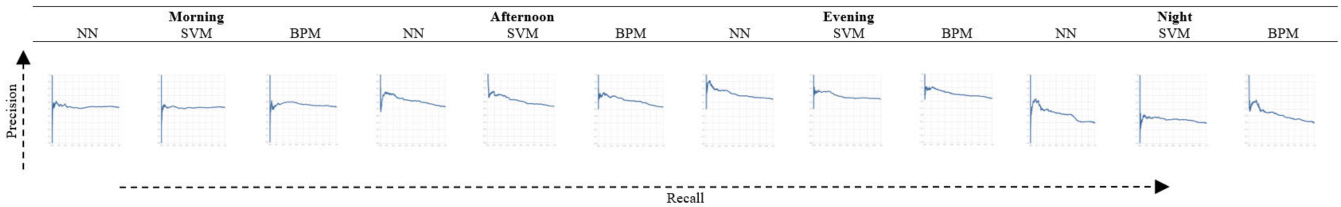

FIGURE 20. Data trends for the precision plots for neural network, SVM and Bayes point machine.

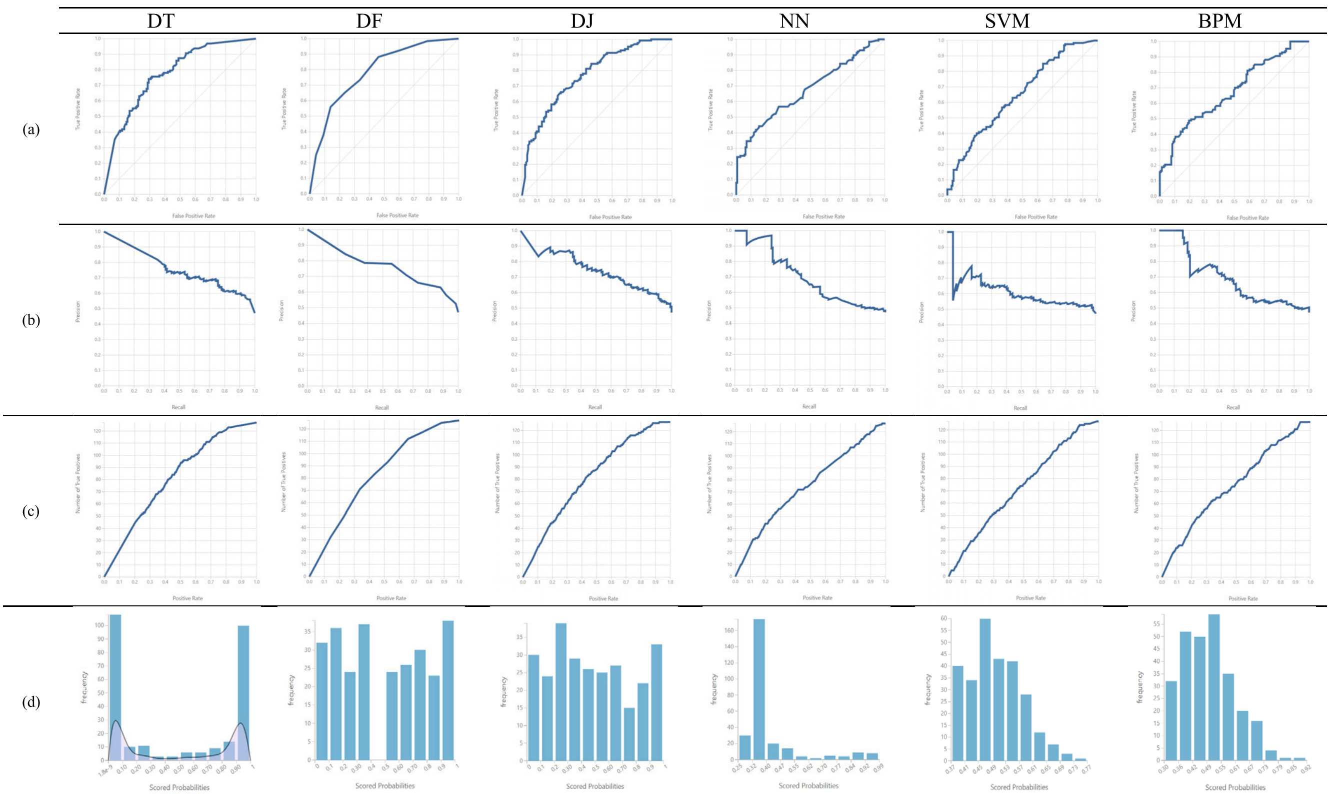

FIGURE 21. Data trends for (a) ROC curve, (b) precision, (c) lift and (d) scored probabilities.

The decision tree is able to detect with an overall mean accuracy of 75.8% across the four different blocks of time, scoring 81.5% at highest for the night time block classification. The decision forest performed higher with a 77.1%

