IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 7 June 2022, accepted 27 June 2022, date of publication 4 July 2022, date of current version 8 July 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3188004

RESEARCH ARTICLE

Influence of Geodemographic Factors on Electricity Consumption and Forecasting Models

JITENDER PAL SINGH[1], OMAR ALAM[1], AND ABDULSALAM YASSINE[2], (Member, IEEE)
[1]Trent University, Peterborough, ON K9L 0G2, Canada
[2]Lakehead University, Thunder Bay, ON P7B 5E1, Canada

Corresponding author: Omar Alam (omaralam@trentu.ca)

ABSTRACT The residential sector is a major consumer of electricity, and its demand will rise by 65 percent by the end of 2050. The electricity consumption of a household is determined by various factors, e.g. house size, socio-economic status of the family, size of the family, etc. These factors play a critical role in analyzing the energy consumption causalities in the residential sector for better energy prediction models, effective price policy implementations, and improved customer engagement in energy efficiency programs. However, determining the effect of demographic factors on energy consumption is a challenging prospect. First, it is not trivial to study the causes of energy consumption variation, even for similar size residential houses, without analyzing the impact of interdependencies between demographic factors on energy consumption behavior. Second, to achieve higher accuracy of energy prediction models, it is necessary to identify key geodemographic factors that influence these models. Previous studies have only identified a limited number of socio-economic and dwelling factors. In this paper, we study the significance of 826 *geodemographic* factors on electricity consumption for 4917 homes in the City of London. *Geodemographic factors* cover a wide array of categories e.g. social, economic, dwelling, family structure, health, education, finance, occupation, and transport. Using Spearman correlation, we have identified 354 factors that are strongly correlated with electricity consumption. We also examine the impact of using *geodemographic* factors in designing forecasting models. In particular, we develop an encoder-decoder LSTM model which shows improved accuracy with geodemographic factors. We believe that our study will help energy companies design better energy management strategies.

INDEX TERMS Socio-economic factors, geodemographic factors, electricity forecasting, encoder-decoder model.

I. INTRODUCTION

Electrical power demand has been on the rise in the last few decades. From 1974 to 2018, energy consumption in the housing sector rose five-folds, from 1203 TWh to 6008 TWh [1]. The US Energy Information Administration [2], expects this number to further rise by 65% in 2050. Therefore, it is essential for the utility companies to forecast the consumer demand and manage its supply.

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

Over the years, researchers have developed several forecasting models using techniques such as, ARIMA [3], Support Vector Regression (SVR) [4] and artificial neural networks [5]. However, these approaches did not incorporate data about consumer demographics. Energy consumption of a home is influenced by several socio-economic factors, e.g. number of occupants, family composition, age, employment status, education and household income. In addition, there factors related to the house (dwelling) itself, e.g., house type, house age, number of rooms, number of floors and size. Jones *et al.* [6] provides a summary of socio-economic and dwelling factors that have been studied in the past.

The aforementioned factors do not count for all characteristics of a household. In this paper, rather than limiting ourselves to few socio-economic and dwelling factors, we expand to a large number of factors. In particular, we use the concept of *geodemographic factors* [7] to understand the energy consumption of a household. Geodemographics factors cover an extensive range of demographic information using 826 different variables. These variables are selected from 15 different categories, including social, economic, dwelling, family structure, health, education, finance, occupation, transport, digital behavior and leisure-time preferences.

We believe that geodemographic information can help power generation companies design better energy management strategies. Because consumer behaviour could be influenced by factors such as their socio-economic conditions, house size, and ethnicity, tailored strategies could be developed that target needs of specific group of consumers, based on their energy consumption behaviour. Furthermore, geodemographic information can help power generation companies design better forecasting models that can help them predict the consumer demand more precisely. This will also ensure that the companies meet their demand supply balance. In this paper, we study the impact of geodemographic factors on energy consumption by addressing the following research questions:

RQ1: What are the geodemographic factors that affect energy consumption?

RQ2: How does the geodemographic factors impact the accuracy of a forecasting model?

We answer the above questions by analyzing the London Smart meter dataset [8]. In particular, we use the geodemographic information for 4917 homes in the City of London to understand which factors are strongly correlated with electricity consumption. We find that students and unemployed people have negative correlation with energy consumption (i.e. they seem to be consuming less electricity), while professionals, managers, and people who work from home have positive correlation. We also find that people who are conscious about the environment have positive correlation (i.e. they seem to be consuming more electricity), while environment and climate change skeptics have negative correlation. This may seem counter-intuitive, suggesting that holding certain opinions on climate change may not impact electricity consumption in the way we hope it would. We suspect that people who are conscious about the environment come from affluent backgrounds, i.e. they are highly educated, live in large homes, have more appliances, thus consume more electricity. We then build a machine learning model for forecasting future energy consumption, and study whether adding geodemographic factors impacts the accuracy of the model. We find that using geodemographic factors significantly improves the accuracy of our forecasting model.

The remaining part of the paper is organised as follows. Section II covers the literature review. Section III discusses the dataset. Section IV discusses the methodology of our research questions. Section V covers our results. Section VI concludes the paper and discusses the future work.

II. LITERATURE REVIEW

Several studies have designed models for forecasting energy consumption. Yuan *et al.* [9] presents a comparative study of 3 univariate models: the ARIMA model, GM (1,1) model, and a hybrid model for forecasting the energy consumption in China. They find that the hybrid model, which uses both the ARIMA and GM (1,1) model, has the least Root Mean Square Error. Akay *et al.* [10] develops a GPRM (*Grey Prediction with Rolling Mechanism*) approach to predict electricity demand in Turkey. GPRM is a simple, computationally inexpensive technique, and it is highly effective in dealing with volatile data. Boran *et al.* [11] uses the Box-Jenkins method (ARIMA model) to predict total electricity consumption in Turkey from 2009 to 2014. Several studies design a forecasting model for electricity consumption using Support Vector Regression (SVR), stochastic or regression models [4], [12]–[17]. Other studies used neural networks for forecasting electricity consumption [18]–[21]. None aforementioned techniques uses geodemographic data as we do in our study.

There are studies that looked into the relationship between economic and socio-economic factors and energy consumption, but not with comprehensive geodemographic factors [22]–[29]. For example, Brounen *et al.* [30] and Permana *et al.* [31] study the relationship between gender and electricity consumption. Other studies looked into home size [32], presence of air conditioning/heating [33], [34], or appliances used at home [35].

Two studies use the dataset that we used in our study. Alonso *et al.* [36] develop a separate LSTM model for each geodemographic group to predict the electricity consumption of houses within that group. In our study, we focus on showing the significance of geodemographic factors. Therefore, we design a multi-step encoder-decoder model and compare the model performance with and without including geodemographic factors. We also identify the geodemographic factors that have significant correlation with the electricity consumption. Singh *et al.* [37] is motivated to improve the efficiency of the *Demand Response (DR)* strategy. They use a multivariate clustering technique on the smart meter data and generate *consumer classes*. These classes reflect the consumer's behavior, demographics, economics, and dwelling characteristics. They prove that demographic factors can improve DR strategies, but they fail to identify those factors that are actually relevant in determining electricity consumption. Our study explicitly identifies those factors. We then design a forecasting model that shows that geodemographic factors can be used to enhance the model accuracy.

