Energy Economics 81 (2019) 709–727

Contents lists available at ScienceDirect

Energy Economics

journal homepage: www.elsevier.com/locate/eneco

Machine learning in energy economics and finance: A review

Hamed Ghoddusi[a,*], Germán G. Creamer[a], Nima Rafizadeh[b]

[a]School of Business, Stevens Institute of Technology, Hoboken, NJ, USA
[b]Independent Researcher, Tehran, Iran

ARTICLE INFO

Article history:
Received 21 October 2018
Received in revised form 6 March 2019
Accepted 7 May 2019
Available online 21 May 2019

JEL classification:
C1
C5
C8
Q4
C4
C11
C14

Keywords:
Machine learning
Energy markets
Energy finance
Support Vector Machine
Artificial Neural Network
Forecasting
Crude oil
Electricity price

ABSTRACT

Machine learning (ML) is generating new opportunities for innovative research in energy economics and finance. We critically review the burgeoning literature dedicated to Energy Economics/Finance applications of ML. Our review identifies applications in areas such as predicting energy prices (e.g. crude oil, natural gas, and power), demand forecasting, risk management, trading strategies, data processing, and analyzing macro/energy trends. We critically review the content (methods and findings) of more than 130 articles published between 2005 and 2018. Our analysis suggests that Support Vector Machine (SVM), Artificial Neural Network (ANN), and Genetic Algorithms (GAs) are among the most popular techniques used in energy economics papers. We discuss the achievements and limitations of existing literature. The survey concludes by identifying current gaps and offering some suggestions for future research.

1. Introduction

Machine learning(ML) refers to a class of data science models that can learn from the data and improve their performance over time. The roots of ML goes back to the scientific community's interest in 1950s and 1960s in replicating human leaning through computer programs. From this perspective, ML extracts knowledge from data, which can then used for prediction and generating new information. This information reduces uncertainty as it indicates how to solve particular problems. ML is particularly useful in dealing with tasks that cannot be explicitly instructed by an analytic solution, such as image and voice processing, pattern recognition, or complex classification tasks.

The superior performance of ML models in processing, classifying, and forecasting using complex and large-scale data, has made them popular in many areas of the energy industry.[1] A sample list of applications in the energy industry includes oil and gas field exploration (Anifowose et al., 2017), oil and gas processes (Zendehboudi et al., 2018), well diagnosis (Fulford et al., 2016), solar radiation forecasting (Voyant et al., 2017), optimization of reactors (Zeng et al., 2018), wind power forecasting (Heinermann and Kramer, 2016), wind energy systems (Marugán et al., 2018), failure prediction (Gupta et al., 2015), power load forecasting (Jurado et al., 2015), and energy-water nexus (Zaidi et al., 2018).[2]

ML has also been widely used in applications related to the economic and financial analysis of energy markets, such as price prediction and risk management. A comparison between the characteristics of ML and traditional econometric models (e.g., ARIMA and GARCH) reveals some of the reasons for the increasing popularity of ML in

* Corresponding author.
E-mail addresses: hghoddus@stevens.edu (H. Ghoddusi), gcreamer@stevens.edu (G.G. Creamer), nima.rafizadeh90@gmail.com (N. Rafizadeh).

[1] Non-expert readers can consult Appendix A for a brief overview of ML methods.
[2] ML has become more popular due to improvements in three major domains: 1) new efficient algorithms; 2) availability of large-scale data sets (aka big data); and 3) improved hardware performance (in particular GPU computing).

https://doi.org/10.1016/j.eneco.2019.05.006
0140-9883/

energy economics. The advantage of ML methods over methods proposed by classical statistics/econometrics is that ML algorithms can manage a large amount of structured and unstructured data and make fast decisions or forecasts. Such superior performance is possible as ML models do not make any pre-specified assumptions about the functional form of the equation, the interaction between variables, and the statistical distribution of parameters. The ML methods, instead, focus on making accurate predictions for some outcome variables given other variables.

Despite widespread use of ML in energy economics, to the best of our knowledge, no review paper exists that systematically reviews the existing literature. Given the benefits of such a study, this paper aims at providing a review of recently published articles in various Energy-related and Economics-related journals. Our review focuses on four major questions: 1) Which methods have been frequently used in energy economics and which ones are underutilized, and why?; 2) Where are the popular and under-explored application domains in energy economics and finance?; 3) What can energy economics/finance learn from other fields?; 4) What are the opportunities for future research in these areas?

Our review reveals that crude oil and power price forecasting are by far the most popular applications in energy economics/finance. As far as methods are concerned, ANN has traditionally been a widely-used technique, following by recent interest in SVM. We note that compared to other fields of science, Deep Learning (DL) has been less common in this area. Thus, a broader set of opportunities exists in applying DL to energy economics and energy finance problems.

We are not the first to write a review paper on the applications of ML in energy in general. There are a few survey papers in related or adjacent scientific fields. For example, Voyant et al. (2017) survey ML methods used for predicting solar radiation, Perera et al. (2014) discuss applications of ML for renewable energy integration, and Zemene and Khedkar (2017) compare ML algorithms used to determine customer electric power consumption. Our focus is mainly the Energy Economics and Energy Finance applications; thus, we do not provide an in-depth analysis of ML/AI in Energy Engineering applications (though we do touch on the topic in our overall analysis).[3] Though not specifically focused on ML techniques, Weron (2014) reviews methods used for electricity price forecasting; the paper devotes an important part of the paper to ML techniques under the title of Computational Intelligence (CI). Thus, it can be considered as a complement to our review. Debnath and Mourshed (2018) review forecasting models used for Energy Planning Models (EPMs) and report Artificial Neural Network (ANN) as the most popular method of forecasting. Mullainathan and Spiess (2017) and Athey (2017) provide a non-technical overview and a review of economic/econometric applications of ML. The economics community is mainly concerned with the capacity and limits of ML for theory testing and causal inference. The energy community, on the other hand, is less concerned with that aspect and is more interested in applied topics.

Our review contains the following components. First, we aim at providing a statistical account of the most popular ML techniques used in Energy Economics papers. The second goal is to identify and uncover under-developed areas in energy markets that have the potential for new ML applications. The final goal is to speculate and suggest some future research topics in this area.

This paper is organized as follows. Section 2 introduces the methodology used in this research. We report the statistical findings in Section 3. A detailed account of ML applications in specific domain areas is provided in Section 4. Section 5 critically reviews the existing literature. Section 6 offers suggestions and an outlook for the future.

[3] Energy Finance is a sub-field of Energy Economics. However, we use both terms next to each other because there is a large concentration of papers using ML for predicting the price of traded energy assets (e.g., crude oil). This particular domain of application is sometimes labeled as Energy Finance.

Table 1
List of journals.

Type of journals	Detailed list of journals
Energy journals	Energy, Applied Energy, Energy Economics, Energy Journal, Energy Policy, Resource, and Energy Economics, Renewable Energy, Journal of Futures Markets, Journal of Electricity Markets, Journal of Commodity Markets
Methodology journals	European Journal of Operational Research, Management Science, Operations Research, Journal of Forecasting, International Journal of Forecasting, Quantitative Finance, Plos One

Finally, we provides a quick overview of major ML techniques and concepts in Appendix A.

2. Research methodology

The focus of this paper is on ML applications in the field of energy economics/finance. The boundary between "energy economics/finance" and "energy engineering" is neither rigid nor well-defined. While some topics belong to one area (e.g., reactor optimization is unambiguously an engineering topic, and futures trading strategy is a clear economics subject), there are gray areas between different sub-fields (e.g., power load and price forecasting). We broadly define energy economics/finance applications when the problem is dealing with any form of market-related variables such as prices, investment, firms and consumers' optimal behavior, and public policy.

We follow multiple strategies to identify the existing literature. We first identify journals that may publish papers using ML techniques to energy problems (see Table 1). The journals are selected from the leading energy-related and economics-related journals listed in the SCImago Journal Rank (SJR), Science Citation Index Expanded (SCIE), Social Sciences Citation Index (SSCI), and Arts & Humanities Citation Index (A&HCI).

We also provide a list of keywords used in our search in Table 2. Given the rich set of keywords and journals, and by double checking the citations of highly-cited papers we hope that our comprehensive search provides an unbiased view of the current literature.

Three specialized electrical engineering journals (IEEE Transactions on Power Systems, Electric Power Systems Research, and International Journal of Electrical Power & Energy Systems) contain a large number of papers on the boundary of energy economics and energy systems engineering. We do not include them in the review because other surveys written on power market forecasts (e.g., Weron, 2014) provide more profound and more specialized coverage of those types of papers.

Our search may ignore some papers published in journals outside of the economics mainstream (e.g., local economics or computer science journals or technical conference proceedings). Thus, the numbers that we report are likely to be below the number one may get from standard bibliometric sources such as the Web of Science and Scopus.[4] This is a minor concern for us because our goal is to create an in-depth review of the content and methods of most relevant papers (published in high impact journals) rather than just providing a statistical analysis.

Our review may also miss progress in the applications of ML in the energy industry (e.g., proprietary algorithms and techniques used for trade or risk management purposes) that are not publicly available. Therefore, the reported results may underestimate the level of frontier knowledge in the field.

3. Statistical summary of findings

We first report a high-level statistical summary of our findings. Table 3 presents a list of papers mapped to different application

[4] We double checked our search results with multiple searches in Scopus to make sure no major publication is missing.

Table 2
Keywords.

Type of keyword	List of keywords
Keywords	Machine Learning, Supervised Learning, Unsupervised Learning, Nearest Neighbor, Clustering, Support Vector Machine, Random Forest, Classification Tree, Deep Learning, Convolutional Neural Network, Artificial Neural Network, Recurrent Neural Network, Long-Short Term Memory, Ensemble Methods, Radial Basis Function Network, Kernel-based Extreme Learning, Feed-Forward Deep Network, Genetic Algorithm, Particle Swarm Optimization, Agent-Based Algorithmic Learning, Wavelet-Based Neural Networks, Ensemble Empirical Mode Decomposition, Data Fluctuation Network, Soft Computing, Simulated-Based Neural Network
Energy keywords	Energy, Crude Oil, Natural Gas, Gasoline, Carbon, Electricity, Power, Renewable

areas. This table also summarizes all the techniques used in that particular domain. We immediately observe in Fig. 1 that the price forecast is the largest category of applications.

We report the number of energy-economics related publications in different journals between 2005 and 2018 in Table 4. The overall number and cumulative number are also plotted in Fig. 2.

4. Detailed review: application domains

The existing application of ML methods in energy-economics related papers can be classified into two major groups: 1) predicting energy commodity prices, and 2) predicting/modeling energy consumption/demand. However, many papers use ML methods in a specific energy-economics related context. Most of the papers mainly use either individual ML techniques or hybrid ML/statistical-econometrics techniques.

4.1. Predicting energy prices

Energy commodity price series typically demonstrate complex features such as non-linearity, lag-dependence, non-stationarity, and volatility clustering which make the use of simple traditional models challenging (Cheng et al., 2018b). ML methods may provide superior forecasting performance because they have higher flexibility in handling complex internal dynamics. A vast majority of papers focusing on price predictions either consider crude oil or power price prediction. Predicting natural gas prices are much less frequent. In the case of coal as a major energy source, we have not found any major paper that uses ML techniques for predicting coal price.

4.1.1. Crude oil price forecasting

Papers dealing with forecasting crude oil prices are predominantly based on advanced and hybrid versions of ANNs and in less degree of SVM models. Also, combining multiple methods (ensemble approach) has become more common in recent years.

As one of the early success stories, Moshiri and Foroutan (2006) forecast the daily series of futures oil price using a nonlinear ANN model that outperforms traditional econometric models. Since then the basic model has been extended in multiple directions.