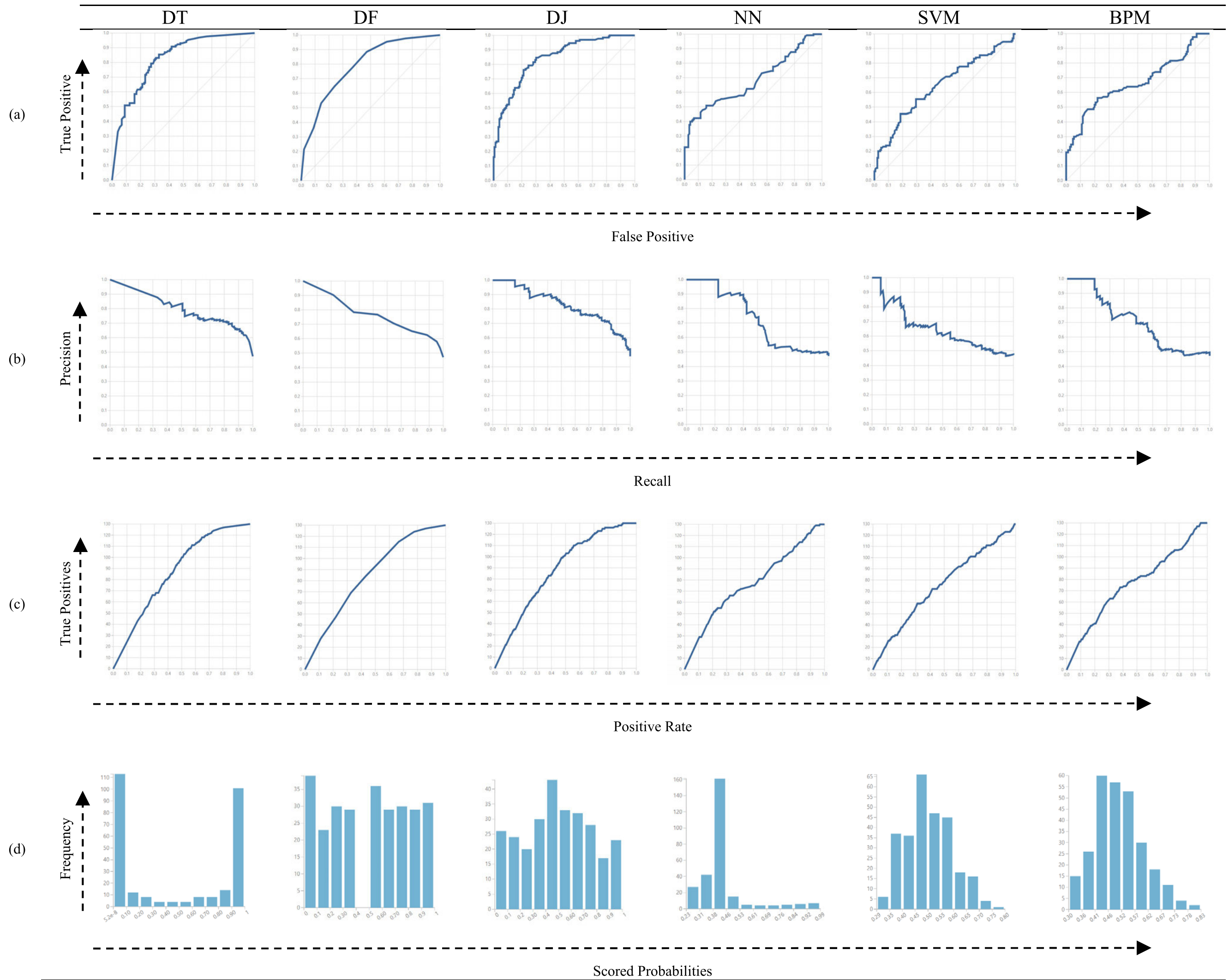

FIGURE 22. Data trends for the (a) ROC curve, (b) precision, (c) lift and (d) scored probabilities.

TABLE 14. Two class boosted decision tree (DT), decision forest (DF) and decision jungle (DJ) statistics.

Statistics	Morning			Afternoon			Evening			Night		
	DT	DF	DJ	DT	DF	DJ	DT	DF	DJ	DT	DF	DJ
Mean	0.5203	0.5250	0.5222	0.5331	0.5272	0.5330	0.7202	0.6604	0.6531	0.3032	0.2971	0.3021
Median	0.5370	0.5000	0.5157	0.5171	0.5000	0.5277	0.8447	0.7500	0.6607	0.1446	0.2500	0.2920
Min	0.0000	0.0000	0.1139	0.0000	0.0000	0.0615	0.0002	0.0000	0.1405	0.0000	0.0000	0.0000
Max	1.0000	1.0000	0.9655	1.0000	1.0000	0.9362	1.0000	1.0000	1.0000	0.9997	1.0000	0.9143
STD	0.3516	0.3078	0.1173	0.3497	0.3122	0.1522	0.2952	0.2874	0.1285	0.3370	0.2981	0.1777

TABLE 15. Two class boosted decision tree (DT), decision forest (DF) and decision jungle (DJ) classification.

Classification	Morning			Afternoon			Evening			Night		
	DT	DF	DJ	DT	DF	DJ	DT	DF	DJ	DT	DF	DJ
Accuracy	0.691	0.729	0.634	0.678	0.707	0.638	0.691	0.719	0.678	0.766	0.770	0.727
Precision	0.704	0.759	0.640	0.706	0.748	0.649	0.714	0.763	0.679	0.606	0.640	0.571
Recall	0.703	0.705	0.683	0.679	0.681	0.701	0.869	0.819	0.950	0.581	0.493	0.287
F1 Score	0.704	0.731	0.661	0.692	0.713	0.674	0.784	0.790	0.792	0.593	0.557	0.382
AUC	0.762	0.781	0.694	0.758	0.765	0.704	0.697	0.736	0.672	0.815	0.802	0.761

TABLE 16. neural network, SVM and Bayes PM results.