III. DATASET

The datasets we use for this study are: *London Smart meter data* [8] and *ACORN geodemographic data* [38]. *London smart meter data* provides the electricity consumption data for the homes in the City of London. Each home is assigned a geodemographic label, called ACORN (*A Classification of Residential Neighbourhoods*) [39]. There are 17 different ACORN labels (ACORN-A to ACORN-Q). Each label groups a certain section of population based on 826 factors. The details about these factors are in *ACORN geodemographic data*. Both datasets are clean and sorted in a table form, therefore, no preprocessing step is involved. In next subsections, we explain each of these datasets.

A. LONDON SMART METER DATA

This dataset is provided by U.K. Power Networks [8]. It contains the electricity consumption of 4917 households in London for 2013.

TABLE 1. Metadata of the london smart meter data.

Feature Name	Description
LCLid	Home Id
stdorToU	Static or Dyanmic Pricing
DateTime	Date and timings of the reading
kWh	Smart meter reading
Acorn	Geodemographic group

The metadata of *London smart meter data* is shown in Table 1. Every household is given a unique identifier (*LCLid*). *stdorToU* denotes the tariff structure (static or dynamic). In previous work, we studied the response of different geodemographic groups to static and dynamic pricing [40]. It can help to identify the consumers who are sensitive to price fluctuations. The electricity consumption for the household is recorded in the form of smart-meter readings. These readings are collected at half hour intervals. *DateTime* provides the timings of these readings, and *kWh* denotes the electricity consumption (in kWh). *Acorn* specifies the ACORN label of the home.

B. ACORN GEODEMOGRAPHIC DATA

This data is provided by CACI (California Analysis Center, Inc.) [41]. CACI has formulated the ACORN labels based on 826 factors. These factors belong to 15 categories shown in Figure 1. Each ACORN label assigns different value for all the factors. For example, the housing category in Figure 1 has 24 factors. One of this factor is *House Type: Detached*. ACORN-A assigns this factor value as 431, whereas ACORN-Q assigns this factor value as 13. A value of 100 on this scale denotes the national average of UK. So, a value of 431 for ACORN-A means that there are 331% more detached homes in ACORN-A compared to the national scale. Similarly, ACORN-Q has 87% less detached homes as compared to the national average.

Overall, CACI has developed the geodemographic classification system of UK with 17 different ACORN labels. These labels are grouped into five major geodemographic groups. The entire classification is shown in the Table 2. In this paper, we have only summarised the characteristics of the five geodemographic groups.

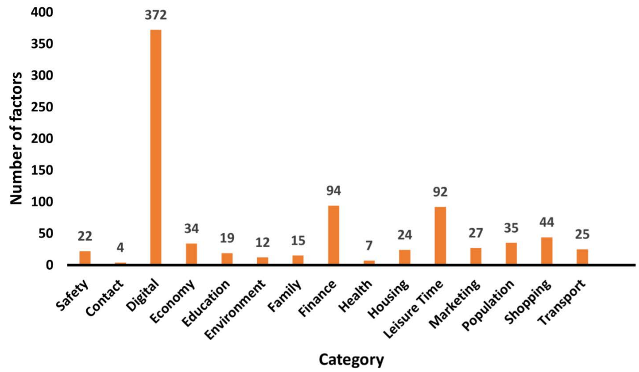

FIGURE 1. Number of factors belonging to each category.

TABLE 2. The complete geodemographic classification system of UK.

	Geodemographic Group	ACORN Label
1	Affluent Achievers	a) **ACORN-A** (*Lavish Lifestyle*) b) **ACORN-B** (*Executive wealth*) c) **ACORN-C** (*Mature Money*)
2	Rising Prosperity	a) **ACORN-D** (*City Sophisticates*) b) **ACORN-E** (*Career Climbers*)
3	Comfortable Communities	a) **ACORN-F** (*Countryside Communities*) b) **ACORN-G** (*Successful Suburbs*) c) **ACORN-H** (*Steady Neighborhoods*) d) **ACORN-I** (*Comfortable Seniors*) e) **ACORN-J** (*Starting Out*)
4	Financially Stretched	a) **ACORN-K** (*Student Life*) b) **ACORN-L** (*Modest Means*) c) **ACORN-M** (*Striving Families*) d) **ACORN-N** (*Poor Pensioners*)
5	Urban Adversity	a) **ACORN-O** (*Young Hardships*) b) **ACORN-P** (*Struggling Estates*) c) **ACORN-Q** (*Difficult Circumstances*)

1) GROUP 1: AFFLUENT ACHIEVERS

This group involves middle-aged and older people, typically born in the *baby boomer generation*. They live in posh suburban and rural areas and generally hold graduate or post-graduate degrees.

2) GROUP 2: RISING PROSPERITY

This group includes the young educated crowd, who work in high-paying professional jobs and live in cities. They prefer executive-style flats, which are generally associated with the corporate culture.

3) GROUP 3: COMFORTABLE COMMUNITIES

This group involves most of the middle-class community in the UK. They reside in suburbs, small towns, and villages.

4) GROUP 4: FINANCIALLY STRETCHED

This group includes households whose incomes are below the national average. They work in lower administrative, clerical,

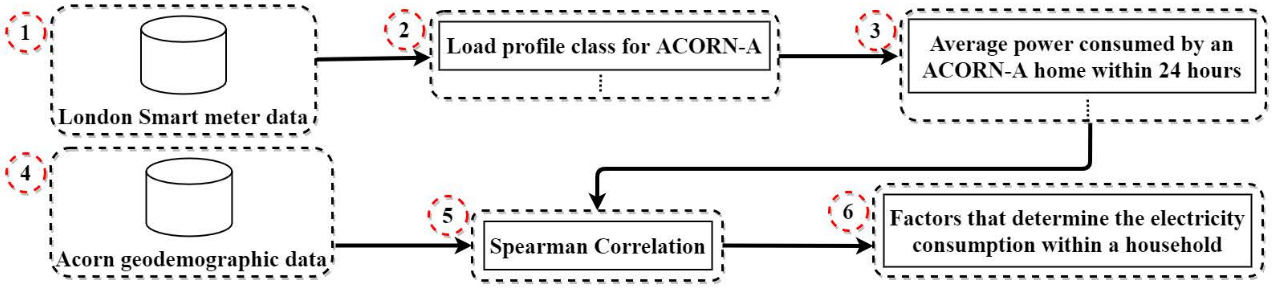

FIGURE 2. Steps involved in answering RQ1.

semi-skilled, and manual jobs. Most of them have O-level education, live in terraced or semi-detached homes.