As one of the first attempts to use ensembles of different ML models, Yu et al. (2008) use Empirical Mode Decomposition (EMD) based Neural Network Ensemble Learning (NNEL) paradigm, and Ding (2018) extends this approach to include a final ensemble step to predict oil prices. Jammazi and Aloui (2012) develop a Multilayer BPNN combined with the Harr A Trous Wavelet decomposition (HTW-MBPNN) to improve the prediction of crude oil price. Yu et al. (2017b) propose an ensemble forecasting approach, which combines Sparse Representation (SR) and Feedforward Neural Network (FNN) to forecast crude oil price. The results confirm the superiority of the

Table 3
List of papers.

Applications	List of papers	List of methods
Predicting energy prices	Conejo et al. (2005), Moshiri and Foroutan (2006), Shambora and Rossiter (2007), Yu et al. (2008), Ghaffari and Zare (2009), Koutroumanidis et al. (2009), Nguyen and Nabney (2010), Lin et al. (2010), Movagharnejad et al. (2011), Jammazi and Aloui (2012), Khosravi et al. (2013), Godarzi et al. (2014), Tang et al. (2015a), Young et al. (2014), Yu et al. (2014), Papadimitriou et al. (2014), Chiroma et al. (2015), Zhang et al. (2015), Yu et al. (2015), He et al. (2015), Ghasemi et al. (2016), Zhu et al. (2016), Yu et al. (2016a), Keles et al. (2016), Dudek (2016), Panapakidis and Dagoumas (2016), Wang and Wang (2016), Yu et al. (2016b), Baruník and Malinska (2016), Yu et al. (2017a), Dagoumas et al. (2017), Mirakyan et al. (2017), Zhu et al. (2017), Wang et al. (2017), Čeperić et al. (2017), Han et al. (2017), Singh et al. (2017), Zhao et al. (2017), Yang et al. (2017), Yu et al. (2017b), Safari and Davallou (2018), Zhu et al. (2018), Cheng et al. (2018a), Cheng et al. (2018b), Marcjasz et al. (2018), Lago et al. (2018a), Chai et al. (2018), Lago et al. (2018b), Huang and Wang (2018), Tang et al. (2018b), Tang et al. (2018a), Peng et al. (2018), Zhao et al. (2018), Wang et al. (2018e), Wang et al. (2018f), Bekiroglu et al. (2018), Bento et al. (2018), Dogah and Premaratne (2018), Ding (2018), Sun et al. (2018)	ANN, EMD-NNEL,SC, ARIMA–ANN, RBFN, AAL, AI, SVM, LSSVM–PSO, CEEMD-EELM, DEL, BED, NLSSVM, FFNN, BPNN, DL, KEL, NARNN, GA, GA-LSSVM, VEC-NARNN, NARX, EEMD-RVFL, PPM-KM, LSTM-DE, SR-FNN, VTFM, DFN, AIC-ANN, IDE
Predicting/modeling energy consumption/demand	Pao (2006), Murat and Ceylan (2006), Azadeh and Tarverdian (2007), Sözen et al. (2007), Sözen and Arcaklioglu (2007), Hamzaçebi (2007), Lai et al. (2008), Azadeh et al. (2008), Ünler (2008), Wang et al. (2009), Geem and Roper (2009), Azadeh et al. (2010), Ekonomou (2010), Kavaklioglu (2011), Kankal et al. (2011), Adam et al. (2011), Limanond et al. (2011), Geem (2011), Wang et al. (2011), Tang et al. (2012), Forouzanfar et al. (2012), Yu et al. (2012), Kialashaki and Reisel (2013), An et al. (2013), Liu et al. (2014), Ardakani and Ardehali (2014), Tang et al. (2014), Szoplik (2015), Antanasijević et al. (2015), Tang et al. (2015b), Castelli et al. (2015), Coelho et al. (2016), Ghasemi et al. (2016), Günay (2016), Kaboli et al. (2016), Liu et al. (2016), Panapakidis and Dagoumas (2017), Kaboli et al. (2017), Bassamzadeh and Ghanem (2017), Zeng et al. (2017), Özmen et al. (2018), Li et al. (2018a), Xiao et al. (2018), Chen et al. (2018), Wang et al. (2018d), Alobaidi et al. (2018), Hong et al. (2018), Wang et al. (2018g), Mohan et al. (2018), Li et al. (2018b)	ANN, GA, SNN, SI, SVR, ϵ-SVR, ANFIS, FFNN, EEMD, LSSVR, FA-LSSVR, SD-LSSVR, PSO–GA, IPSO-ANN, BPNN, RBFNN, EL, FFDN, WNN, DMD, GRNN, SSVRE
Model calibration	Amjady and Keynia (2010), Sun et al. (2011), Genc (2017)	MLPNN, NN, SVM
Trading strategies	Moreno (2009), Wang et al. (2016), Pinto et al. (2016)	AI, GA, SVM
Structure of energy systems	Ermis et al. (2007), Sözen (2009), Fang et al. (2013), Wang and Tian (2015), Ju et al. (2016a), Zhang et al. (2016), Ju et al. (2016b), Skiba et al. (2017), Farajzadeh and Nematollahi (2018)	WNN, One-Class SVM, AI, ANN
Policy analysis	Azadeh et al. (2007), Cinar et al. (2010), Mahmoud and Alajmi (2010), Granell et al. (2014), Dagoumas et al. (2017), Skiba et al. (2017), Mashhadi and Behdad (2018), Wang et al. (2018a)	NN, GA, SVM, ANN, LASSO, EEMD-LSSVM-ARIMA
Data management	Li et al. (2017), Zhang et al. (2018)	AL, SVM, AdaBoost

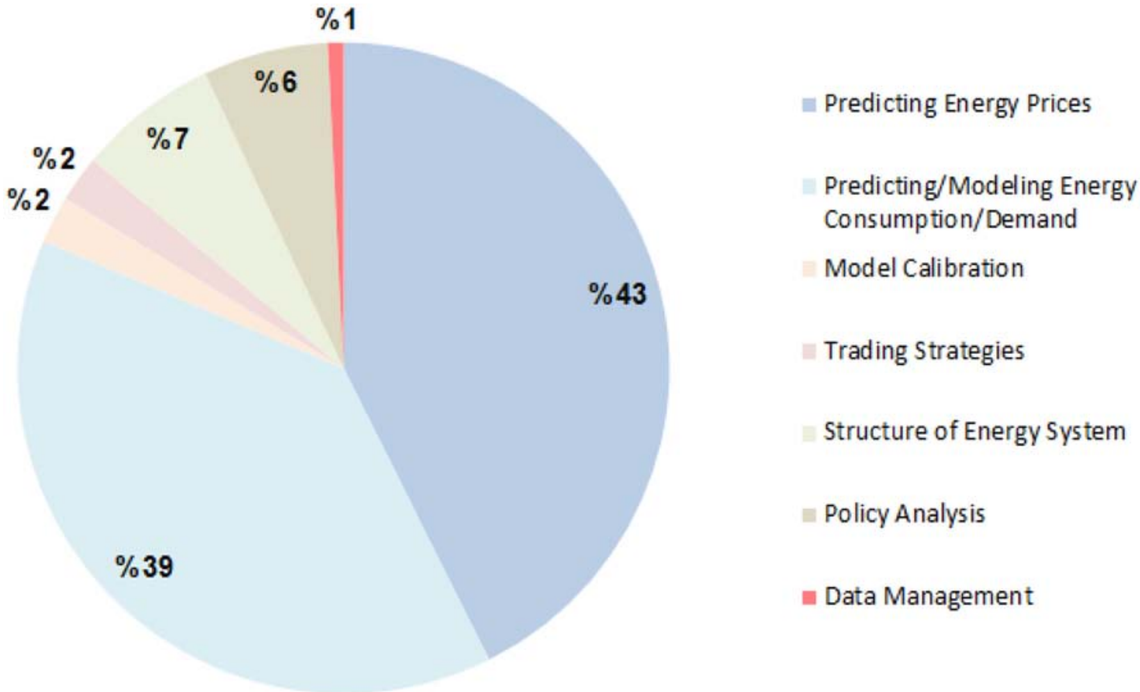


Fig. 1. Relative frequency of application domains.

proposed method over other popular forecasting models and similar ensemble models with other decomposition tools.

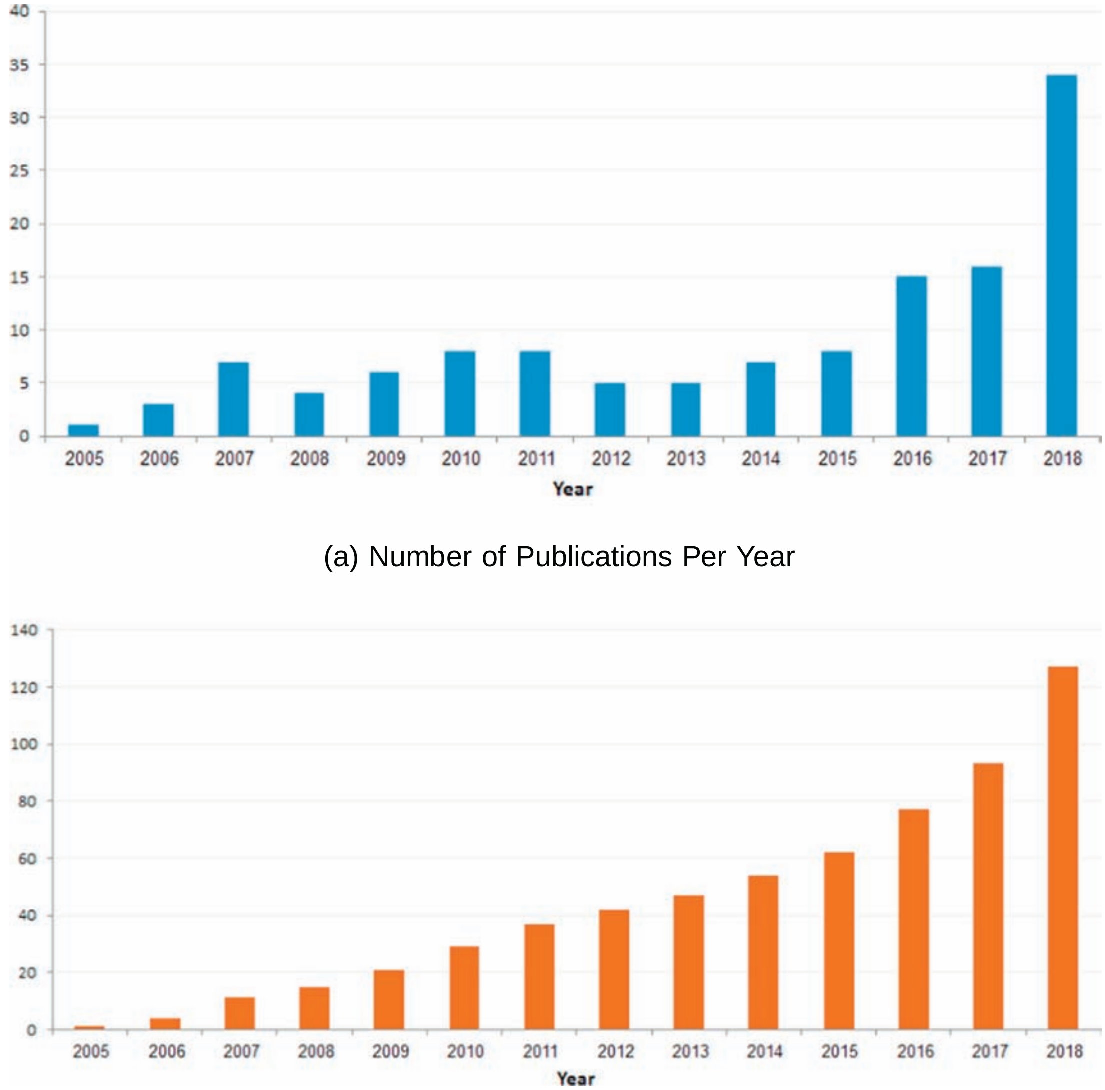
Combining ML and econometric models is also a common approach. For example, Godarzi et al. (2014) develop a dynamic Non-linear Autoregressive model with Exogenous inputs (NARX). Zhang et al. (2015) apply the Ensemble Empirical Mode Decomposition (EEMD) approach to decompose international crude oil price into a series of independent Intrinsic Mode Functions (IMFs) and the residual term. They also develop the Least Square SVM together with the PSO (LSSVM-PSO) method and the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model to predict the nonlinear and time-varying components of crude oil prices, respectively.