Classification	Morning			Afternoon			Evening			Night		
	NN	SVM	BPM	NN	SVM	BPM	NN	SVM	BPM	NN	SVM	BPM
Accuracy	0.540	0.533	0.533	0.579	0.556	0.584	0.644	0.645	0.644	0.713	0.705	0.710
Precision	0.534	0.530	0.545	0.616	0.567	0.594	0.646	0.645	0.645	0.622	0.000	0.564
Recall	0.939	0.940	0.875	0.558	0.713	0.694	0.990	1.000	0.998	0.055	0.000	0.052
F1 Score	0.681	0.678	0.671	0.586	0.631	0.640	0.782	0.784	0.783	0.100	0.000	0.095
AUC	0.520	0.502	0.556	0.606	0.586	0.605	0.587	0.554	0.589	0.637	0.586	0.636

TABLE 17. Two class boosted decision tree (DT), decision forest (DF), decision jungle (DJ), neural network (NN), support vector machine (SVM) and Bayes point machine (BPM) classification results.

	DT	DF	DJ	NN	SVM	BPM
Accuracy	0.719	0.711	0.704	0.622	0.596	0.600
Precision	0.689	0.709	0.694	0.821	0.575	0.580
Recall	0.732	0.654	0.661	0.252	0.543	0.543
F1 Score	0.710	0.680	0.677	0.386	0.559	0.561
AUC	0.771	0.782	0.769	0.673	0.650	0.676

overall mean accuracy and the decision jungle is the least accurate of the decision tree processes with a 70.78% overall mean score.

D. RESULTS OF EXPERIMENT 4

The final experiment offers one of the most potentially impactful methodologies that can have implications in the development of systems outside of the typical load balancing and data error domains. The detection of age groups has been briefly outlined in our previous work [7], but here the results are expanded on. The ability to detect age groups and their consumption patterns offers significant benefits to the health care domain. The results from Experiment 4 are presented in Figure 21 and detailed in Table 17.

The scores produced in the early stages of experiment are consistent with the results from previous experiments; yet, in this case the neural network, SVM and Bayes PM classifiers have an improved AUC score. The decision forest process scores the highest with a 78.2% AUC accuracy. However, on inspection of the dataset, during the data cleaning process an imbalance is created in the two-class dataset.

Class 1 accounts for 51% of the dataset and Class 2 the remaining 49%.

Whilst this might seem like an insignificant change in the dataset, the introduction of SMOTE to balance the data improves the classification accuracy significantly.

The results of the classification after conducting the SMOTE stage are presented in Figure 22 and Table 18.

Remarkably, from the SMOTE process, the AUC results are improved for the decision tree and decision jungle classification, which performs with 83% and 84.1% accuracy respectively.

A. DISCUSSION OF THE RESULTS

Experiment one begins with a direct vs in-direct classification comparison. The focus is also on the detection of the tariff grouping. A comparison of the results is presented in Figure 23. The results display a significant improvement, as expected, when cleaning, normalisation and feature extraction have been applied to the dataset. The decision forest shows the highest increase from 51.6% to 77.4%, which is a 25.8% percent increase.

Being able to predict the gas tariff the end user is on has beneficial applications. A demonstration of the change in

TABLE 18. Two class boosted decision tree (DT), decision forest (DF), decision jungle (DJ), neural network (NN), support vector machine (SVM) and Bayes point machine (BPM) classification results.

	DT	DF	DJ	NN	SVM	BPM
Accuracy	0.743	0.707	0.772	0.623	0.609	0.623
Precision	0.719	0.706	0.752	0.882	0.585	0.597
Recall	0.746	0.464	0.769	0.231	0.585	0.615
F1 Score	0.732	0.675	0.760	0.366	0.585	0.606
AUC	0.830	0.787	0.841	0.680	0.644	0.665

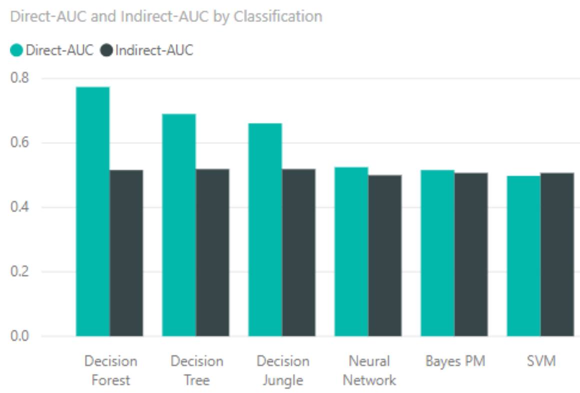

FIGURE 23. Experiment one results.

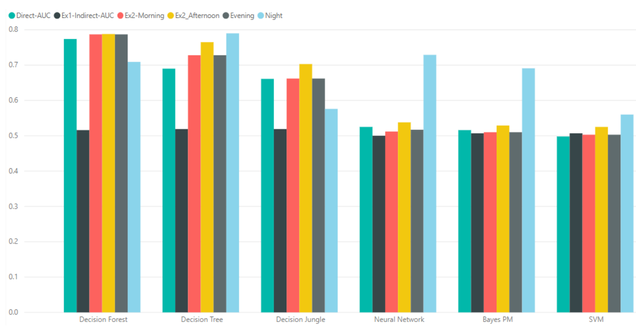

FIGURE 24. Experiment one vs experiment two.

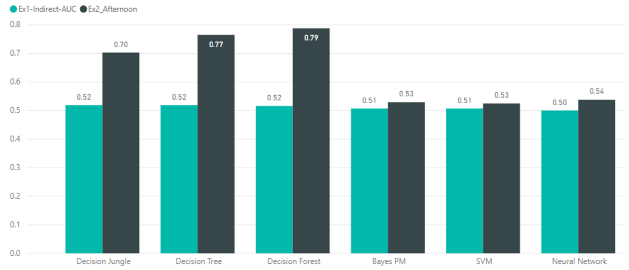

FIGURE 25. AUC results for experiment 2.