5) GROUP 5: URBAN ADVERSITY

This group contains people mainly from towns and cities. They do not have a good financial standing and face difficulties while applying for loans/credit-cards. Their education level is very low, mainly work in unskilled jobs, earn little income, and live in social housing, which are often small and crowded.

In our study, both the ACORN the London Smart Meter datasets are available for free for the City of London. Electricity companies may face challenges when they want to use geodemograpic data to understand electricity consumption of other cities. Possible constrains, e.g. privacy laws or lack of consumer cooperation, may make the use of gedemographic factors difficult.

IV. METHODOLOGY

In this section, we discuss the methodology used to answer the research questions. In RQ1, we use Spearman correlation between the power consumption of each home and its geodemographic factors. Spearman Correlation is a useful statistic to measure the association between different variables. In RQ2, we develop a machine learning model that uses the power consumption for each home and its associated ACORN label. As discussed previously, the ACORN label contains the geodemographic information for the home. We study the accuracy of our model when geodemographic factors are used.

Next subsections discuss the methodology in detail.

A. RQ1: WHAT ARE THE MAJOR GEODEMOGRAPHIC FACTORS THAT INFLUENCE ELECTRICITY CONSUMPTION?

The methodology to answer this research question is summarised in Figure 2. In step 1, we start with the *London smart meter data*. It contains the power consumption reading for each home. As mentioned before in section 3, each home is labelled with an ACORN. In step 2, we group all the homes for each ACORN together and get the average of their consumption. This step generates a *load profile class* for that ACORN label. In step 3, we use the *load profile class* to calculate the amount of electricity consumed over a period of 24 hours. Step 4 reads the 826 factors for each ACORN label (as provided by *ACORN geodemographic data*). In step 5, we correlate the factors (from step 4) with the electricity consumption (calculated in step 3) using Spearman Correlation. Finally, step 6 outputs all the factors that are correlated with the electricity consumption.

B. RQ2: HOW DOES GEODEMOGRAPHIC INFORMATION AFFECT THE ACCURACY OF A FORECASTING MODEL?

Energy companies need to forecast their consumer demand to meet the demand-supply balance. They employ different forecasting algorithms to achieve better forecasting accuracy. Only a limited number of factors, e.g. house size, household income, were used in those forecasting algorithms [6], [30], [34]. An ACORN label carries a large number of demographic factors and has the potential to alter a forecasting model's performance. In this research question, we explore this possibility. In particular, we build a neural network-based (Encoder-Decoder LSTM) forecasting model and study whether using ACORN labels improve its accuracy. This subsection starts with an introduction to LSTM (Long Short Term Memory) [42] and Encoder-Decoder LSTM model [43]. Then we will discuss our forecasting model.

1) LONG SHORT TERM MEMORY (LSTM)

LSTM is a variant of RNN (Recurrent Neural Networks). RNN suffers from the vanishing gradient problem [44]. LSTM overcomes this problem by maintaining a cell state which acts as a memory of the network [42].

An LSTM cell consists of three gates: input gate, output gate, and forget gate. The forget gate identifies the part of the information that is no longer needed and removes it from the cell state. The input gate identifies the new information that is relevant and stores it in the cell state. The output gate decides the part of the cell state that will be passed to the next LSTM cell as the hidden state.

LSTM is described by the following equations [42]:

$$Forget\ gate : f_g = \beta(W_{fg} \cdot [h_{t-1}, x_t] + b_{fg}) \quad (1)$$

$$Input\ Gate : i_g = \beta(W_{ig} \cdot [h_{t-1}, x_t] + b_{ig}) \quad (2)$$

$$Cell\ state : C_t = f_g * C_{t-1} + i_g *$$

$$\times\, tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (3)$$
$$Output\ gate : o_g = \beta(W_{og}[h_{t-1}, x_t] + b_{og}) \quad (4)$$
$$hidden\ state : h_t = o_g * \tanh(C_t) \quad (5)$$

W_{fg}, W_{ig}, and W_{og} denote the weight matrix for the forget gate, the input gate, and the output gate, respectively. b_{fg}, b_{ig} and b_{og} denote the bias vector for the forget gate, the input gate and the output gate, respectively. x_t denotes the input at time t. C_t denotes the cell state vector, and h_t denotes the hidden state vector. β is the sigmoid function and $\beta = \frac{1}{1+e^{-x}}$.

2) ENCODER-DECODER LSTM MODEL

This model is predominantly used in sequence learning problems like language translation and image captioning [45]. The model learns to generate an output sequence based on an input sequence. The encoder converts the input $(x_1, x_2, \ldots, x_n)$ into a *fixed length vector* c_T [43]. The vector c_T is a summary of the input sequence. It is passed to the decoder, which generates an output o_i $(i \in (1, n'))$, one time-step at a time. Subsequently, it generates the entire output sequence $(o_1, o_2, \ldots, o_{n'})$.

The encoder-decoder model computes the conditional probability of generating the output sequence $(o_1, o_2, \ldots, o_{n'})$ given the input sequence $(x_1, x_2, \ldots, x_n)$. It is represented by $p(o_1, o_2, \ldots, o_{n'}|x_1, x_2, \ldots .x_n)$ and calculated using equation 6 [46].

$$p(o_1, o_2, \ldots, o_{n'}|x_1, x_2, \ldots .x_n) \approx \prod_{n=1}^{n=n'} p(o_n|c_T, o_1, o_2, \ldots, o_{n-1}) \quad (6)$$

Normally, both the encoder and decoder consist of a single layer of LSTM cells. We can also stack multiple layers to achieve better accuracy while requiring more training time. In this paper, we are mainly interested in studying the significance of geodemographic factors. Therefore, we keep the design of our encoder and decoder relatively simple.

Our study uses an Intel Core i5-9300H CPU (@2.4GHz) machine with 16GB RAM. Therefore, a simple encoder-decoder design helps us to retrain the model within 48 hours.

3) ELECTRICITY FORECASTING MODEL

We use the encoder-decoder LSTM architecture to design our model. For simplicity, we consider only one layer of 200 LSTM cells to design our encoder and another layer of 200 LSTM cells to design our decoder. Since the model uses 200 cells in our encoder, we get the vector c_T of length 200.

Figure 3 describes the architecture of the proposed model. The overall goal is to predict the total energy consumption of a city. Since the dataset contains 4917 homes only, we assume that there are 4917 homes in the city. Accordingly, we compute the prediction of energy consumption of the city.