Earlier models typically use an autoregressive structure; whereas, while more recent papers use a hybrid approach that can manage a large set of input variables such as demand, supply, the level of stocks, and financial market indicators (Chai et al., 2018). Dogah and Premaratne (2018) examine the exposure of sectoral equity returns to changes in oil risk factors among BRICS (Brazil, Russia, India, China, and South Africa) markets. They combine the VAR model together with the Random Forest technique to provide a framework which overcomes some weaknesses in VAR modeling and help in the selection of oil-risk factors considered.

Yu et al. (2014) present a compressed sensing based learning paradigm by integrating Compressed Sensing-based Denoising (CSD) and an Artificial Intelligence approach (AI), i.e., CSD-AI. Results show that the CSD-AI learning paradigm is significantly superior to other models including single models without the CSD process and hybrid models with other denoising techniques. Wang and Wang (2016) propose an ANN architecture combining multi-layer perceptron and Elman Recurrent Neural Networks (ERNNs), which is a time-varying predictive control system, along with stochastic time

Table 4
Number of energy-economics related publication in each journal between 2005 and2018.

Journal	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	Total
Energy	0	1	0	1	0	3	4	1	2	1	3	4	7	12	39
Applied Energy	0	0	0	0	0	3	2	2	3	2	3	6	4	9	34
Energy Economics	0	0	1	1	1	1	0	1	0	3	2	1	5	10	26
Energy Policy	0	1	6	2	5	1	2	1	0	1	0	1	0	1	21
Energy Journal	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1
Resource and Energy Economics	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Renewable Energy	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Journal of Futures Markets	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Journal of Commodity Markets	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
European Journal of Operations Research	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Management Science	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Operations Research	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Journal of Forecasting	0	0	0	0	0	0	0	0	0	0	0	2	0	0	2
International Journal of Forecasting	1	0	0	0	0	0	0	0	0	0	0	1	0	2	4
Quantitative Finance	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Plos One	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Total	1	3	7	4	6	8	8	5	5	7	8	15	16	34	127



(a) Number of Publications Per Year

(b) Cumulative Number of Publications

Fig. 2. Energy-economics related publications between 2005 and2018.

effective function. They develop ERNN with the ability to keep a memory of recent events to predict future prices.

DL models, as a more advanced version of ANNs, have still not been widely applied for crude oil price forecasting. One exception is Zhao et al. (2017) who use a DL algorithm for crude oil forecasting. The paper combines Stacked Denoising Autoencoders (SDAE) – a DL technique – with bootstrap aggregation (bagging). The bagging routine generates multiple data sets for training a set of base models SDAEs.

Almost all recent papers in this space combine multiple advanced techniques. For example, Tang et al. (2015a) propose an ensemble learning paradigm coupling Complementary Ensemble Empirical Mode Decomposition (CEEMD) and Extended Extreme Learning Machine (EELM) to enhance the prediction accuracy for crude oil price. The results indicate that the model can be used as a promising forecasting tool for complicated time series data with high volatility and irregularity. Zhu et al. (2016) develop an Adaptive Multiscale Ensemble Learning (AMEL) paradigm incorporating EEMD, PSO, and Least Square Support Vector Machines (LSSVM) with kernel function prototype. Cheng et al. (2018a) predict oil price turning points with a log-periodic power law and multi-population GA. Safari and Davallou (2018) combine the Exponential Smoothing Model (ESM), ARIMA, and NARNN in a state-space model framework. Cheng et al. (2018b) propose a hybrid Vector Error Correction and NARNN (VEC-NARNN) model to predict future crude oil price. The results show that the VEC-NARNN model provided better forecasting accuracy than traditional models such as GARCH class models, VAR, VEC, and NARNN models in the multi-step ahead short-term forecast. Yu et al. (2016b) propose a hybrid learning paradigm integrating LSSVR with a hybrid optimization searching approach for the parameters selection in the LSSVR, consisting of the grid method and Genetic Algorithm for crude oil price forecasting. The empirical results demonstrate that the proposed hybrid grid-GA-based LSSVR learning paradigm can outperform its benchmarking models, including some popular forecasting techniques and similar LSSVRs with other parameter searching algorithms, in terms of both prediction accuracy and time-savings. Huang and Wang (2018) combine WNN with a random time effective function. Finally, Zhao et al. (2018) develop the Vector Trend Forecasting Method (VTFM) to oil price forecasting.

Recently, Sun et al. (2018) proposed an interval decomposition ensemble (IDE) learning approach to forecast interval-valued crude oil price. The proposed IDE learning approach significantly outperforms some other benchmark models in term of forecasting accuracy and hypothesis test, which indicates that the strategy of "divide and conquer" can effectively improve the forecasting performance of crude oil price.

4.1.2. Predicting electricity prices

The spot market for power is reduced to a day-ahead, meaning that it does not allow for price bids beyond this horizon. This is due to the verification requirement by system operators to ensure the feasibility of proposed bids within the aggregate transmission constraints (Weron, 2014)). Prediction price in electricity markets is a complicated process because the price is subject to physical constraints on electricity generation and transmission, as well as the potential for market power (Young et al., 2014). The prediction of power prices using ML techniques is one of the oldest applications of these techniques in energy economics. Indeed, in the early 2000s, a wave of papers emerged that were trying to forecast electricity prices using conventional ANN techniques (for a review, see Weron, 2014 and Aggarwal et al., 2009).

Another difference between crude oil and electricity price prediction is that electricity price models use many physical variables – such as temperature, wind speed, and production and transmission capacity – to significantly improve the accuracy of forecasting. ML techniques turn out to be particularly useful because they can handle a large number of inputs even without requiring them to be pre-processed and cleaned.

In general, models used to predict electricity prices are technically more advanced, and they contain complex and novel features. This is partly because electricity price models have typically been developed by engineers, who have a high degree of proficiency in CI methods.

One of the first papers in energy economics is Conejo et al. (2005), who predict the 24 market-clearing prices of the day-ahead electric energy market. The paper compares the performance of time series analysis, ANNs, and wavelets. In this research, wavelet models show similar results to ARIMA models while ANN models underperform. Lin et al. (2010) propose an Enhanced RBFN (ERBFN) for the solving process combining the Radial Basis Function Network (RBFN) and Orthogonal Experimental Design (OED). Khosravi et al. (2013) employ the delta and bootstrap methods to construct electricity price Prediction Intervals (PIs) used for uncertainty quantification. The results suggest that constructed PIs perform better compared to ANN. Papadimitriou et al. (2014) investigate the efficiency of a SVM-based forecasting model for the next-day directional change of electricity prices. The results demonstrate SVM is a powerful method for short-term electricity prices forecast, as it shows a forecast accuracy of 76.12% over a 200 day period. He et al. (2015) propose a Bivariate EMD Denoising (BED) based forecasting methodology to track and predict the electricity price movement. Empirical studies conducted in the Australian electricity markets demonstrate the significant performance improvement of the proposed BED algorithm incorporating the heterogeneous market characteristics.

4.1.2.1. Ensemble methods. Similar to the case of crude oil, ensemble methods are also widely used in the electricity market (e.g. Mirakyan et al., 2017). Ghasemi et al. (2016) propose a hybrid algorithm consisting of three main parts: a Flexible Wavelet Packet Transform (FWPT) to decompose a signal into multiple terms at different frequencies combined with a new feature selection method, Nonlinear Least Square Support Vector Machine (NLSSVM), ARIMA and, finally an Artificial Bee Colony (ABC) algorithm based on time-varying coefficients and stumble generation operator, called TV-SABC, in order to optimize NLSSVM parameters in a learning process.

4.1.2.2. Multi-layer NN. Dudek (2016) establishes a forecasting approach based on a Feed Forward Neural Network (FFNN) for probabilistic electricity price forecasting. The advantage of this approach is that it does not require any special data pre-processing, such as detrending, seasonal adjustment, or decomposition of the time series. Panapakidis and Dagoumas (2016) examine ANN-based models for Day-ahead price forecasting. Wang et al. (2017) first propose a two-layer decomposition technique and then develop a hybrid model based on Fast Ensemble Empirical Mode Decomposition (FEEMD), Variational Mode Decomposition (VMD) and BPNN optimized by a swarm intelligence based algorithm to forecast electricity price. The results show that the model has superior accuracy to both one-step and multi-step ahead forecasting of electricity price. Singh et al. (2017) apply generalized neuron model for forecasting the short-term electricity price of the Australian electricity market to overcome the limitations of the classical ANN model. Yang et al. (2017) propose a hybrid approach combining the wavelet transform, the KELM based on Self-Adapting PSO (SAPSO) and an ARMA. The results demonstrate that the method has a higher level of accuracy, superior generality and practicability to individual methods and other hybrid methods. Bento et al. (2018) propose a method for short-term electricity price forecasting. The method is based on Bat Algorithm (BA), Wavelet Transform and ANNs.

4.1.2.3. Deep Learning. Contrary to the case of oil prices, DL is used more extensively to forecast electricity prices. Lago et al. (2018a) achieve a high degree of accuracy by four different DL models. Lago et al. (2018b) use Deep Neural Networks (DNNs) as a based model to include market integration in electricity price forecasting. Peng et al. (2018) apply Long Short-Term Memory (LSTM) with the Differential Evolution (DE) algorithm to predict electricity prices.

4.1.2.4. Agent-based models(ABM). ABM is a method used to study complex systems characterized by interacting agents and by emergent properties that arise from the interaction of these agents. ABM simulates, through computational programs, the behavior and interaction of these agents that may or may not use machine learning algorithms. This approach has led to a new paradigm in economics known as agent-based computational economics. See Chen (2012) for a comprehensive review and the origins of this area of research.

After the pioneering work of the Santa Fe Institute during the 1990s on ABM and the financial market, ABM has also been extended to other markets such as the electricity market as described by Weidlich and Veit (2008) and Guerci et al. (2010). The key advantage of this approach is that it enables the modeling of strategic behavior over a realistic electricity network without having to fully solve for game theoretic best responses. Young et al. (2014) investigate whether a constructed and calibrated agent-based model using the modified Roth and Erev algorithm (Erev and Roth, 1998) could predict short-run prices from critical inputs in a realistic electricity market. The result shows the model can predict the price to a startling degree of accuracy. Dehghanpour et al. (2018) employ an agent-based framework for studying the behavior of a day-ahead retail electrical energy market with demand response from air conditioning loads.

4.1.3. Predicting other energy commodity prices

A few papers predict other energy commodity prices such as fuelwood (Koutroumanidis et al., 2009), natural gas (Nguyen and Nabney, 2010; Čeperić et al., 2017), and carbon prices (Fan et al., 2015; Zhu et al., 2017; Sun et al., 2016; Zhu et al., 2018) using ML methods.