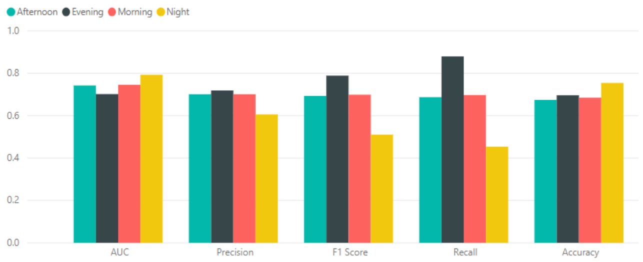

FIGURE 26. Mean classification performance of four time blocks.

behaviour that occurs when the customer is aware of their consumption habits in high detail.

Being able to predict the gas tariff the end user is on has beneficial applications. A demonstration of the change in behaviour that occurs when the customer is aware of their consumption habits in high detail. Tariff 4 is selected for this experiment for this purpose, as the customer is provided with an IHD and bi-monthly billing service. They are the consumers who are most aware of their gas usage. As a result, this serves as an ideal example of how a change is consumption occurs, which is beneficial for the environment and economically.

The second experiment employs the same methodology as experiment one, however with time blocks factored in to the classification. A comparison of the results achieved in both experiment one and two is displayed in Figure 24. The results are presented in order of highest mean classification score for the experiment to lowest.

The inclusion of the different time blocks improved the mean classification accuracy from between 61.10% to 63.80% for all the classifiers combined. However, when focusing on the decision tree approach, the classification mean average improved from 70.83% to 72.38%. The most significant increase is evident in the decision tree which improves from 69.0% to 76.5% mean AUC, when the data from the afternoon time block is assessed, as displayed in Figure 25.

In experiment three, the focus changes from tariff detection to time block comparison. Across the entire experiment, the six classifiers perform with a 74.56% AUC mean accuracy. However, during the night period the classification

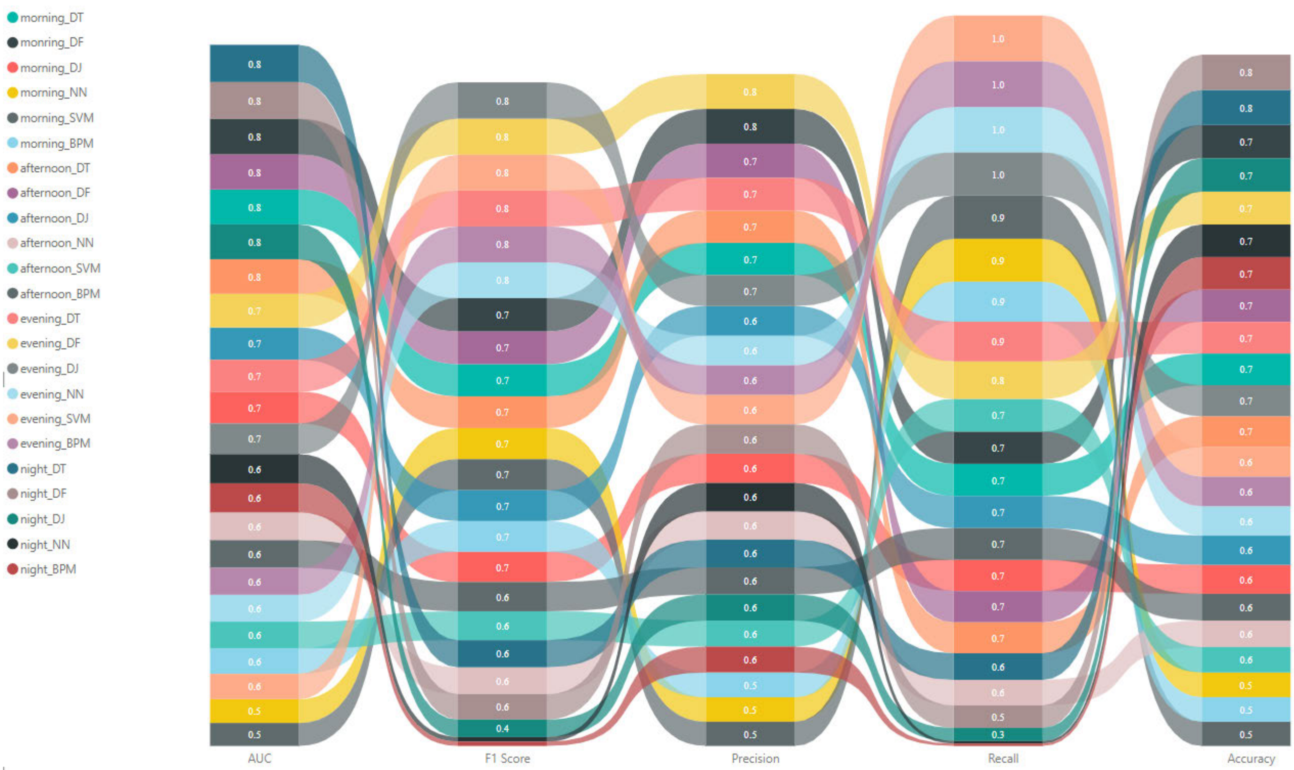

FIGURE 27. Classification performance of four time blocks.