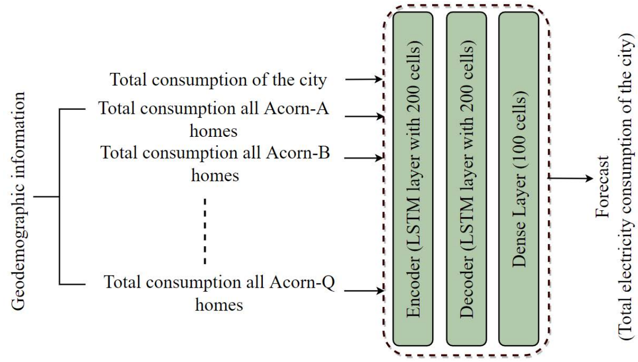

FIGURE 3. **Architecture of our forecasting model.**

a: MODEL FEATURES

There are 23 features used in this model: *Day of the week*, *Weekend or Weekday*, *Day of the month*, *Hour of the day*, *month*, *Hour of the day*, *Total consumption of the city*, and *total consumption for homes of each geodemographic group (17 ACORN group)*.

b: MODEL TRAINING

The layer of LSTM cells expect a 3 dimensional input in the form of [*samples, time-steps, features*]. One *sample* includes a pair of input and output sequence i.e. $[(x_1, x_2, \ldots, x_n), (o_1, o_2, \ldots, o_{n'})]$. *Time-steps* specifies the number of observations in a sample. We use 16 time-steps in the input sequence $(x_1, x_2, \ldots, x_{16})$ and 16 time-steps in the output sequence $(o_1, o_2, \ldots, o_{16})$. Every element in the input-sequence $(x_1, x_2, \ldots, x_{16})$ is a vector of 23 elements (features). Every element in the output sequence $(o_1, o_2, \ldots, o_{16})$ is a vector of only one element (i.e. *total consumption of the city*).

For training this model, we use 51 weeks of smart meter data starting on Jan 06, 2013 (Sunday) and ending on Dec 28, 2013 (Saturday). The initial 42 weeks of the data are used in the training phase, and the remaining 9 weeks of data are used for testing the model. Table 3 describes the values of various parameters used in the training phase.

TABLE 3. **Hyper-parameter values.**

Parameter Name	Value chosen for the model
Activation Function	ReLU
Optimizer	Adam
Epochs	40
Batch Size	16

V. RESULTS

In this section, we present our findings for RQ1 and RQ2. In section 4.1, we used Spearman Correlation between the power consumption of a home and its geodemographic factors. Section 5.1 summarises its results. In Section 4.2.3, we designed an electricity forecasting model with *geodemographic information*. We provide discussion on its accuracy in section 5.2.

A. RQ1: WHAT ARE GEODEMOGRAPHIC FACTORS THAT INFLUENCE ENERGY CONSUMPTION?

Out of 826 geodemographic factors, we find that 354 factors share strong correlation with power consumption. They have a Spearman correlation in the range ($0.6 \leq |\rho| \leq 1.0$) and their p-values are statistically significant ($p \leq 0.05$). Figure 4 shows which categories these 354 factors belong to. For the sake of brevity, the next subsections discuss the highly correlated factors from each category. We analyze which geodemographic group in Table 2 these factors belong to. Interested readers may consult the complete list of these factors in following the link [47].

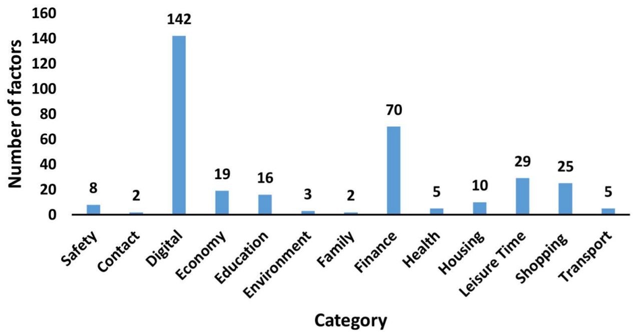

FIGURE 4. Results of spearman correlation (RQ1).

1) HOUSING

In our analysis, we find that 10 housing factors have a significant correlation with the electricity consumption. Table 4 shows the results, where the top three entries correspond to the 3 highest correlations. Overall, we can make the following observations:

TABLE 4. Correlation results for the *Housing* category.

Property	Correlation	P-value
Number of beds: *5 plus*	0.792	0.00015
Number of beds: *4*	0.767	0.00033
House Value: *Upto £100,000*	-0.757	0.00043
House Tenure: *Social Renting*	-0.738	0.00072
House Value: *greater than £1 million*	0.723	0.00105

a) Large and expensive homes share very strong positive correlation with the electricity consumption. A big house generally has more consumer appliances leading to an increased electricity consumption. Similar observation has also been recorded in Gonzalez *et al.* [48].

b) Cheap houses (value £0 - £100,000) and social renting share strong negative correlation with the electricity consumption. These houses are small in size, and are occupied by poor families who are inclined to be economical in electricity consumption [49].

2) ECONOMY

19 economic factors have strong correlation with the electricity consumption. Table 5 shows the top 3 rows corresponding to the three highest correlations with the following observations:

TABLE 5. Correlation results for the *Economy* category.

Property	Correlation	P-value
Higher managerial	0.89	0
Social Grade: *B*	0.887	0
Social Grade: *E*	-0.899	0
Occupation: *Student*	-0.623	0.00753
Economic Activity: *Self-employed*	0.693	0.00205

a) NS-SEC (National Statistics Socio-economic classification) [50] system contains eight categories. The *Higher managerial*, *Social grade B* (professional jobs), and *Self-employed* categories show strong positive correlation as people in this categories live in very big houses and premium flats. People in *Higher Managerial* tend to work from home often [51], therefore, consuming more electricity.

b) Social grade E has a strong negative correlation. This group mainly includes unemployed people who depend on income support. They try to reduce their electricity bills by maintaining low consumption. Similarly, students also have strong negative correlation.

3) EDUCATION

Based on the results in Table 6, we can draw the following observations:

TABLE 6. Correlation results for the *Education* category.

Property	Correlation	P-value
Pupils at the end of KS2 (Key Stage 2) level	0.871	0.00001
Pupils at the end of KS4 (Key Stage 4) level	0.86	0.00001
Education: *No formal qualification*	-0.863	0.00001
Education: *Degree or Higher Degree*	0.846	0.00002

a) Families with children in KS2 (Key Stage 2) and KS4 (Key Stage 4) who have academic performance have a strong positive correlation with electricity consumption. Parents in these families work in professional or managerial jobs [52].

b) An adult with no formal education has a strong negative correlation with electricity consumption. Contrarily, a university/college degree holder has a strong positive correlation with electricity consumption. Having a university degree helps in securing a well-paid job [53] and affording a lifestyle that consumes higher electricity.

4) FINANCE

The finance category has 70 significant factors. This category is related to the Economy category discussed before but focuses more on spending behaviours. Based on the results from Table 7, we can make the following observations:

TABLE 7. Correlation results for the *Finance* category.