4.2. Predicting/modeling energy consumption /demand

Understanding the future level of energy demand or consumption is essential for short-run and long-run planning purposes. A wide range of users – from government agencies to local development authorities to financial and trading institutions – are interested in having a realistic forecast of future consumption portfolio. Energy consumption predictions are typically handled by using lagged values of consumption and a set of exogenous socio-economic and technological variables such as GDP per capita, population, and technology trends.

The outcome of the models is either the aggregate energy demand for a particular area, or the decomposition of demand to different energy sources including electricity demand, natural gas demand, and refined products demand. The frequency of forecasting may also differ significantly. While electricity load forecast models need to provide a short-term (e.g., daily) prediction, long-term energy policy models require long-term forecast horizons.

A critical difference between "price" prediction versus "consumption" prediction is that the latter is not subject to market efficiency dynamics. The prediction of consumption has little effect on the actual consumption of agents; whereas, price predictability tends to offset itself by creating opportunities for traders who use that information. Additionally, price data is available at a second and minute frequency; whereas, consumption data is collected monthly. Thus, the number of observations feed to the ML algorithm is much smaller.

4.2.1. Long-range forecasts

ML has been used for long-range forecasts of aggregate and sectoral energy demand. A typical paper in this domain uses a set of potential predictors (e.g., GDP and population) and then builds an ML model to find the relationship between energy consumption and structural variables. ML models might outperform a simple linear regression model if the relationship between the variables is complex. The challenge is that the number of observations in such application tends to be small, weakening the inherent advantage of ML-based methods (see Kaytez et al., 2015 for a comparison of multiple methods.)

In a recent review, Debnath and Mourshed (2018) report that ANN is the most popular technique, among all conventional and CI methods, used in EPMs. Within the sub-category of ML methods, SVM holds ranks second in terms of popularity and frequency of use.

Several papers use ANN models to predict energy consumption at the national level. As a few examples, Sözen and Arcaklioglu (2007) and Sözen et al. (2007) predict the net energy consumption and greenhouse gas emissions in Turkey using sectoral energy consumption using the ANN approach. Similarly, Geem and Roper (2009) suggest an ANN-based model for estimating the energy demand for South Korea. At the sector level, Liu et al. (2016) combine Grey forecasting method and BPNN model to forecast energy consumption in Spanish economic sectors.

Tang et al. (2012) propose a hybrid Ensemble Learning (EL) paradigm integrating EEMD and Least Squares Support Vector Regression (LSSVR) for nuclear energy consumption forecasting. Tang et al. (2014) propose a data-characteristic-driven modeling methodology for nuclear energy consumption forecasting. Li et al. (2018a) forecast oil consumption in China based on a combination of models optimized by AI algorithms. Xiao et al. (2018) predict the nonlinear variation of energy consumption. They use AdaBoost to improve the forecasting accuracy of single nonlinear prediction models, including BPNN, SVR, GA, and RBFNN. Finally, Wang et al. (2018g)) forecast energy consumption in China by building a Self-Adaptive Multi-Verse Optimizer (AMVO) algorithm to optimize the SVM parameters and applied the rolling cross-validation to improve its performance. Wang et al. (2011) employ a seasonal decomposition (SD) based LSSVR ensemble learning model for Chinese hydropower consumption forecasting. Results reveal that the proposed SD-based LSSVR ensemble learning paradigm is a very promising approach for complex time series forecasting with seasonality. Tang et al. (2015b) also propose a hybrid learning paradigm to predict Chinese hydropower consumption, by incorporating the Firefly Algorithm (FA) into LSSVR (FA-LSSVR). The results statistically confirm the superiority of the FA-based LSSVR model to other benchmark models, including existing popular traditional econometric models, AI models and similar hybrid LSSVRs with other popular parameter searching tools, in terms of level and directional accuracy.

4.2.2. Predicting electricity demand

Electricity demand forecasting is one of the traditional domains for ML in energy economics.

Many papers (e.g. Pao, 2006; Azadeh et al., 2008; Lai et al., 2008; Ardakani and Ardehali, 2014; Castelli et al., 2015) have built hybrid models to predict low-frequency energy demand trends in different countries. The overall structure and components of the majority of these papers look similar: a traditional time-series model, SVM, ANN model (in particular a WNN), and more recently feedforward and feedback multi-layer models. GAs and PSO are also often used.

4.2.3. Predicting smart grid load

Compared to the national level forecast, ML is much better suited for short-term electricity demand forecast because it involves a large number of high-frequency observations (e.g., daily or even hourly) on a large set of potential input variables.

There are a few papers taking advantages of ML methods to study electricity load forecasting in smart grids, such as Liu et al. (2014), Coelho et al. (2016), Bassamzadeh and Ghanem (2017), Mohan et al. (2018), Anderson et al. (2011) and Li et al. (2018b).

A prominent study is Ghasemi et al. (2016) who propose a hybrid algorithm for electricity load forecasting in smart grids, classified into three main parts. The first part applies an FWPT to decompose a signal into multiple terms at different frequencies, and a feature selection method that uses Conditional Mutual Information (CMI) and adjacent features to select valuable input data. The next part includes a Multi-Input Multi-Output (MIMO) model based on NLSSVM and ARIMA to model the linear and nonlinear correlation between price and load in for two stages. The final part applies a modified version of Artificial Bee Colony (ABC) algorithm based on time-varying coefficients and a stumble generation operator, called TV-SABC, to optimize NLSSVM parameters in a learning process.

A related line of study is the forecast of the probability of failures of electric grids. Rudin et al. (2012) propose a ranking algorithm to rank electrical grid components based on their likelihood of failure. Special cases of this ranking algorithm are obtained using SVM, boosting and P-norm push for the ranking objective function.

4.2.4. Predicting natural gas demand

Papers that use ML to predict natural gas demand include Azadeh et al. (2010), Szoplik (2015), and Panapakidis and Dagoumas (2017), among others.

Özmen et al. (2018) apply Multivariate Adaptive Regression Splines (MARS) and Conic MARS (CMARS) for natural gas consumption forecasting. The results show that MARS and CMARS methods for natural gas prediction are superior to the existing methods based on ANN and Linear Regression. Collado and Creamer (2016) use an approximate dynamic programming approach to combine a time series method (ARIMA) with two machine learning algorithms (Support Vector Machine and Random Forests) to predict natural gas prices. This method outperforms logistic regression which is the benchmark of this study.

4.2.5. Predicting transport energy demand

Murat and Ceylan (2006) propose an ANN approach based on supervised ANNs for the transport energy demand forecasting using socio-economic and transport related indicators. Limanond et al. (2011) develop Log-Linear Regression (LLR) models and ANN models to project transport energy consumption in Thailand. Geem (2011) develops ANN models to forecast South Korea's transport energy demand. Forouzanfar et al. (2012) propose a Multi-Level Genetic Programming (MLGP) approach for predicting transport energy demand in Iran.

4.2.6. Predicting coal demand

There are few papers that use ML techniques for predicting coal consumption such as Yun-cai (2003), Xuemian and Guohao (2008), and Yang et al. (2014). Jia et al. (2007) employ an SVM model with a multi-input and single output to forecast the coal demand of China from 1980 to 2002. The results show that the SVM predictor has higher precision and greater generalization ability in comparison to RBFNN.

4.3. Other applications

4.3.1. Model calibration

Many papers employ ML methods for model calibration. Sun et al. (2011) identify the parameters for the energy resources demand-supply system by using the ANN method. Amjady and Keynia (2010) propose a new learning algorithm for deregulated electricity markets. Finally, Genc (2017) uses the SVM model to split the data before and after the final crisis of 2009 and then conduct an econometric analysis of the crude oil market.

4.3.2. Trading strategies

Papers on trading strategies belong to a small subset where ML and optimization models are combined. Moreno (2009) presents a model based on fuzzy logic and ML to model trading agents' strategies in the Colombian power market. Wang et al. (2016) select trading rules in the crude oil futures market using GAs. Pinto et al. (2016) use an SVM model for strategic bidding in the electricity markets.

4.3.3. Structure of the energy system

We find that ML methods are commonly used to estimate the structure of energy systems. These papers are aiming at uncovering the impact of various internal and external forces on the level and composition of energy consumption.

Fang et al. (2013) examine the effects of levying a carbon tax on the outcome of energy intensity and economic growth in a four-dimensional energy-saving and emission-reduction system with carbon tax constraints. They use ANN to identify the quantitative coefficients of the actual system. Farajzadeh and Nematollahi (2018) use WNNs to examine the ability of regression models in forecasting energy intensity and its components. Ermis et al. (2007) focus on the analysis of world green energy consumption through ANNs. Ju et al. (2016a) use one-class SVM to provide complementary explanations for nineteen major oil-related countries/regions' macroeconomic effects caused by unexpected oil price changes. Ju et al. (2016b) use AI to predict the co-movements between macroeconomy and oil price shocks in China. Skiba et al. (2017) estimate the distribution of potential energy savings using an ANN model. Sözen (2009) uses the ANN method to estimate Turkey's energy dependency based on basic energy indicators and sectoral energy consumption. Wang and Tian (2015) employ FFNN to build an energy price–energy supply–economic growth dynamic system. Finally, Zhang et al. (2016) propose a dynamic system model of electricity supply-consumption-price. They apply ANN to identify the parameters for establishing the dynamic system of the electricity market.

4.3.4. Policy analysis

A related topic to the modeling of the energy structure is the application of ML for policy analysis.

Azadeh et al. (2007) propose an ML method to measure efficiency as a complementary tool for the typical methods of the efficiency studies in the previous studies. The method finds a stochastic frontier of input-output observational data, and there is no need for explicit assumptions regarding the functional form of the stochastic frontier. Cinar et al. (2010) contribute to the demonstration of using GAs in the development of future energy scenarios and also to the strategic energy studies in Turkey. Granell et al. (2014) apply several ML methods to a data set of electrical power use by 12,000 businesses (in 44 sectors) to evaluate the businesses' benefits or losses of switching tariffs. Dagoumas et al. (2017) take advantage of day-ahead electricity price forecasting using ANN models to study risk management of electricity traders.

Skiba et al. (2017) use ANNs to estimate the distribution of potential energy savings and Mashhadi and Behdad (2018) apply the Least Absolute Shrinkage and Selection Operator (LASSO) regression to analyze the energy consumption of residential units. Mahmoud and Alajmi (2010) quantitatively assess energy conservation due to public awareness campaigns using ANNs.

Wang et al. (2018a) propose a hybrid ML method, i.e., EEMD-LSSVM-ARIMA, to quantitatively analyze and forecast coal overcapacity in China. They suggest several policy recommendations to limit coal overcapacity.

4.3.5. Data management

Results obtained from available data for a particular application depends on the quality and completeness of the data used in the analysis. Therefore, the accuracy of results obtained from a complete data set is most likely to be superior compared to incomplete data sets (Abdella and Marwala, 2005).

Missing data poses various problems in many applications depending on proper access to accurate data. ML is a powerful tool to identify outlier observations or provide a conjecture for missing data as well as to mitigate potential problem related to incomplete data (Abdella and Marwala, 2005; Nelwamondo et al., 2007). Li et al. (2017) employ AI methods in big data-driven models for oil price trends prediction. Zhang et al. (2018) apply several ML methods such as SVM and AdaBoost for mining and analyzing data and conclude that ML algorithms can effectively address energy consumption data gaps.

4.4. Comparison of ML techniques

In Table 5, we provide a short review of the merits and limitations of each major group of ML techniques, as well as, the typical application domains where such a method is recommended.[5]

5. Critical evaluation

5.1. Achievements and advantages of ML techniques

We summarize some achievements that ML methods have delivered over standard econometric models.

5.1.1. Superiority in prediction accuracy

5.1.1.1. Performance evaluation metrics. The literature tends to report somewhat significant improvement in the out-of-sample performance of predictions. Papers use a set of standard performance metrics, such as Root Mean Square Error (RMSE), Mean Absolute

[5] In Table 5, we benefit from the review of the classification of supervised ML techniques conducted by Kotsiantis et al. (2007).