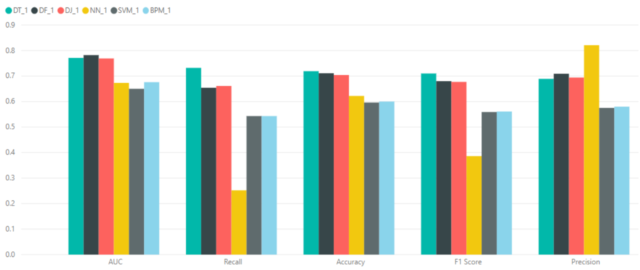

FIGURE 28. Classification performance of four time blocks.

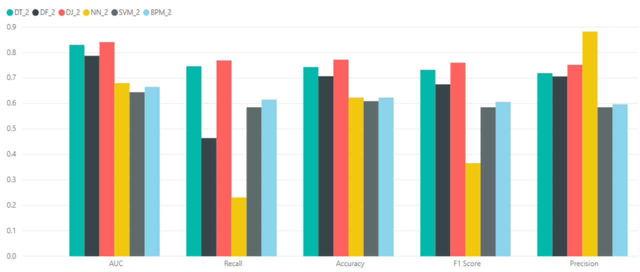

FIGURE 29. Classification performance of four time blocks.

is highest and able to achieve 79.27% accuracy; with the boosted decision tree scoring the highest of all the classifiers with 81.50% accuracy. A comparison of the four different time periods is displayed in Figure 26.

Figure 27 displays a comparison of the AUC, F1 Score, Precision, Recall and Accuracy scores for each of the classifiers for the different blocks of time. The visualisation effectively shows how the classifiers compare with each other for each of the different evaluation metrics.

In the final experiment, the mean AUC classification is 72.02%, as displayed in Figure 28. However, once again the three decision tree algorithms outperformed the others and were able to detect with a 77.4% accuracy combined compared to a 66.63% accuracy.

As previously mentioned, during the pre-processing stage, an imbalance is created in the dataset. As a result, SMOTE is used to restore the balance. As displayed in Figure 29, the overall mean AUC score increases from 72.02% to 74.12%.

However, individually the highest increase is noticeable for the boosted decision tree and the decision forest algorithms which increase to 83% an 84.10% respectively.

A comparison of the two sets of results from experiment four are presented in the ribbon chart in Figure 30. Each of the evaluation criteria are displayed on the x-axis. From the visualisation it is clear that the decision jungle and decision forest after SMOTE retain the highest classification accuracy.

B. RECOMMENDED ENSEMBLE DETECTION MODEL

Overall, based on the classification AUC scores, the recommended ensemble detection model framework is defined as follows.

The model combines the highest scoring techniques detected in the experiments and combines them to produce improved results.

For the Age Detection process, the Decision Forest Model is recommended. Stage 2 recommends two models, the DF for the morning, afternoon and evening detection and DT for the night detection. For the tariff detection, the DT is recommended for when a full 24-hour period is assessed, and

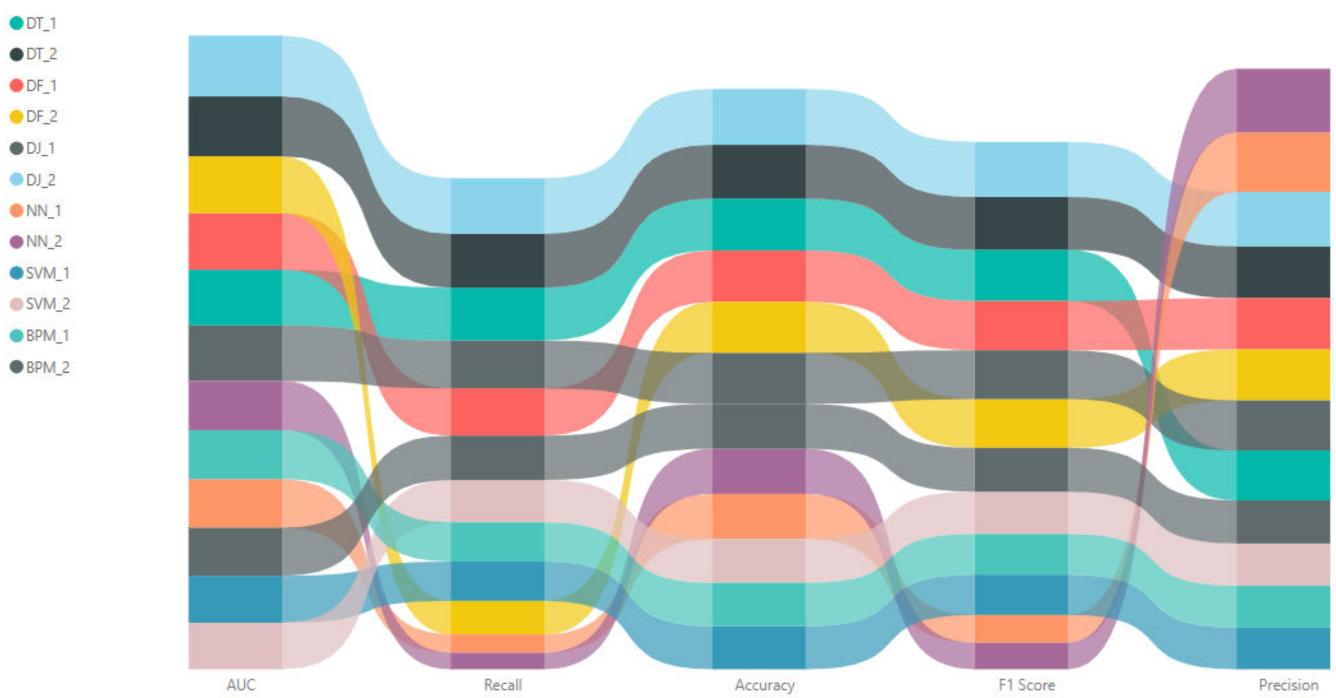

FIGURE 30. Ribbon plot of classifier performance for experiment 4.