Property	Correlation	P-value
Household Annual Income: *£100,000+*	0.91	0
Household Annual Income: *£0 - £20,000*	-0.895	0
Manage *current account* in branch	-0.899	0
Weekly expenses for alcohol/tobacco	-0.696	0.00192
Weekly expenses for *Clothing/Footwear*	0.677	0.00284
Benefits: *Job-Seeker's allowance*	-0.743	0.00063

a) People with high income (i.e. Affluent Achievers in Table 2) regularly shop for clothing/footwear [54] have a strong positive correlation with electricity consumption.

b) Low-income households maintain a low electricity consumption, which leads to a strong negative correlation. People with low income prefer social renting and live in small houses. Their salaries are inadequate to meet the basic expenses, so they depend on social benefit schemes like job-seeker allowance, income support, and disability allowance. We also observe that they spend more money for alcohol/tobacco.

c) Pensioners with little income prefer to visit the bank branch instead of doing online banking. They use social renting and maintain a low electricity consumption, which leads to a strong negative correlation.

5) TRANSPORT

Table 8 shows the 5 significant factors belonging to the Transport category. We observe the following:

TABLE 8. Correlation results for the *Transport* category.

Property	Correlation	P-value
Main Car Class: *Luxury or Executive*	0.844	0.00002
Car Ownership: *Number of Cars 2*	0.706	0.00154
Car Ownership: *Number of Cars 0*	-0.612	0.00904
Work mainly from home	0.605	0.01002

a) People who own more than one car or luxury cars have strong positive correlation with electricity consumption. The same applies to people who do not commute because they work from home.

b) People who do not own cars use public transit. They have negative correlation with electricity consumption.

6) SHOPPING

The shopping category contains 25 significant factors. Table 9 shows the results for some shopping habits. We can make the following observations:

TABLE 9. Correlation results for the *Shopping* category.

Property	Correlation	P-value
Preferred Supermarket: *M & S*	0.891	0
Clothing & Footwear Stores: *Mass Market*	0.88	0
Checks product reviews before purchasing	0.867	0.00001
Food shopping: *Budget Ranges*	-0.744	0.00061
Footwear & Clothing Stores: *Premium*	0.8	0.00011
Footwear & Clothing Stores: *Value*	-0.781	0.00022
Spends extra money on quality products	0.817	0.00006

a) Affluent Achievers and Rising Prosperity groups go premium supermarkets and clothing/footwear stores. The quality of a product is important for these shoppers. We observe that these people have high positive correlation with electricity consumption.

b) People who shop on budget have strong negative correlation with electricity consumption.

7) ENVIRONMENT

The environment category has three significant factors. Table 10 shows the results. Overall, we can make the following observations:

TABLE 10. Correlation results for the *Environment* category.

Property	Correlation	P-value
Member of an Environmental Group	0.886	0
Think that there is too much concern with the environment	-0.852	0.00001
Closes tap while brushing teeth	0.825	0.00005

a) People who are concerned about the environment, e.g. they join different environmental groups or their actions support environmental conservation (e.g. *closing the tap while brushing teeth*) have strong positive correlation with electricity consumption. This may seem counter-intuitive. However, these people tend to come from affluent/well-educated backgrounds who live in big houses/flats and own multiple appliances (like TV, computer, air-conditioner).

b) People who hold the opinion that *there is too much concern about the environment* have strong negative correlation with electricity consumption. This also may seem counter-intuitive. However, these people often come from the lower middle class, live in small houses and use less electricity.

8) DIGITAL

The Digital category has 142 significant factors covering behaviours regarding the usage of digital technology. Table 11 summarises its results. Overall, we can make the following observations:

a) People who use gambling websites and old people/pensioners who do not use internet at all have negative correlation. Gamblers tend to come from the economically weaker sections of society who consume less electricity [55].

TABLE 11. Correlation results for the *Digital* category.

Property	Correlation	P-value
Regularly research for *Holidays*	0.906	0
Own Tablet Devices (like iPad)	0.891	0
Regularly research for *Hotel reservations*	0.887	0
Internet usage in the last week: *Not at all*	-0.768	0.00032
Internet usage in the last week: *8-19 hours*	0.828	0.00004
Check stocks, shares and investments	0.87	0.00001
Regularly research on internet: *Beer/Wine*	0.821	0.00005
Website regularly visited: *bet365.com*	-0.789	0.00017

b) People who spend long hours on internet or seem to come from affluent backgrounds (i.e. they search for shares/stocks prices, use tablets, research for vacations) have positive correlation with electricity consumption.

9) LEISURE TIME

Based on the Table 12, we can draw the following observations:

TABLE 12. Correlation results for the *Leisure time* category.

Property	Correlation	P-value
Prefer Outdoor Sports	0.888	0
Daily Newspapers: *Daily Star*	-0.875	0
Magazines Read: *Sport*	0.866	0.00001
Interests/Hobbies: *Gambling*	-0.845	0.00002
Restaurants visited most often: *Premium*	0.812	0.00008
Charities: *Overseas development*	0.8	0.00011

a) People who read tabloids or interested in gambling have strong negative correlation. Tabloids are popular among lower socio-economic class [56].

b) People who are interested in sports, visit premium restaurants, or contribute to charity have strong positive correlation. These people often come from affluent backgrounds [57].

10) COMMUNITY SAFETY

We summarise the results in Table 13 and derive the following observations:

TABLE 13. Correlation results for the *Community safety* category.

Property	Correlation	P-value
Have confidence in the police	0.888	0
Incidents reported: Rape	-0.758	0.00042
Incidents reported: physically attacked	-0.747	0.00057
Incidents reported: mugged/robbed	-0.627	0.00707

a) Neighborhoods dominated by wealthy people (i.e. Affluent Achievers in Table 2) report high confidence levels in the police [58]. They have strong positive correlation with electricity consumption.

b) Homes in unsafe neighbourhoods have negative correlation with electricity consumption.

11) CONTACT

Contact specifies the people's preferred channels (phone/SMS, mail, email, or a personal visit) via which they want unacquainted people to approach them (e.g. by marketing agencies). Meaning, these preferences are not applicable to their family, friends, colleagues, or other people in their social circle.

Table 14 summarises the significant correlations and we can obtain the following conclusions:

TABLE 14. Correlation results for the *Contact* category.

Property	Correlation	P-value
Preferred Channel: *Phone or SMS*	-0.755	0.00045
Preferred Channel: *Mail*	0.683	0.00249

a) People belonging to Urban Adversity and Financially Stretched groups (as shown in Table 2) prefer unknowns to contact them by phone/SMS. These people belong to the lower levels of social stratification, consume less electricity and thus, show a strong negative correlation.

b) People belonging to Rising Prosperity and Affluent Achievers groups, do not prefer unknown people contacting them via phone/SMS or surprise visits, instead, they prefer mails. They show a positive correlation with electricity consumption and belong to the higher levels of socially stratified groups.

12) FAMILY STRUCTURE

We can make the following observations based on the results in Table 15:

TABLE 15. Correlation results for the *Family* category.