Table 5
Comparison of ML techniques.

Techniques	Merits	Limitations	Typical applications
SVM	– Accuracy in general – Speed of classification – Tolerance to irrelevant attributes – Tolerance to redundant attributes – Tolerance to highly interdependent attributes – Model linear and non-linear problems – Dealing with binary and continuous attributes	– Speed of learning – Tolerance to missing values – Tolerance to noise – Dealing with danger of overfitting – Attempts for incremental learning – Explanation ability – Model parameter handling	– Classification – Regression – Time series forecasting
Decision Trees	– Speed of learning – Speed of classification – Tolerance to missing values – Tolerance to highly interdependent attributes – Explanation ability – Model parameter handling – Dealing with danger of overfitting – Dealing with binary and continuous attributes – Handles attributes with different costs	– Accuracy in general – Tolerance to redundant attributes – Tolerance to highly interdependent attributes – Tolerance to noise – Attempts for incremental learning – Upper bound on error	– Classification – Regression – Credit risk models – Market segmentation – Time series forecasting
Random Forest	– No overfitting – Combination of several tree predictors – Upper bound on error – Robust to noise – Fastest tree-based technique – Not very sensitive to node split variable	– Black box – Explanation ability	– Classification – Regression – Time series forecasting
Adaboost	– No overfitting – Feature selection capability – Minimize bias of the error term – The tree version (alternating decision trees) facilitates interpretation of results – Upper bound on error	– Increase variance of responses – Performance depends on the quality and quantity of data and weak learner – Tolerance to noise	– Classification – Ranking
ANN	– Speed of classification – Dealing with binary/continuous attributes – Attempts for incremental learning	– Tolerance to missing values – Tolerance to irrelevant attributes – Tolerance to redundant attributes – Tolerance to noise – Dealing with danger of overfitting – Explanation ability – Model parameter handling – Black box	– Time-series forecasting
Deep Learning	– Same as ANN	– Black box – Heavy data requirement	– Vision and voice processing – Time series forecasting

Error (MAE), Mean Absolute Percentage Error (MAPE), Directional Accuracy (DA), Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) for regression or forecasting of continuous variables. Papers that use ML algorithms for classification typically include the Matthews correlation coefficient, test error or the accuracy rate (1 − test error) to evaluate the predictive accuracy of ML methods compared to other statistical or econometric techniques. RMSE, MAE, and MAPE indicate the differences between actual and predicted values. Thus, if there are smaller amounts of these metrics, there would be superior forecasting performance. DA that is expressed in percentage provides the correctness of the predicted direction. As DA increases, the accuracy of forecasting becomes stronger. AIC and BIC reward accuracy of the prediction and penalize the complexity of the model. The accuracy rate is expressed in percentage and provides the correctness of the predicted direction using the test sample. As the accuracy rate increases, the forecast quality improves. Matthews correlation coefficient (Matthews, 1975) is a correlation coefficient between the predicted and observed values. It is more appropriate with unbalanced samples as it avoids the bias of accuracy due to data skew.

Among recent papers evaluating the predictive accuracy of methods, Safari and Davallou (2018) is a particular case. As an example, in Table 6, we report results of Safari and Davallou (2018) for crude oil price forecasting based on some of the metrics mentioned above to compare the performance of different methods. It can be seen that based on all the performance metrics, a hybrid (EWH) of ESM, ARIMA, and NAR with time-varying weights using the Kalman-filter is better than the other methods including ARIMA that is a popular traditional econometric model.

5.1.1.2. Statistical tests. There are sets of statistical tests such as Wilcoxon Signed Rank Test (WSRT), Forecast Encompassing Test (FET), and Reality Check (RC) that have been applied for comparing forecasting outcomes by the ML and econometric models.

WSRT is the best known and most widely used nonparametric inference test (Gibbons and Chakraborti, 2011). The null hypothesis of WSRT is that the loss differential series $d(t) = g(e_A(t)) - g(e_B(t))$ has zero median (Diebold and Mariano, 2002), where $e_A(t)$ and $e_B(t)$ are the forecast error series of model A and model B respectively, and $g(.)$ is a loss function (e.g., mean square error). For FET, the Harvey, Leybourne and Newbold (HLN) test (Harvey et al., 1998) is used. The null hypothesis is that the forecast of the model A encompasses that of model B (i.e., all information in the model B is contained in model A). For RC, test for Superior Prediction Ability (SPA) (Hansen, 2005) is performed. It is based on White's reality check test (White, 2000) and is more powerful and less sensitive to poor and irrelevant alternatives. The null hypothesis of the SPA test is that the predictive performance of model A is no better than that of model B.

Zhao et al. (2017) is a good example of recent papers applying all three aforementioned tests to statistically assess some ML models versus some econometric models (i.e., Random Walk (RW) and Markov Regime Switching (MRS)) for crude oil price forecasting. To save space, we only report the results of Zhao et al. (2017) for WSRT in Table 7. In the table, the p-values of relevant statistics between pairs of models are listed. In the last column, all the p-values are smaller than 0.1, except for those in row 7 column 9, indicating that there are significant differences between the forecasts of SDAE-Bagging with that of six competing models, under the confidence level of 90%. Note that it has already been shown in the paper that

Table 6
Results of Safari and Davallou (2018) for crude oil price forecasting.

Performance metrics	ESM	ARIMA	NAR	EWH	GWH	PHM	ZHM
RMSE	3.64	3.68	2.72	3.24	2.72	2.54	3.33
MAE	2.87	3.01	2.34	2.61	2.34	2.23	2.55
MAPE (%)	3.13	3.29	2.55	2.86	2.55	2.44	2.79
DA (%)	63.63	81.81	81.81	72.72	81.81	81.81	72.72

EWH: Hybrid of ESM, ARIMA, and NAR with time-varying weights using the Kalman-filter.
GWH: Hybrid of ESM, ARIMA, and NAR with constant weights using GA.
PHM: Hybrid of ESM, ARIMA, and NAR with equal weights.
ZHM: Zhang (2003)'s Hybrid Model (ARIMA-ANN).

Table 7
Results of WSRT for crude oil price forecasting, conducted by Zhao et al. (2017).

Model	RW	MRS	SVR	SVR-B	FNN	FNN-B	SDAE	SDAE-B
RW	–	0.082	0.102	0.101	0.035	0.037	0.033	0.025
MRS	0.082	–	0.332	0.252	0.694	0.074	0.071	0.027
SVR	0.102	0.332	–	0.443	0.654	0.076	0.086	0.033
SVR-Bagging	0.101	0.252	0.443	–	0.670	0.081	0.085	0.033
FNN	0.035	0.694	0.654	0.670	–	0.017	0.062	0.016
FNN-Bagging	0.037	0.074	0.076	0.081	0.017	–	0.443	0.326
SDAE	0.033	0.071	0.086	0.085	0.062	0.443	–	0.051
SDAE-Bagging	0.025	0.027	0.033	0.033	0.016	0.326	0.051	–

SDAE-Bagging has the highest DA, lowest MAPE and lowest RMSE, which indicate the model has the best prediction performance in comparison to other models; thus, it can also be concluded that the prediction accuracy of SDAE-Bagging is better than that of six models from a statistic point of view.

5.1.2. Ability to handle heterogeneous and large number of inputs

ML models, especially DL models, are not bound by the number of input variables. This feature relieves the model builder from the task of picking a small number of informative input variables. The ML algorithm can accept hundreds of candidate input variables, without being concerned with issues such as co-linearity, and can select the right factors (or features) to be used for forecasting purposes.

ML algorithms typically can manage a diverse group of quantitative and qualitative variables. This capability is especially important for the energy sector as it can use different sources such as text coming from reports or news and mixes with continuous time series to improve the forecast.

5.1.3. Ability to uncover complex relationships

ML models can examine different topologies for possible relationships between input and output data. In the space of traditional models, Bayesian Model Averaging (BMA) techniques allow one to run a family of models on a data set. However, the BMA approach requires that the modeler specifies the structure of each model. ML models do not need any prior specification of the structure. They can also go beyond linear relationships to uncover complex, non-linear, and high-dimensional relationships between many input variables and the desired outcome.

5.1.4. Lower sensitivity to data quality

Data sparsity problem (e.g., data sets with missing observations) has always been an issue for conventional econometric models. Specific techniques (e.g., Fuzzy and GNN models) have been successfully developed within the ML community to allow feeding the model with even less than perfect quality data. For example, Alobaidi et al. (2018) provide insights regarding the ability of ensemble models to produce improved prediction performance with limited data.

5.1.5. Limited data pre-processing needed

Energy time-series data are known to contain specific characteristics such as seasonality, unit-root, structural break, regime switching, and heteroskedasticity that must be resolved before an econometric model is used. ML models, on the other hand, do not require major pre-processing of the data because they can consider those characteristics as additional features of the data. Of course, the performance of ML algorithms improves if the features are already adapted to the requirements of every forecast using meaningful ratios or kernel transformations. However, most of the ML methods might be able to capture these particular features and incorporate them into the final forecasting algorithm.

Only for comparative reasons, when several models are evaluated, the same procedure to pre-process the data should be applied in all cases. For instance, if the growth of a variable, rather than its level, is used in an econometric model because of the unit root problem, the same data transformation should also be applied when an ML model is used. In other words, pre-processing the data is not an option for ML models whereas it is necessary for the competing model. However, once an ML model is selected for a particular forecast, the modeler can pre-process the data or let the ML algorithm find the optimal transformation of the data.

5.1.6. Flexible application: regression, classification or ranking

The output of many supervised ML algorithms can be used for regression, classification or ranking. Depending on the problem under study, the output can change. Many forecasting problems in energy economics can be reformulated as classification problems to simplify the calculation. For instance, an energy trader may prefer to have a strong indication of the direction of the oil price for the next day rather than an imprecise forecast of its price.

For risk management problems, ML methods may rank the likelihood of failure of the different components of an energy grid. In this way, extreme events such as blackouts can be anticipated, and preventive actions can be taken.

5.2. Limitations and challenges

5.2.1. Performance skepticism

If the underlying variable is a traded asset in a highly efficient market, providing an accurate prediction is challenging by definition. In these cases, even fancy and sophisticated ML techniques may not cause a substantial improvement in the forecasting power because the structure of the market prices is already very close to what the model can predict.

For example, Čeperić et al. (2017) claim that in the context of natural gas prices the success of ML results are somehow exaggerated because they demonstrate only slight improvements over the conventional time series approaches.

5.2.2. Over-fitting

Any statistical prediction method is subject to the risk of overfitting; i.e., trying to achieve a very high accuracy using the training data set by even including the sample noise in the fitted model despite receiving a very low accuracy for the test data set.

A typical approach to deal with this problem and compare different algorithms is calculating the test error using cross-validation in which case the test error approximates the true test error of the underlying distribution. However, several ML algorithms such as Random Forests avoid overfitting running its internal cross-validation process.

5.2.3. Generalization capability

Another issue, closely related to overfitting, is the problem of generalization. In short, one main purpose of an ML algorithm is to minimize non-computable expected risk by minimizing the computable empirical risk with the aim of obtaining low generalization gap (Kawaguchi et al., 2017). The difference between empirical risk and expected risk is known as the generalization gap. The generalization gap explains the dependency of a trained model on the unseen training data set (Wickramasinghe et al., 2018). Poor generalization performance is the bottleneck that restricts the application of many ML algorithms such as feedforward neural network (Xie et al., 2018).