Algorithm 1 *Ensemble Detection Model*

```
Function ageDetection
        Pass In: data block
            train classifier
            ageResults = AUC evaluation
        Pass Out: ageResults
Endfunction
Function routineDetection
        Pass In: data block
        split data block into four time periods
        FOR each time period
            train decision forest classifier (time period)
            routineResults[time period] = AUC
            evaluation
        ENDFOR
        Pass Out: routineResults
Endfunction
Function tariffDetection
        Pass In: data block
        train boosted decision tree classifier
        blockTariffResults = AUC evaluation
        split data block into four time periods
        FOR each time period
            train decision forest classifier
            periodTariffResults[time period] = AUC
                                                evaluation
        ENDFOR
        Pass Out: blockTariffResults, periodTariffResults
Endfunction
FOR each time period
        extract data block
        preprocess data block
        ageTraining = Call:ageDetection(arguments:
          preprocessed data block)
        IFageTraining result > 70THEN
              routineTraining = Call:
              routineDetection(arguments: preprocessed
data block)
            IFany routineTraining result > 70 THEN
                tariffTraining =
                  Call:tariffDetection(arguments:
                    preprocessed data block)
                IFany periodTariffResults or
                  blockTariffResults > 70THEN
                    IFmost active period != cheapest
period THEN
        recommendedTariff = tariff where most active
period == cheapest period
                ELSE
                    goto next time period
                ENDIF
            ELSE
                    goto next time period
            ENDIF
        ELSE
            goto next time period
        ENDIF
ENDFOR
```

a combination of DF and DT is recommended when time blocks are factored in.

V. CONCLUSION AND FUTURE WORK

In order to detect and support individual households that are at risk during seasonal periods (due to financial challenges, and the rising cost of bills), it is essential to adopt more advanced analytics at the service provider end. Within the future smart cities domain, the autonomous detection of such households at risk is of growing importance in order to face up to the impacts of energy and fuel poverty on energy, economy, quality of life and health and environmental quality of low-income housing.

As smart gas meters will eventually phase out the traditional analogue meter as society moves increasingly further towards a holistic smart city, the amount of information relating to consumer behaviour will increase significantly. However, with this, new opportunities for both providing more innovative services, and modernise exiting ones, helping researchers to understanding the behaviour of customers and gaining intelligent insight into the data patterns will grow. Yet, this must be conducted within the constraints of opt-in services to prevent data misuse and ensure that the privacy of consumer data is considered. The deployment of smart meters will also be key to the reduction in CO2 levels and will help towards reducing the carbon footprint [42].

In this paper, a method has been proposed to improve wellbeing monitoring using smart gas meter readings. There are four different observational characteristics involved in this process, each has proven successful in the experiments presented. The classifiers are able to establish the detection of certain patterns and trends within a population, not evident through visual inspection. This is particularly beneficial for health-based resource allocation and understanding how trends in health conditions are connected in a specific demographic. In the future, the approach could be built upon to help understand and visualise the health patterns that can be seen within an urban area. This offers an effective insight into the type of intervention that should be in place to help people with the most needs. This facilitates early intervention and the allocation of medical resources to key demographic areas. Future investigations will also include experimenting with other classification techniques; for example, clustering to identify all ages groups at the same time.

ACKNOWLEDGMENT

Owing to the ethical sensitive nature of this research, the data underlying this publication cannot be made openly available. However, it is available for request from the Commission for Energy Regulation (CER) (2012), the CER Smart Metering Project-Gas Customer Behaviour Trial, from 2009 to 2010 [dataset] (1st Edition), and the Irish Social Science Data Archive (SN: 0013-00. www.ucd.ie/issda/CER-gas).

REFERENCES

[1] J. Robinson, K. Lee, K. Appiah, and R. Yousef, "Energy-aware systems for improving the well-being of older people by reducing their energy consumption," *Int. J. Adv. Life Sci.*, vol. 9, no. 3, pp. 163–175, 2017.

[2] *Fuel Poverty: A Framework for Future Action, Department of Energy and Climate Change*, TSO, London, U.K., 2013.

[3] N. Hodges, D. Goaman, N. Banks, J. Thumim, and A. Lamley, "Supporting vulnerable consumers to benefit from their smart meters," Joseph Rountree Foundation, York, U.K., Tech. Rep., 2018, pp. 1–60.

[4] Y. Wang, Q. Chen, T. Hong, and C. Kang, "Review of smart meter data analytics: Applications, methodologies, and challenges," *IEEE Trans. Smart Grid*, vol. 10, no. 3, pp. 3125–3148, May 2019.

[5] A. Visvizi, M. D. Lytras, E. Damiani, and H. Mathkour, "Policy making for smart cities: Innovation and social inclusive economic growth for sustainability," *J. Sci. Technol. Policy Manage.*, vol. 9, no. 2, pp. 126–133, Jul. 2018.

[6] M. Mwansa, W. Hurst, C. Chalmers, Y. Shen, and A. Boddy, "A study into smart grid consumer-user profiling for security applications," in *Proc. 9th Int. Conf. Cloud Comput., GRIDs, Virtualization*, 2017, pp. 7–12.

[7] W. Hurst, C. A. C. Montañez, and D. Al-Jumeily, "Age group detection in stochastic gas smart meter data using decision-tree learning," in *Proc. Int. Conf. Intell. Comput. (ICIC)*, Nanchang, China, 2019, pp. 593–605.