Property	Correlation	P-value
Lone parent family with dependent children	-0.693	0.00205
Couple family no children	0.635	0.00615

a) Family structure (Couple family with no children) is more prevalent in affluent families residing (Affluent Achievers in Table 2) in cities and towns.

b) Family structure (Lone parent family with dependent children) is frequently found in the *Urban Adversity* group [59]. This group consumes less electricity consumption and reveals a negative correlation.

13) HEALTH

Based on our results in Table 16, we can draw the following conclusions:

a) *Affluent Achievers* and *Rising Prosperity* groups are health conscious, and take regular exercise [60]. These groups have strong positive correlation with electricity consumption.

TABLE 16. Correlation results for the *Health* category.

Property	Correlation	P-value
Takes regular exercise	0.874	0
Eats vegetables 3 or less days per week	-0.838	0.000029
Eats fruit 3 or less days per week	-0.811	0.00008
Current Cigarette Smoker	-0.743	0.00063

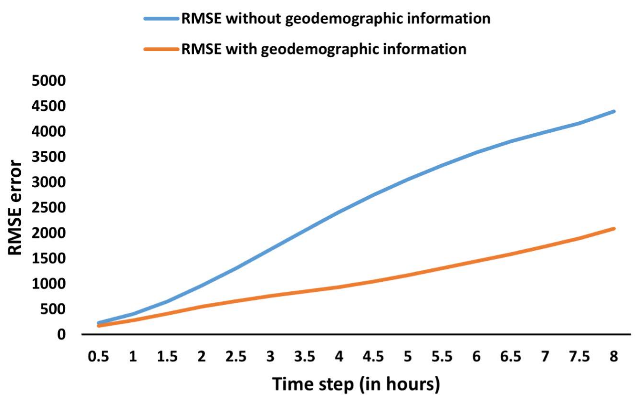

FIGURE 5. Influence of geodemographic information on encoder-decoder LSTM model.

b) The *Urban Adversity* group pay less attention to their physical/mental well-being, neglect their diet and resort to high consumption of tobacco/alcohol [60]. They have strong negative correlation with electricity consumption.

We identified 354 factors that have high correlation with electricity consumption. These factors cover consumer behaviors/habits such as, usage of digital technology, opinions on the environment, newspaper readership, and weekly tobacco/alcohol consumption. Overall, people who exhibit behaviours/habits that are prevalent in the high socio-economic class seem to consume more electricity. Energy companies can use this list of factors to better plan electricity management and distribution.

B. RESULTS RQ2: HOW DOES THE GEODEMOGRAPHIC FACTORS IMPACT THE ACCURACY OF A FORECASTING MODEL?

In section IV (B), we presented the encoder-decoder LSTM model with geodemographic information. We measure its accuracy using the following formulae [61]:

$$R.M.S.E. = \sqrt{\frac{\sum_{i=1}^{n}(y_{predicted} - y_{actual})^2}{n}}$$

where $y_{predicted}$, y_{actual} and n denote the predicted value, the actual value, and the number of observations, respectively. The RMSE values across the different time-steps are shown in the Figure 5. The x-axis shows future time-steps with 0 denoting the present time, i.e. 1 denotes the 1st hour in future. The y-axis shoes the RMSE errors.

The blue curve indicates the model's performance without the geodemographic factors, and the orange curve denotes its performance with geodemographic factors. We observe that the error increases gradually for both the curves, i.e. we can predict the first time-steps with the highest accuracy. However, as we predict further into the future, the accuracy drops. The increase in the RMSE error shows this drop. However, the rate of increase of error is higher for the blue curve. The orange curve is consistently below the blue curve. It suggests that the model that includes the geodemographic factors consistently outperforms the model without the geodemographic factors.

This finding could be beneficial for utility companies that employ different forecasting techniques to estimate future demand. Using geodemographic factors, they can predict the demand more accurately and better arrange for supply.

We find that including geodemographic factors improve the accuracy of our machine-learning based forecasting model. The model that uses geodemographic factors greatly outperforms the one that does not when forecasting longer into the future.

VI. CONCLUSION AND FUTURE WORK

This paper explores the significance of geodemographic factors by asking two research questions. Firstly, it identifies 354 factors that impact the power consumption of a home. Secondly, it suggests an energy forecasting model using geodemographic factors. The approach suggested in this paper could benefit energy companies. Both research questions show that gedemographic factors are important for understanding and forecasting electricity consumption. Electricity companies can incorporate the geodemographic factors to implement better energy management strategies. We have also shown that geodemographic factors can be used to improve the accuracy of a forecasting model. This will further help the companies to estimate their consumer demand more precisely. Companies can use these factors to forecast consumer demand and achieve the load-supply balance.

In the future, we plan to extend this study and explore other methods to analyze the socio-economic and demographic factors that affect energy consumption as follows:

(1) Explore and analyze datasets of different regions, and compare the influencing features on energy consumption pertaining to specific geographic locations, e.g., countries, urban centers, rural areas, etc. We believe the factors that affect energy consumption behavior vary from one place to another based on the different socio-economic and demographic features. (2) Examine the impact of demographic features on various forecasting models and compare them. One direction would be to study forecasting ensemble models that adapt to the influence of the demographic factors to enhance the accuracy of the model. (3) Examine data analytics methods such as [62] to extract features from unstructured data using unsupervised machine learning techniques to identify

features and behavior patterns that are otherwise hidden or not included in structured datasets. (4) Compare the energy consumption of the different geodemographic groups across seasons.

REFERENCES

[1] International Energy Agency. (Jul. 2020). *Electricity Information 2019—Analysis*. [Online]. Available: https://www.iea.org/reports/electricity-information-overview

[2] *U.S. Energy Information Administration—EIA—Independent Statistics and Analysis*.

[3] J. Contreras, R. Espinola, F. J. Nogales, and A. J. Conejo, "ARIMA models to predict next-day electricity prices," *IEEE Trans. Power Syst.*, vol. 18, no. 3, pp. 1014–1020, Aug. 2003.

[4] K. Kavaklioglu, "Modeling and prediction of Turkey's electricity consumption using support vector regression," *Appl. Energy*, vol. 88, no. 1, pp. 368–375, 2011.

[5] B. R. Szkuta, L. A. Sanabria, and T. S. Dillon, "Electricity price short-term forecasting using artificial neural networks," *IEEE Trans. Power Syst.*, vol. 14, no. 3, pp. 851–857, Aug. 1999.

[6] R. V. Jones, A. Fuertes, and K. J. Lomas, "The socio-economic, dwelling and appliance related factors affecting electricity consumption in domestic buildings," *Renew. Sustain. Energy Rev.*, vol. 43, pp. 901–917, Mar. 2015.