Many strategies have been proposed to achieve better generalization. These strategies collectively refer to the term "regularization". Regularization is any modification which we make to the algorithm so that it reduces the generalization error, not the training error (Goodfellow et al., 2016). Regularization can be separated into two categories: implicit and explicit. It should be noted that this categorization is subjective. Both control the effective capacity of the network with the purpose of reducing the overfitting (Hernández-García and König, 2018):

– Implicit regularization: These regularization methods use characteristics of the network architecture, the learning algorithm or the data in order to control the effective capacity of an ML algorithm. Examples: Stochastic Gradient Descent Algorithm, Convolution Layers, Batch Normalization.
– Explicit regularization: Regularization methods which are not structural parts of the network architecture, the algorithms or the data and typically can be added or removed easily. Examples: Weight Decay, Dropout, Data Augmentation, Stochastic Depth.

5.2.4. Black box nature

Some ML methods, such as ANN and SVM, are known to be black box types and therefore are more difficult to understand how the results were obtained compared to more transparent linear regression models. However, some ML algorithms such as decision trees offer the capacity to identify the impact of each feature, and the linear and nonlinear relationships between the features.

The main question that the modeler must solve is if the emphasis is in prediction or interpretation. For policy decisions or scientific research, it is much more important to understand the behavior and the relationship among the different variables while for specific industrial applications such as trading or corporate planning, the emphasis is on prediction.

5.2.5. Requirement of large data sets

More complex models require a larger number of observations to be properly trained and tested. This is a major barrier for the application of ML in areas such as macroeconomics, in which only a limited number of observations (e.g., 50–100) exist.

Dietterich (2000) argues that applying an ensemble learning method can mitigate the risk related to the lack of sufficient data to properly represent the data distribution. Without sufficient data, many hypotheses which give the same training accuracy may be chosen as the learning algorithm. Ensemble methods can thus reduce the risk of selecting the wrong model by aggregating all these candidate models.

Moreover, one of the remedies suggested is the use of economic theory as a guide to variable selection. This approach has been applied to forecasting macroeconomic variables such as inflation and unemployment (e.g., Moshiri et al., 1999; Moshiri and Cameron, 2000; Moshiri and Brown, 2004) and can also be used in cases such as energy consumption when data frequency may not be high.

Finally, a few papers in other fields such as biomedical engineering (Shaikhina et al., 2015) and materials science (Zhang and Ling, 2018) propose techniques to apply ML methods to small data sets, which can be potentially applicable in the burgeoning literature of economics.

5.2.6. Lack of statistical inference

It is common to believe that while econometric methods are primarily concerned with the statistical behavior of the coefficients β in a regression, ML is mainly concerned with the properties of the predicted outcome (i.e., $\hat{y}$). Driven by an interest in theory and in uncovering relationships with variables, economists focus on the statistical significance of each independent variable in a regression. On the other hand, the ML community is mostly seeking to improve the accuracy of forecasting/predicting, using all possible input variables. Therefore, there is little interest in the statistical significance of independent variables (or features in the language of ML). Despite this, in recent years, attempts have been made to develop statistical tests on the results of ML estimates (Demšar, 2006).

5.2.7. Robustness of estimation

The non-linear and dynamic nature of ML algorithms also makes them less robust with regard to specifications and training sets. Special attention to this matter is needed when setting up and calibrating ML algorithms, as even small changes in the data set can lead to different results. Because of this property, it is highly recommended to run multiple specification and robustness tests when ML is used as modeling tools splitting the training data set into a training and validation data set. So, the model is calibrated using the validation data set, and only once the model is calibrated, it is evaluated with the test data set.

5.2.8. Concentration on predicting market prices

As we show in this paper, a significant fraction of papers is focused on predicting market prices. Many ML experts who are not trained as economists do not appreciate the point that the widespread application of ML for predicting traded assets is a different game than predicting physical phenomena such as temperature. Predicting tomorrow's temperature will not change the behavior of the climate; however, forecasting tomorrow's crude oil prices will immediately affect the current prices because it will trigger trade activities.

In this sense, there is an externality from every ML algorithm trying to predict prices on other algorithms. Each successful prediction model makes it more difficult for the next algorithm.

Based on this discussion, there is a limit to the application of ML for common energy prices, or the algorithms become more sophisticated capturing new dynamics of the energy market.

6. Suggestions for future directions

We approach the discussion of future directions from three angles. First, we discuss how the current exercise can be moved a step ahead. We then introduce a few application domains. Finally, we discuss methods used in other fields that can also be applied to the energy industry.

6.1. Improvement in current exercises

An important limitation of the current literature on energy and ML is that some papers emphasize the computer science perspective optimizing computational parameters such as the accuracy rate while the economic or finance intuition might be ignored. Other papers concentrate on the economic or financial perspective without fully exploring the capacity of the algorithms to solve or explain the problem under study.

Another aspect is that most of the papers reviewed use supervised learning algorithms while the application of unsupervised learning methods to different areas such as marketing and risk management is very limited.

Some problems related to marketing and customer management could be solved using clustering techniques such as K-means to segment clients by different demographic or behavioral characteristics and by their likelihood to default or switch companies.

In the area of energy risk management, extreme events could be identified as outliers using principal component analysis or ranking algorithms.

6.2. Opportunities for applying under-utilized methods

1. Theory-driven ML: Combining a theoretical model with ML techniques is a general agenda for the ML research in general. For instance, Gu et al. (2018) evaluate several ML algorithms and conclude that they can improve the description of asset price behavior, in particular, the measurement of risk premia, in comparison to the traditional econometric methods.
   In several domains of energy economics (e.g., electricity markets, demand rebound, optimal investment), the theory provides clear guidance for the empirical analysis. We did not find any major paper that contains both a micro-founded theory model and an ML model. Thus, it is an open question and an opportunity to better integrate the two to achieve a more transparent and robust data-driven analysis.
2. Deep Learning: Recent progress in Deep Learning techniques are practically revolutionizing the entire space of ML. Thanks to their multi-layer structure, DL methods allow the algorithm to process a much larger number of inputs more robustly. They also do not require a prior specification of features. On the downside, ML techniques typically need a large number of input observations and take more time and effort to be calibrated. Thus, they are not recommended for Energy Economics applications that only have a limited number of observations (e.g., monthly data over a few years); however, ML can be beneficial when processing a large set of observations (e.g., intra-day trade data or GIS-based image data).
3. Natural language processing: Processing text and unstructured data to generate quantitative proxies for various entities is becoming a standard method in many disciplines of economics and management. However, this approach has not yet been widely used for energy applications.
4. Social network analysis: The energy market can be studied as a vast network or a combination of many networks. The effect of a change of one network is quickly extended into others such as has been observed during the blackouts. So, an area under-explored and with great potential is the use of social network analysis to evaluate energy networks either physical networks (i.e., smart grids) or human networks (trader networks) to capture risks as well as trends in energy markets (Creamer, 2016; Creamer and Creamer, 2018).

6.3. Unexplored domains

Our review reveals that ML techniques have not yet been fully utilized in the following application domains.

1. Volatility prediction: ML algorithms have been used for modeling the volatility of financial data. Examples include SVM (Tang et al., 2009) and DL (Xiong et al., 2015; Chatzis et al., 2018). However, except for very few examples (e.g., Rode and Fischbeck, 2018), we did not find a significant body of papers devoted to modeling volatility and risk management in energy markets.
   Modeling volatility is both challenging and promising. The challenge is to include the concept of second-moment clustering in a standard ML model. The advantage is that volatility is not subject to market efficiency effect (i.e., volatility will not disappear as a result of prediction).
   Though not directly using ML techniques, papers such as Afkhami et al. (2017) and Wang et al. (2018c) use Internet sentiment to predict energy price volatility. ML techniques are powerful tools in processing such inputs for improving volatility forecasts.
2. Quantifying unstructured and qualitative information: ML provides powerful tools to quantify and classify unstructured and qualitative data for prediction and causal inference purposes. Such methods have been used in the energy domain; however, there is a great deal of potential to further expand the application of ML for combining quantitative, qualitative, and non-structured data. Examples include the conversion of various companies' text-based financial reports and media outcomes to determine their energy and environmental positions, the use of GIS-based data to estimate energy production and consumption patterns, and the use of sentiment and social media information to better predict electricity consumption.
3. Causal and cross-sectional inference: Statistical analysis of cross-sectional and panel data using ML will discover patterns among many different variables. See Grimmer (2015) and Athey (2015) for theoretical discussions on the interface of ML and causal inference.
   We find that time-series models are the dominant form of analysis in the energy economics area. One of the very few exceptions is Hajko (2017) that tries to identify the causal relationship between energy consumption and economic growth in research papers.
   Examples from other domain include student performance evaluation (Masci et al., 2018), credit risk modeling (Zhang et al., 2014), customer purchase decisions (Martínez et al., 2018), and medical diagnosis (Wang et al., 2018b).
4. Trading strategies: Academicians and corporate researchers are frequently using ML techniques to formulate methods to predict the future economic market and devise an effective trading system to maximize the profit (e.g., Choudhury et al., 2014). In particular, over the last few years, a growing number of researchers have been exploiting ML methods in the price and volatility movement of different kinds of financial instruments (see Cavalcante et al., 2016).
   Except for very few papers such as Wang et al. (2016), the literature on the use of ML techniques for forecasting and trading in energy markets is still very limited.

5. ML-based optimization: In our review, we found almost no paper that combines ML-based analyzes with optimization techniques to solve a managerial or policy problem. Such combinations have been used in other areas such as portfolio optimization (Ban et al., 2016) and supply chain (Chi et al., 2007). Analyzing the optimal behavior under strategic interactions (e.g., bidding in electricity markets) is a potentially fruitful area of research.

6.4. Economic and organizational impacts of ML/AI

Similar to other sectors of the economy, the energy industry has been and continues to be affected by the ML revolution. The long-term impact of ML on the structure of the energy industry is yet to be observed and analyzed. It is well understood that ML techniques can both substitute and complement human skills. If they are substitutes, they will replace current human resources and add to the level of automation. In the case of a complementary effect, ML will augment the skills of existing workers. It is not clear which outcome will prevail in different sub-sectors of the energy industry.

These forces may result in the restructuring of the industry and also the emergence of the new forms of market structures and market players. Future research should identify and analyze the impact of ML on areas like energy efficiency, smart networks, cost of producing energy, the efficiency of energy exchanges and markets, and also the labor force in the industry.

In particular, with the increasing climate change concerns, the penetration of renewable sources, and the diffusion of smart grids, ML/AI can be used to improve the predicting and integration of volatile renewable energy resources, balancing energy grids, more profound and quicker understanding of consumers' needs and demand pattern.

To conclude, a non-exhaustive list of open questions, covering both direct as well as indirect effects of ML/AI, is presented here:

1. Untapped potentials of ML to mitigate the intermittent nature of renewable energy sources.
2. The impact of autonomous vehicles on the demand for various forms of energy.
3. The impact of smart buildings and energy management systems on the overall demand, as well as the distribution of intra-day energy demand pattern.[6]
4. Positive (e.g., energy efficiency) and negative (e.g., electricity used for computational purposes) impacts of ML on energy consumption and climate change targets
5. The magnitude of a possible rebound effect as a result of more smart energy systems
6. New skills needed for the current and future energy industry workforce.
7. The optimal combination of human and machine to efficiently manage energy systems.

7. Conclusion

We reviewed a large body of literature on the energy economics/finance applications of various ML methods. We draw the following broad conclusions from the extensive review of the literature: 1) crude oil and power price prediction are the most popular application domains; 2) SVM, ANN, GA, and PSO are the most popular techniques; 3) the majority of papers use structured data (e.g., price time-series). Very few papers use ML to process unstructured and qualitative inputs, and 4) there is an untapped opportunity to use DL methods.

We hope that the application domains and methods discussed in this paper provide the readers with the big picture of the state-of-the-art in the field. In particular, young researchers may leverage the reported status quo to initiate new and innovative research projects.