[8] C. Chalmers, W. Hurst, M. Mackay, and P. Fergus, "Identifying behavioural changes for health monitoring applications using the advanced metering infrastructure," *Behav. Inf. Technol.*, vol. 38, no. 11, pp. 1154–1166, Nov. 2019.

[9] Z. Ma, J. Xie, H. Li, Q. Sun, Z. Si, J. Zhang, and J. Guo, "The role of data analysis in the development of intelligent energy networks," *IEEE Netw.*, vol. 31, no. 5, pp. 88–95, Sep. 2017.

[10] M. A. Devlin and B. P. Hayes, "Load identification and classification of activities of daily living using residential smart meter data," in *Proc. IEEE Milan PowerTech*, Jun. 2019, pp. 1–6.

[11] G. R. Barai, S. Krishnan, and B. Venkatesh, "Smart metering and functionalities of smart meters in smart grid—A review," in *Proc. IEEE Electr. Power Energy Conf. (EPEC)*, Oct. 2015, pp. 138–145.

[12] A. Visvizi and M. D. Lytras, "Rescaling and refocusing smart cities research: From mega cities to smart villages," *J. Sci. Technol. Policy Manage.*, vol. 9, no. 2, pp. 134–145, Jul. 2018.

[13] Z. A. Khan and D. Jayaweera, "Approach for forecasting smart customer demand with significant energy demand variability," in *Proc. 1st Int. Conf. Power, Energy Smart Grid (ICPESG)*, Apr. 2018, pp. 1–5.

[14] O. Elma and U. S. Selamogullar, "A survey of a residential load profile for demand side management systems," in *Proc. IEEE Int. Conf. Smart Energy Grid Eng. (SEGE)*, Aug. 2017, pp. 85–89.

[15] K. Basu, A. Hably, V. Debusschere, S. Bacha, G. J. Driven, and A. Ovalle, "A comparative study of low sampling non intrusive load dis-aggregation," in *Proc. 42nd Annu. Conf. IEEE Ind. Electron. Soc. (IECON)*, Oct. 2016, pp. 5137–5142.

[16] H. G. C. P. Dinesh, D. B. W. Nettasinghe, G. M. R. I. Godaliyadda, M. P. B. Ekanayake, J. V. Wijayakulasooriya, and J. B. Ekanayake, "A subspace signature based approach for residential appliances identification using less informative and low resolution smart meter data," in *Proc. 9th Int. Conf. Ind. Inf. Syst. (ICIIS)*, Dec. 2014, pp. 1–6.

[17] G. Dudek, A. Gawlak, M. Kornatka, and J. Szkutnik, "Analysis of smart meter data for electricity consumers," in *Proc. 15th Int. Conf. Eur. Energy Market (EEM)*, Jun. 2018, pp. 1–5.

[18] T. Sirojan, T. Phung, and E. Ambikairajah, "Intelligent edge analytics for load identification in smart meters," in *Proc. IEEE Innov. Smart Grid Technol.-Asia (ISGT-Asia)*, Dec. 2017, pp. 1–5.

[19] P. Gupta, T. T. T. Zan, J. Dauwels, and A. Ukil, "Flow-based estimation and comparative study of gas demand profile for residential units in Singapore," *IEEE Trans. Sustain. Energy*, vol. 10, no. 3, pp. 1120–1128, Jul. 2019.

[20] M. Lytras, A. Visvizi, and A. Sarirete, "Clustering smart city services: Perceptions, expectations, responses," *Sustainability*, vol. 11, no. 6, p. 1669, Mar. 2019.

[21] D. Gros, P. Wiest, K. Rudion, and A. Probst, "Parametrization of stochastic load profile modeling approaches for smart grid simulations," in *Proc. IEEE PES Innov. Smart Grid Technol. Conf. Europe (ISGT-Europe)*, Sep. 2017, pp. 1–6.

[22] R. Granell, C. J. Axon, and D. C. H. Wallom, "Impacts of raw data temporal resolution using selected clustering methods on residential electricity load profiles," *IEEE Trans. Power Syst.*, vol. 30, no. 6, pp. 3217–3224, Nov. 2015.

[23] I. Benítez, A. Quijano, J.-L. Díez, and I. Delgado, "Dynamic clustering segmentation applied to load profiles of energy consumption from Spanish customers," *Int. J. Electr. Power Energy Syst.*, vol. 55, pp. 437–448, Feb. 2014.

[24] *ER Smart Metering Project—Gas Customer Behaviour Trial*, Certified Emission Reduction, Dublin, Ireland, 2010.

[25] *Report On Smart Metering Technology Trials for Commission for Energy Regulation*, Certified Emission Reduction, Dublin, Ireland, 2011.

[26] A. Yassine, S. Singh, and A. Alamri, "Mining human activity patterns from smart home big data for health care applications," *IEEE Access*, vol. 5, pp. 13131–13141, 2017.

[27] J. Peppanen, X. Zhang, S. Grijalva, and M. J. Reno, "Handling bad or missing smart meter data through advanced data imputation," in *Proc. IEEE Power Energy Soc. Innov. Smart Grid Technol. Conf. (ISGT)*, Sep. 2016, pp. 1–5.

[28] R. Moghaddass and J. Wang, "A hierarchical framework for smart grid anomaly detection using large-scale smart meter data," *IEEE Trans. Smart Grid*, vol. 9, no. 6, pp. 5820–5830, Nov. 2018.

[29] G. Beaufort Samson, M.-A. Levasseur, G. Gagnon, and F. Gagnon, "Autocalibration of Hall effect sensors for home energy consumption monitoring," *Electron. Lett.*, vol. 50, no. 5, pp. 403–405, Feb. 2014.