[7] *Acorn User Guide*. Accessed: 2020. [Online]. Available: https://acorn.caci.co.U.K./downloads/Acorn-User-guide.pdf

[8] *U.K. Power Networks*. Accessed: 2022. [Online]. Available: https://data.london.gov.U.K./dataset/smartmeter-energy-use-data-in-london-households

[9] C. Yuan, S. Liu, and Z. Fang, "Comparison of China's primary energy consumption forecasting by using ARIMA (the autoregressive integrated moving average) model and GM(1,1) model," *Energy*, vol. 100, pp. 384–390, Apr. 2016.

[10] D. Akay and M. Atak, "Grey prediction with rolling mechanism for electricity demand forecasting of Turkey," *Energy*, vol. 32, no. 9, pp. 1670–1675, Sep. 2007.

[11] K. Boran, "The box Jenkins approach to forecast net electricity consumption in Turkey," *Energy Sour., A: Recovery, Utilization, Environ. Effects*, vol. 36, no. 5, pp. 515–524, Mar. 2014.

[12] V. Bianco, O. Manca, and S. Nardini, "Electricity consumption forecasting in Italy using linear regression models," *Energy*, vol. 34, no. 9, pp. 1413–1421, Sep. 2009.

[13] J. F. M. Pessanha and N. Leon, "Forecasting long-term electricity demand in the residential sector," *Proc. Comput. Sci.*, vol. 55, pp. 529–538, Jan. 2015.

[14] F. Tanrisever, K. Derinkuyu, and M. Heeren, "Forecasting electricity infeed for distribution system networks: An analysis of the Dutch case," *Energy*, vol. 58, pp. 247–257, Sep. 2013.

[15] L. Di Persio, A. Cecchin, and F. Cordoni, "Novel approaches to the energy load unbalance forecasting in the Italian electricity market," *J. Math. Ind.*, vol. 7, no. 1, pp. 1–15, Dec. 2017.

[16] L. Di Persio and I. Perin, "An ambit stochastic approach to pricing electricity forward contracts: The case of the German energy market," *J. Probab. Statist.*, vol. 2015, Oct. 2015, Art. no. 626020.

[17] F. McLoughlin, A. Duffy, and M. Conlon, "A clustering approach to domestic electricity load profile characterisation using smart metering data," *Appl. Energy*, vol. 141, pp. 190–199, Mar. 2015.

[18] M. Kankal, A. Akpınar, M. I. Kömürcü, and T. S. Özsahin, "Modeling and forecasting of Turkey's energy consumption using socio-economic and demographic variables," *Appl. Energy*, vol. 88, pp. 1927–1939, May 2011.

[19] N. Cetinkaya, "Long-term electrical load forecasting based on economic and demographic data for Turkey," in *Proc. IEEE 14th Int. Symp. Comput. Intell. Informat. (CINTI)*, Nov. 2013, pp. 219–223.

[20] Y. E. Unutmaz, A. Demirci, S. M. Tercan, and R. Yumurtaci, "Electrical energy demand forecasting using artificial neural network," in *Proc. 3rd Int. Congr. Human-Comput. Interact., Optim. Robotic Appl. (HORA)*, Jun. 2021, pp. 1–6.

[21] A. Deka, N. Hamta, B. Esmaeilian, and S. Behdad, "Predictive modeling techniques to forecast energy demand in the United States: A focus on economic and demographic factors," *J. Energy Resour. Technol.*, vol. 138, no. 2, Mar. 2016.

[22] A. Y. Davankov, D. Y. Dvinin, and Y. A. Postnikov, "Methodological tools for the assessment of ecological and socio-economic environment in the region within the limits of the sustainability of biosphere," *Ekonomika Regiona*, vol. 4, no. 1, pp. 1029–1039, Dec. 2016.

[23] A. Y. Davankov, D. Y. Dvinin, and Y. A. Postnikov, "Methodological tools for the assessment of ecological and socio-economic environment in the region within the limits of the sustainability of biosphere," *Economy Region*, vol. 4, pp. 1029–1039, Dec. 2016.

[24] J. Mei, Y. Goude, G. Hebrail, and N. Kong, "Spatial estimation of electricity consumption using socio-demographic information," in *Proc. IEEE PES Asia–Pacific Power Energy Eng. Conf. (APPEEC)*, Oct. 2016, pp. 753–757.

[25] A. Sancho-Tomás, M. Sumner, and D. Robinson, "A generalised model of electrical energy demand from small household appliances," *Energy Buildings*, vol. 135, pp. 350–366, Jan. 2017.

[26] G. P. Saha and J. Stephenson, "An evaluation of residential energy conservation strategies in New Zealand," *Energy*, vol. 5, no. 5, pp. 445–450, May 1980.

[27] B. Malmberg, "Future forest trends: Can we build on demographically based forecasts?" in *The Future Use of Nordic Forests*. Springer, 2015, pp. 25–42.

[28] L. Roy and E. Guimond, "Demography perspectives and forecasts of the demand for electricity," *Cahiers Quebecois de Demographie*, vol. 24, no. 1, pp. 87–108, 1995.

[29] E. I. Villacis, E. Yontz, R. Musa, D. McElroy, and K. Roger, "Integrated regional economic and demographic modeling system for developing long-range forecast inputs to sectoral energy and peak demand forecasting models," in *Proc. Amer. Power Conf.*, 1986, pp. 382–398.

[30] D. Brounen, N. Kok, and J. M. Quigley, "Residential energy use and conservation: Economics and demographics," *Eur. Econ. Rev.*, vol. 56, no. 5, pp. 931–945, Jul. 2012.

[31] A. S. Permana, N. A. Aziz, and H. C. Siong, "Is mom energy efficient? A study of gender, household energy consumption and family decision making in Indonesia," *Energy Res. Social Sci.*, vol. 6, pp. 78–86, Mar. 2015.

[32] C. Kofod and E. Piano, "Different everyday lives: Different patterns of electricity use1," Tech. Rep., 2004.

[33] J. Singh, S. S. Mantha, and V. M. Phalle, "Characterizing domestic electricity consumption in the Indian urban household sector," *Energy Buildings*, vol. 170, pp. 74–82, Jul. 2018.

[34] R. V. Jones and K. J. Lomas, "Determinants of high electrical energy demand in U.K. homes: Socio-economic and dwelling characteristics," *Energy Buildings*, vol. 101, pp. 24–34, Aug. 2015.

[35] W. Kleebrang, P. Bunditsakulchai, and W. Wangjiraniran, "Household electricity demand forecast and energy savings potential for Vientiane, Lao PDR," *Int. J. Sustain. Energy*, vol. 36, no. 4, pp. 344–367, Apr. 2017.

[36] A. M. Alonso, F. J. Nogales, and C. Ruiz, "A single scalable LSTM model for short-term forecasting of massive electricity time series," *Stat*, vol. 1050, p. 6, Mar. 2020.

[37] S. Singh, A. Yassine, and R. Benlamri, "Consumer segmentation: Improving energy demand management through households socio-analytics," in *Proc. IEEE Int. Conf. Dependable, Autonomic Secure Comput., Int. Conf. Pervasive Intell. Comput., Int. Conf. Cloud Big Data Comput., Int. Conf. Cyber Sci. Technol. Congr. (DASC/PiCom/CBDCom/CyberSciTech)*, Aug. 2019, pp. 1038–1045.