Similar to other fields of economics, one sizeable next step for the energy economics/finance community is to combine more transparent economic-driven models with the black-box type ML systems. Moreover, the field can benefit from adopting under-utilized state-of-the-art techniques such as DL methods used in other scientific areas. However, a word of caution: given the highly efficient markets for major energy commodities, the price forecasting benefits of using more sophisticated methods can be limited. The benefits can even become smaller if some large players implement such systems and cause the market to become more efficient by better incorporating latent information in the equilibrium prices. This is a serious barrier for many forecasting applications of ML in the energy economics domain. We believe ML can potentially be more useful in modeling and forecasting market risks. In particular, if ML techniques are combined with network models to better capture the propagation of shocks on equilibrium variables, ML can improve the efficacy of risk management measures.

In this study, we do not provide a performance comparison or ranking of different methods, because the accuracy of each model crucially depends on factors such as the type and nature of the problem, available data, and forecasting horizon. Future research can focus on comparing various methods in a particular domain or application.

Appendix A. Brief overview of supervised machine learning algorithms

We provide a brief overview of the major supervised ML algorithms used for energy economics applications as indicated in the papers explored in this survey. Even though some of these papers use very sophisticated algorithms, a majority of them can be reduced to the individual algorithms included in this section or a combination of them. Our goal is not to present a comprehensive tutorial or review of the vast and evolving world of ML techniques. However, in order to provide a conceptual framework – especially for readers who may not be familiar with the technical foundations of ML – we briefly discuss major classes of ML algorithms that are frequently used in Energy Economics/Finance and some of the key concepts in this area.[7]

The description of these algorithms uses a vector of inputs $X = (X_1, X_2, \ldots, X_p)$, and an output Y that may have two or more categories for classification and a continuous value for regression models. We can assume that $Y = f(X) + \epsilon$ where ϵ is a white noise. Typically, every model uses a data set with the variables (X, Y) and the observations $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$, where x_i and y_i represent the set of features i and output i respectively. The data set is at least split between a training and a testing data set.

A.1. Learning

To clarify the concept of learning, we note that ML algorithms typically generate a function $\hat{f}(X)$ that approximates the true function $f(X)$. The function $\hat{f}(X)$ changes in different iterations to reduce the error term $y_i - \hat{f}(x)$ for the instance i. Through this learning process,

[6] As an example, the DeepMind team claims that their ML algorithms could cut electricity usage at Google's data centers by 15%.

[7] Note that despite the complex mathematical structure behind many ML methods, implementing them is getting easier over time. One can download many packages in platforms such as R or Python to run various ML algorithms on the input data.

the algorithm continuously improves, while approximating the true value, based on its errors. The learning process stops when the error term is small enough to be acceptable.

As the learning algorithms have improved and the computational power has substantially increased, many new algorithms reduce the duration of the learning process by the simultaneous combination of many methods (as in the case of hybrid and ensemble methods presented in Section A.5). Some of these algorithms do not learn only through approximations; they also explore a large search space using random variations of the function $\hat{f}(X)$ or with random variations of the training sample (bootstrap) as the expected value of these variations should approximate the true function.

A.2. Support Vector Machine

Support Vector Machine(SVM) is a classification method that classifies data through a hyperplane that maximizes the distance between observations that belong to each category (Vapnik, 1995).

The hyperplane is $\{X : F(X) \doteq X^T\beta + \beta_0 = 0\}$ where $||\beta|| = 1$. The final prediction of the algorithm is $\hat{f}(X) = \text{sign}(F(X))$ (see Hastie et al., 2008).

SVM is one of the main algorithms used to forecast energy time series such as crude oil and electricity. Xie et al. (2006) presented one of the earliest and pioneering papers that used SVM for crude oil pricing.

The use of kernels allows the solution of nonlinear problems with SVM, and also has derived new versions of this algorithm such as Kernel Extreme Learning Machine (KELM). Zhu et al. (2016) used a kernel function with SVM to forecast nonstationary and nonlinear crude oil prices. Subsequently, Yang et al. (2017) utilized KELM in the prediction of electricity prices.

A.3. Artificial Neural Networks

Artificial Neural Network(ANN) was initially introduced by McCulloch and Pitts (1943) as an algorithm that simulates how the brain works based on the connection between neurons. For this reason, this algorithm is known as the connectionist approach. For many years, ANN has been one of the most common ML algorithms used in the industry.

A basic ANN linearly combines the inputs of the vector X into one or more layers of hidden nodes that become the derived features contained in vector V, and the target Y is obtained as a combination of these derived features.

For a regression model, Y represents only one continuous output Y. However, for a classification model, Y includes k classes that are represented as nodes at the top of the network. Typically, the main function $\hat{f}(X)$ is obtained using the following sigmoid function as the activation function $\sigma(h) = \frac{1}{1+e^{-h}}$, that receives the weighted sum of the inputs as input, where W represents the vector of weights applied either to the original inputs X for the hidden nodes or the derived features V for the output nodes: $h = W^TV$.

The most common method to calibrate an ANN model is back-propagation. This method back propagates the output errors to the hidden node errors and corrects the weights until the model converges to a solution.

Predicting power prices is one of the oldest areas of applications for ML techniques. ANN models were widely used to forecast electricity prices in the early 2000s (Amjady, 2006, see Aggarwal et al., 2009 for a review of early papers). Moshiri and Foroutan (2006) and Shambora and Rossiter (2007) are among the first papers using an ANN model to forecast crude oil prices.

Variations of this algorithm have led to the development of many new algorithms such as Radial Basis Function Network (RBFN), Back-Propagation Neural Network (BPNN), Nonlinear Autoregressive Neural Network (NARNN), Wavelet Neural Network (WNN) with random time effective function, Deep Learning (DL) and Adaptive Network-based Fuzzy Inference System (ANBFIS). We briefly describe two of the most essential sub-categories relevant for our review.

A.3.1. Wavelet Neural Network

It is often useful to decompose a time-series to its various frequency components (e.g. short-term and long-term variations). This decomposition makes it possible to produce separate predictions for each frequency layer. In simple terms, a Wavelet Neural Network (WNN) combines the idea of wavelet decomposition and ANN to improve the performance of the learning algorithm via a dimension-reduction layer. The ANN contains an additional hidden layer that first pre-processes the original time-series by producing its wavelet bases and then uses the estimated coefficients to feed the next layers of the network. See Alexandridis and Zapranis (2013) for a detailed description.

Starting with papers such as Liang et al. (2005), WNN technique has been used extensively in the energy economics space. Nguyen and Nabney (2010) are among the first to combine wavelet transformations with ML models to forecast gas price.

A.3.2. Deep Learning

Deep Learning (DL) is quickly becoming the dominant form of ML methods in terms of accuracy and speed. Its multi-layer structure allows the algorithm to model a much higher degree of complexity in the interaction between input variables (LeCun et al., 2015; Hatcher and Yu, 2018).

Compared to other domains (e.g. image and voice recognition, medical diagnosis, Natural Language Processing (NLP), and self-driving cars), DL seems to be less frequently used in energy economics papers. Still, a few recent papers are utilizing this technique. One of the pioneering works is Zhao et al. (2017) who applies a DL algorithm for crude oil forecasting. The article combines a DL technique with bagging. Bagging generates multiple data sets for training a set of base DL models called stacked denoising autoencoders. DL methods will provide a higher degree of flexibility to better address features as including a large number of exogenous and exogenous variables in time-series models.

A.4. Evolutionary Computation (EC)

The objective of EC methods is to offer a heuristic approach to optimize an objective function (e.g. the error or loss) when the search space is large and complex.

Two major methods in supervised learning which we found to be popular in the energy economics community are Genetic Algorithms (GAs) and Particle Swarm Optimization (PSO). We briefly introduce them.

A.4.1. Genetic Algorithm and Genetic Programming

Genetic Algorithm (GA) is a method inspired by Darwin's Theory of Evolution; for this reason, it is part of the emergent or evolutionary approach. This algorithm uses a chromosome whose values are the outcome of Boolean functions as a data structure. These Boolean functions evaluate the values of the features. The chromosomes evolve following the evolution rules such as mutation, crossover and selection until the combination of the chromosomes converges into a solution. The chromosomes and the final output represent a set of decision rules. These rules are interpretable and can change for each problem. In the area of energy economics, Azadeh and Tarverdian (2007) use GA to forecast monthly electrical energy consumption and show that GA outperforms conventional time series.

Genetic Programming (GP) is a variation of GA where computing programs are split into chromosomes and their evolution leads to a new program with an optimal solution. As part of this family of evolutionary algorithms, optimized gene expression programming encodes complex algorithms such as SVM or decision trees in chromosomes which evolve until a solution is found.

A.4.2. Particle Swarm Optimization

Particle Swarm Optimization (PSO) is another evolutionary metaheuristic optimization method which is based on population-based search (aka swarms). PSO and GA share the general approach of evolving, evaluating, and comparing different solutions. However, unlike GA – in which solutions do not have a memory of the past – in the PSO approach, each solution remembers the "path" it has taken and the "velocity" of reaching the current state. Particles also share the experience and learn from the aggregate state of the swarm. PSO does not allow for "mutation" as in GA. Ünler (2008) can be considered a seminal energy economics paper proposing a model based on PSO to forecast the energy demand of Turkey.

A.5. Hybrid and ensemble models

Hybrid and ensemble models combine multiple models to improve the forecasting accuracy and robustness compared to individual models (Kourentzes et al., 2014). The difference is that hybrid models combine entirely different and heterogeneous methods while ensemble algorithms mix the output of multiple versions of the same method or weak learner[8] when it is applied with variations of the data or of the parameters.

A basic hybrid model combines conventional econometric models (e.g ARIMA) with a set of ML-based methods and integrates the individual forecasts using a simple or weighted mean, median or mode. In the space of energy economics, Wang et al. (2005) presented one of the early hybrid models to forecast crude oil price. The model used a combination of ANN and Rule-based Expert Systems (RESs). More recent papers that combine econometrics and ML techniques include Zhang et al. (2015) for crude oil price and Mirakyan et al. (2017) for electricity price prediction.

The three most well-known ensemble methods are bagging, AdaBoost and Random Forests. Bagging averages the results of applying a learning algorithm, such as decision trees, to many bootstrap samples with replacement, AdaBoost increases the weight of the misclassified observations in every iteration, and Random Forests randomly select different samples and features to build many decision trees. We explain the first two as they are used in several algorithms explored in this survey.

A.5.1. Adaboost

Adaboost is a learning algorithm proposed by Freund and Schapire (1997). This algorithm applies a weak base learner to the same training data set with n observations, generates a set of weights $w_1^t, \ldots, w_n^t$, produces a prediction rule h_t, and increases the weights of the misclassified observations. In a new iteration, the weak learner generates a new prediction rule with the new weights and the process repeats. After a fixed number of iterations, the weak prediction rules are combined into a single *strong* rule using a weighted majority vote. The algorithm generates a prediction score $F(x)$ based on the quality of the forecast and the final prediction rule is $\hat{f}(x) = \text{sign}(F(x))$.

A.5.2. Bagging

Bagging or bootstrap aggregation is a method proposed by Breiman (1996) to reduce the variance of single learning algorithms. Using the training data setΥ , this method generates uniform bootstrap samples with replacement $\Upsilon^{(B)}$. Their predictions are obtained by the function $\psi(X, \Upsilon^{(B)})$.

If the final output is a continuous value, the forecast is the average of the predictors of the bootstrap samples as in $\hat{f}(x) = av_B\psi(x, \Upsilon^{(B)})$.

If the predictor is a categorical variable, the forecast $\hat{f}(x)$ is based on the majority vote of $\psi(x, \Upsilon^{(B)})$.