[30] H. Clougherty, A. Brown, M. Stonerock, M. Trepte, M. Whitesell, and R. Bailey, "Home automation and personalization through individual location determination," in *Proc. Syst. Inf. Eng. Design Symp. (SIEDS)*, Apr. 2017, pp. 300–305.

[31] V. Vadakattu and S. Suthaharan, "Feature extraction using apparent power and real power for smart home data classification," in *Proc. 17th IEEE Int. Conf. Mach. Learn. Appl. (ICMLA)*, Dec. 2018, pp. 1290–1295.

[32] Y. Zhou, X. Chen, A. Y. Zomaya, L. Wang, and S. Hu, "A dynamic programming algorithm for leveraging probabilistic detection of energy theft in smart home," *IEEE Trans. Emerg. Topics Comput.*, vol. 3, no. 4, pp. 502–513, Dec. 2015.

[33] M. D. Lytras and A. Visvizi, "Who uses smart city services and what to make of it: Toward interdisciplinary smart cities research," *Sustainability*, vol. 10, no. 6, p. 1998, Jun. 2018.

[34] A. Guez, V. Protopopsecu, and J. Barhen, "On the stability, storage capacity, and design of nonlinear continuous neural networks," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-18, no. 1, pp. 80–87, Feb. 1988.

[35] J. Nalepa and M. Kawulok, "Selecting training sets for support vector machines: A review," *Artif. Intell. Rev.*, vol. 52, no. 2, pp. 857–900, Aug. 2019.

[36] R. Herbrich, T. Graepel, and C. Campbell, "Bayes point machines," *J. Mach. Learn. Res.*, vol. 1, pp. 245–279, Aug. 2001.

[37] F. Wellmer, "Standard deviation and variance in the mean," in *Statistical Evaluations in Exploration for Mineral Deposits*. Springer, 1998.

[38] B. McNeese, "Are the skewness and kurtosis useful statistics?" *SPC Stat.*, 2016. [Online]. Available: https://www.spcforexcel.com/knowledge/basic-statistics/are-skewness-and-kurtosis-useful-statistics

[39] P. H. Westfall, "Kurtosis as peakedness, 1905–2014. *R.I.P.*," *Amer. Stat.*, vol. 68, no. 3, pp. 191–195, 2014.

[40] C. Croux and G. Haesbroeck, "Influence function and efficiency of the minimum covariance determinant scatter matrix estimator," *J. Multivariate Anal.*, vol. 71, no. 2, pp. 161–190, Nov. 1999.

[41] A. J. Boddy, W. Hurst, M. Mackay, and A. El Rhalibi, "Density-based outlier detection for safeguarding electronic patient record systems," *IEEE Access*, vol. 7, pp. 40285–40294, 2019.

[42] P. Carroll, T. Murphy, M. Hanley, D. Dempsey, and J. Dunne, "Household classification using smart meter data," *J. Off. Statist.*, vol. 34, no. 1, pp. 1–25, Mar. 2018.

WILLIAM HURST received the B.Sc. (Hons.) degree in product design from Sussex University, in 2007, and the M.Sc. degree (Hons.) in web computing and the Ph.D. degree in computer science from Liverpool John Moores University (LJMU), in 2010 and 2014, respectively. He is a Reader of creative technologies with the Department of Computer Science, LJMU. He has over 60 international publications in the areas of data analytics, machine learning, computer graphics, simulation, virtual and augmented reality, cyber security, and behavioural observation. He has also had successful grants from EPSRC, Innovate U.K., SBRI, and UKAIS for multiple projects within his research areas.

CASIMIRO ADAY CURBELO MONTAÑEZ received the B.Eng. degree in telecommunications from Alfonso X el Sabio University, Madrid, Spain, and the M.Sc. degree in wireless and mobile computing and the Ph.D. degree from Liverpool John Moores University (LJMU), in 2014 and 2019, respectively. He is currently a Research Assistant with the Department of Computer Science, LJMU. He is currently a part of a research team, working on strategies to detect anomalies utilizing the smart metering infrastructure. His research interests include various aspects of data science, and machine learning and their applications in bioinformatics.

NATHAN SHONE received the B.Sc. (Hons.) in computer forensics and the Ph.D. degree in network security from Liverpool John Moores University (LJMU) in 2010 and 2014, respectively. He is a Senior Lecturer with the Department of Computer Science, LJMU. His experience is within the areas of monitoring, focusing on real-time behavioural monitoring, and statistical behavioural modeling and analysis. He has over 20 publications. He has been involved in writing and securing numerous successful grants including H2020, HEA, EPSRC, Newton Fund, and UKAIS. His previous research related projects include the development of deep learning-based intrusion detection system, a productivity accelerator platform, and a critical infrastructure research testbed. His current research interest includes applications of deep learning for botnet detection.

DHIYA AL-JUMEILY (Senior Member, IEEE) is a Professor of artificial intelligence and the Associate Dean of external engagement with the Faculty of Engineering and Technology, Liverpool John Moores University. He has extensive research interests covering a wide variety of interdisciplinary perspectives concerning the theory and practice of applied computing in medicine, human biology, and health care. He has published over 170 peer-reviewed scientific publications in multidisciplinary research areas including technology enhanced learning, applied artificial intelligence, neural networks, signal prediction, AI-based clinical decision-making, medical knowledge engineering, human–machine interaction, intelligent medical information systems, and wearable and intelligent devices and instruments. He is a Chartered IT Professional. He is also a fellow of the U.K. Higher Education Academy.