[38] (2019). *Smart Meters in London*. [Online]. Available: https://www.kaggle.com/jeanmidev/smart-meters-in-london

[39] M. Morgan and S. Chinn, "ACORN group, social class, and child health," *J. Epidemiol. Community Health*, vol. 37, no. 3, pp. 196–203, 1983.

[40] J. P. Singh, O. Alam, and A. Yassine, "Response of socioeconomic groups to dynamic and static tariffs of electricity," in *Proc. IEEE Int. Conf. Dependable, Autonomic Secure Comput., Int. Conf. Pervasive Intell. Comput., Int. Conf. Cloud Big Data Comput., Int. Conf. Cyber Sci. Technol. Congr. (DASC/PiCom/CBDCom/CyberSciTech)*, Edmonton, AB, Canada, Oct. 2021, pp. 276–281.

[41] *Caci*. Accessed: 2022. [Online]. Available: https://www.caci.com/about-caci

[42] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[43] K. Cho, B. van Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder–decoder for statistical machine translation," 2014, *arXiv:1406.1078*.

[44] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE Trans. Neural Netw.*, vol. 5, no. 2, pp. 157–166, Mar. 1994.
[45] X. Xiao, L. Wang, K. Ding, S. Xiang, and C. Pan, "Deep hierarchical encoder–decoder network for image captioning," *IEEE Trans. Multimedia*, vol. 21, no. 11, pp. 2942–2956, Nov. 2019.
[46] S. H. Park, B. Kim, C. M. Kang, C. C. Chung, and J. W. Choi, "Sequence-to-sequence prediction of vehicle trajectory via LSTM encoder–decoder architecture," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2018, pp. 1672–1678.
[47] J. P. Singh. (Apr. 2021). *Results IEEE*. [Online]. Available: https://github.com/jjitenderpalsingh/AMOD_program/blob/main/Results_IEEE.csv
[48] A. González González, J. García-Sanz-Calcedo, and D. Rodríguez Salgado, "Evaluation of energy consumption in German hospitals: Benchmarking in the public sector," *Energies*, vol. 11, no. 9, p. 2279, Aug. 2018.
[49] J.-P. Lévy and F. Belaïd, "The determinants of domestic energy consumption in france: Energy modes, habitat, households and life cycles," *Renew. Sustain. Energy Rev.*, vol. 81, pp. 2104–2114, Jan. 2018.
[50] *NS-Sec Classification*. Accessed: 2022. [Online]. Available: https://www.ons.gov.U.K./methodology/classificationsandstandards/otherclassifications/thenationalstatisticssocioeconomicclassificationnssecrebasedonsoc2010
[51] J. I. Dingel and B. Neiman, "How many jobs can be done at home?" *J. Public Econ.*, vol. 189, Sep. 2020, Art. no. 104235.
[52] M. I. Usaini and N. B. Abubakar, "The impact of parents' occupation on academic performance of secondary school students in Kuala Terengganu," *Multilingual Academic J. Educ. Social Sci.*, vol. 3, no. 1, pp. 112–120, 2015.
[53] A. Sullivan, S. Parsons, F. Green, R. D. Wiggins, and G. Ploubidis, "Elite universities, fields of study and top salaries: Which degree will make you rich?" *Brit. Educ. Res. J.*, vol. 44, no. 4, pp. 663–680, Aug. 2018.
[54] J. Scheerder, S. Vos, and M. Taks, "Expenditures on sport apparel: Creating consumer profiles through interval regression modelling," *Eur. Sport Manage. Quart.*, vol. 11, no. 3, pp. 251–274, Jun. 2011.
[55] M. Aasved, *The Sociology of Gambling*. vol. 2. Charles C Thomas Publisher, 2003.
[56] S. Johansson, *Reading Tabloids: Tabloid Newspapers Their Readers*. Huddinge, Sweden: Södertörns Högskola, 2007.
[57] K. McDonald, "How do philanthropists select a charity?" in *Literature Review Summary Report*. 2016, p. 19.
[58] E. Alda, R. R. Bennett, and M. S. Morabito, "Confidence in the police and the fear of crime in the developing world," *Policing: Int. J. Police Strategies Manage.*, vol. 40, no. 2, pp. 366–379, 2017.
[59] B. Schlesinger, "Lone-parent families in cross-cultural perspectives: Ethnic and immigration issues," *Int. J. Sociol. Family*, vol. 26, no. 1, pp. 89–105, 1996.
[60] J. Wardle, "Socioeconomic differences in attitudes and beliefs about healthy lifestyles," *J. Epidemiol. Community Health*, vol. 57, no. 6, pp. 440–443, Jun. 2003.
[61] A. G. Barnston, "Correspondence among the correlation, RMSE, and Heidke forecast verification measures; refinement of the Heidke score," *Weather Forecasting*, vol. 7, pp. 699–709, Dec. 1992.
[62] M. J. Hasan, M. Sohaib, and J.-M. Kim, "An explainable AI-based fault diagnosis model for bearings," *Sensors*, vol. 21, no. 12, p. 4070, Jun. 2021.

JITENDER PAL SINGH received the bachelor's degree in computer science and engineering from Guru Jambheshwar University, Hisar, and the master's degree in applied modeling and quantitative methods from Trent University, Peterborough, where his research focused on developing machine learning models. He is currently working as a Machine Learning Engineer, where he is responsible for developing solutions for clients belonging to different domains like insurance, pharma, and retail.

OMAR ALAM is currently an Associate Professor with the Department of Computer Science, Trent University. He has published in premier venues in model-driven engineering and software engineering, such as MODELS, SPE, SLE, ICSR, and SAM. He served as a reviewer and program committee member for various journals and conferences in the field of model-driven engineering and software engineering. His research interest includes software engineering. In particular, he is interested in model-driven engineering, aspect-oriented modeling and software reuse.

ABDULSALAM YASSINE (Member, IEEE) received the B.Sc. degree in electrical engineering from Beirut Arab University, Lebanon, in 1993, and the M.Sc. and Ph.D. degrees in electrical and computer engineering from the University of Ottawa, Canada, in 2004 and 2010, respectively. From 2001 and 2013, he was a member of the Technical Staff in the Wireless Communication Division, Nortel Networks, and later at Alcatel-Lucent, Ottawa, Canada. From 2014 to 2016, he was a Postdoctoral Fellow with the DISCOVER Laboratory, School of Electrical Engineering and Computer Science, University of Ottawa. He is currently an Associate Professor and the Chair of the Software Engineering Department, Lakehead University, Thunder Bay, ON, Canada. He is also the Principal Investigator of several research projects that are funded by the public and private. His research interests include energy informatics, artificial intelligence, blockchain technology, IoT systems, smart cities, and smart grids research and applications.