[8] A weak learner is an algorithm with performance slightly better than random guessing.

Appendix B. List of acronyms

AAL	Agent-based Algorithmic Learning
ABC	Artificial Bee Colony
ABM	Agent-Based Model
AI	Artificial Intelligence
AIC	Akaike's Information Criterion
AMEL	Adaptive Multiscale Ensemble Learning
AMVO	Adaptive Multi-Verse Optimizer
ANFIS	Adaptive Network-based Fuzzy Inference System
ANN	Artificial Neural Network
BA	Bat Algorithm
BED	Bivariate EMD Denoising
BMA	Bayesian Model Averaging
BPNN	Back-Propagation Neural Network
CI	Computational Intelligence
CEEMD	Complementary Ensemble Empirical Mode Decomposition
CMARS	Conic Multivariate Adaptive Regression Splines
CMI	Conditional Mutual Information
CNN	Convolutional Neural Network
CSD	Compressed Sensing based Denoising
DE	Differential Evolution
DEL	Decomposition–Ensemble Learning
DFN	Data Fluctuation Network
DL	Deep Learning
DMD	Dynamic Mode Decomposition
DNN	Deep Neural Network
EBP	Error Back-Propagation
EC	Evolutionary Computation
EELM	Extended Extreme Learning Machine
EEMD	Ensemble Empirical Mode Decomposition
EL	Ensemble Learning
EMD	Empirical Mode Decomposition
EPM	Energy Planning Model
ERBFN	Enhanced Radial Basis Function Network
ERNN	Elman Recurrent Neural Network
ESM	Exponential Smoothing Model
FA	Firefly Algorithm
FEEMD	Fast Ensemble Empirical Mode Decomposition
FET	Forecast Encompassing Test
FFDN	Feed-Forward Deep Network
FFMP	Feed-Forward Multi-layer Perceptron
FFNN	Feed-Forward Neural Network
FNN	Feedforward Neural Network
FWPT	Flexible Wavelet Packet Transform
GA	Genetic Algorithm
GARCH	Generalized Autoregressive Conditional Heteroskedasticity
GNN	Grey Neural Network
GP	Genetic Programming
GRNN	General Regression Neural Network
HTW-MBPNN	Multi-layer Back Propagation Neural Network combined with the Harr A Trous Wavelet
IDE	interval decomposition ensemble
IMF	Intrinsic Mode Function
IO	Input-Output
IPSO	Improved Particle Swarm Optimization

KEL Kernel-based Extreme Learning
KELM Kernel Extreme Learning Machine
LASSO Least Absolute Shrinkage and Selection Operator
LLR Log-Linear Regression
LSSVM Least Square Support Vector Machine
LSSVM–PSO Least Square Support Vector Machine and Particle Swarm Optimization
LSSVR Least Squares Support Vector Regression
LSTM Long-Short Term Memory
MAE Mean Absolute Error
MAPE Mean Absolute Percentage Error
MARS Multivariate Adaptive Regression Spline
MIMO Multi-Input Multi-Output
MLGP Multi-Level Genetic Programming
MLPNN Multi-Layer Perceptron Neural Network
MRS Markov Regime Switching
NAR Nonlinear Autoregressive
NARNN Nonlinear Autoregressive Neural Network
NARX Nonlinear Autoregressive model with Exogenous input
NLP Natural Language Processing
NLSSVM Nonlinear Least Square Support Vector Machine
NNEL Neural Network Ensemble Learning
OED Orthogonal Experimental Design
PI Prediction Interval
PPM-KM Product Partition Model-K-Means
PSO Particle Swarm Optimization
RW Random Walk
RBFN Radial Basis Function Network
RBFNN Radical Basis Function Neural Network
RC Reality Check
RES Rule-based Expert System
RMSE Root Mean Square Error
RNN Recurrent Neural Network
RVFL Random Vector Functional Link
SAPSO Self-Adapting Particle Swarm Optimization
SC Soft Computing
SD Seasonal Decomposition
SDAE Stacked Denoising Autoencoders
SI Swarm Intelligence
SPA Superior Prediction Ability
SR Sparse Representation
SNN Simulated-based Neural Network
SSVRE Subsampled Support Vector Regression Ensemble
SVM Support Vector Machine
SVR Support Vector Regression
TV-SABC Artificial Bee Colony algorithm based on Time-Varying coefficients and Stumble generation operator
VEC-NARNN Vector Error Correction and Nonlinear Autoregressive Neural Network
VMD Variational Mode Decomposition
VTFM Vector Trend Forecasting Method
WNN Wavelet Neural Network
WSRT Wilcoxon Signed Rank Test

References

Abdella, M., Marwala, T., 2005. The use of genetic algorithms and neural networks to approximate missing data in database. Computational Cybernetics, 2005. ICCC 2005. IEEE 3rd International Conference on. IEEE., pp. 207–212. https://doi.org/10.1109/ICCCYB.2005.1511574.

Adam, N.B., Elahee, M., Dauhoo, M., 2011. Forecasting of peak electricity demand in Mauritius using the non-homogeneous Gompertz diffusion process. Energy 36 (12), 6763–6769.

Afkhami, M., Cormack, L., Ghoddusi, H., 2017. Google search keywords that best predict energy price volatility. Energy Econ. 67, 17–27.

Aggarwal, S.K., Saini, L.M., Kumar, A., 2009. Electricity price forecasting in deregulated markets: a review and evaluation. Int. J. Electr. Power Energy Syst. 31 (1), 13–22.

Alexandridis, A.K., Zapranis, A.D., 2013. Wavelet neural networks: a practical guide. Neural Netw. 42, 1–27.

Alobaidi, M.H., Chebana, F., Meguid, M.A., 2018. Robust ensemble learning framework for day-ahead forecasting of household based energy consumption. Appl. Energy 212, 997–1012.

Amjady, N., 2006. Day-ahead price forecasting of electricity markets by a new fuzzy neural network. IEEE Trans. Power Syst. 21 (2), 887–896.

Amjady, N., Keynia, F., 2010. A new spinning reserve requirement forecast method for deregulated electricity markets. Appl. Energy 87 (6), 1870–1879.

An, N., Zhao, W., Wang, J., Shang, D., Zhao, E., 2013. Using multi-output feedforward neural network with empirical mode decomposition based signal filtering for electricity demand forecasting. Energy 49, 279–288.

Anderson, R.N., Boulanger, A., Powell, W.B., Scott, W., 2011. Adaptive stochastic control for the smart grid. Proc. IEEE 99 (6), 1098–1115. June.

Anifowose, F.A., Labadin, J., Abdulraheem, A., 2017. Ensemble machine learning: an untapped modeling paradigm for petroleum reservoir characterization. J. Petrol. Sci. Eng. 151, 480–487.

Antanasijević, D., Pocajt, V., Ristić, M., Perić-Grujić, A., 2015. Modeling of energy consumption and related GHG (greenhouse gas) intensity and emissions in Europe using general regression neural networks. Energy 84, 816–824.

Ardakani, F., Ardehali, M., 2014. Long-term electrical energy consumption forecasting for developing and developed economies based on different optimized models and historical data types. Energy 65, 452–461.

Athey, S., 2015. Machine learning and causal inference for policy evaluation. Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 5–6. https://doi.org/10.1145/2783258.2785466.

Athey, S., 2017. The impact of machine learning on economics. Economics of Artificial Intelligence. University of Chicago Press., http://www.nber.org/chapters/c14009.

Azadeh, A., Asadzadeh, S., Ghanbari, A., 2010. An adaptive network-based fuzzy inference system for short-term natural gas demand estimation: uncertain and complex environments. Energy Policy 38 (3), 1529–1536.

Azadeh, A., Ghaderi, S., Anvari, M., Saberi, M., 2007. Performance assessment of electric power generations using an adaptive neural network algorithm. Energy Policy 35 (6), 3155–3166.

Azadeh, A., Ghaderi, S., Sohrabkhani, S., 2008. A simulated-based neural network algorithm for forecasting electrical energy consumption in Iran. Energy Policy 36 (7), 2637–2644.

Azadeh, A., Tarverdian, S., 2007. Integration of genetic algorithm, computer simulation and design of experiments for forecasting electrical energy consumption. Energy policy 35 (10), 5229–5241.

Ban, G.-Y., El Karoui, N., Lim, A.E., 2016. Machine learning and portfolio optimization. Manag. Sci. 64 (3), 1136–1154.

Baruník, J., Malinska, B., 2016. Forecasting the term structure of crude oil futures prices with neural networks. Appl. Energy 164, 366–379.

Bassamzadeh, N., Ghanem, R., 2017. Multiscale stochastic prediction of electricity demand in smart grids using Bayesian networks. Appl. Energy 193, 369–380.

Bekiroglu, K., Duru, O., Gulay, E., Su, R., Lagoa, C., 2018. Predictive analytics of crude oil prices by utilizing the intelligent model search engine. Appl. Energy 228, 2387–2397.

Bento, P., Pombo, J., Calado, M., Mariano, S., 2018. A bat optimized neural network and wavelet transform approach for short-term price forecasting. Appl. Energy 210, 88–97.

Breiman, L., 1996. Bagging predictors. Mach. Learn. 24, 123–140.

Castelli, M., Vanneschi, L., De Felice, M., 2015. Forecasting short-term electricity consumption using a semantics-based genetic programming framework: the South Italy case. Energy Econ. 47, 37–41.

Cavalcante, R.C., Brasileiro, R.C., Souza, V.L., Nobrega, J.P., Oliveira, A.L., 2016. Computational intelligence and financial markets: a survey and future directions. Expert Syst. Appl. 55, 194–211.

Čeperić, E., žiković, S., Čeperić, V., 2017. Short-term forecasting of natural gas prices using machine learning and feature selection algorithms. Energy 140, 893–900.

Chai, J., Xing, L.-M., Zhou, X.-Y., Zhang, Z.G., Li, J.-X., 2018. Forecasting the WTI crude oil price by a hybrid-refined method. Energy Econ. 71, 114–127.

Chatzis, S.P., Siakoulis, V., Petropoulos, A., Stavroulakis, E., Vlachogiannakis, N., 2018. Forecasting stock market crisis events using deep and statistical machine learning techniques. Expert Syst. Appl. 112, 353–371.

Chen, K., Jiang, J., Zheng, F., Chen, K., 2018. A novel data-driven approach for residential electricity consumption prediction based on ensemble learning. Energy 150, 49–60.

Chen, S.-H., 2012. Varieties of agents in agent-based computational economics: a historical and an interdisciplinary perspective. J. Econ. Dyn. Control. 36 (1), 1–25.

Cheng, F., Fan, T., Fan, D., Li, S., 2018a. The prediction of oil price turning points with log-periodic power law and multi-population genetic algorithm. Energy Econ. 72, 341–355.

Cheng, F., Li, T., Wei, Y.-m., Fan, T., 2018b. The VEC-NAR model for short-term forecasting of oil prices. Energy Econ. https://doi.org/10.1016/j.eneco.2017.12.035.

Chi, H.-M., Ersoy, O.K., Moskowitz, H., Ward, J., 2007. Modeling and optimizing a vendor managed replenishment system using machine learning and genetic algorithms. Eur. J. Oper. Res. 180 (1), 174–193.

Chiroma, H., Abdulkareem, S., Herawan, T., 2015. Evolutionary Neural Network model for West Texas Intermediate crude oil price prediction. Appl. Energy 142, 266–273.

Choudhury, S., Ghosh, S., Bhattacharya, A., Fernandes, K.J., Tiwari, M.K., 2014. A real time clustering and SVM based price-volatility prediction for optimal trading strategy. Neurocomputing 131, 419–426.
Cinar, D., Kayakutlu, G., Daim, T., 2010. Development of future energy scenarios with intelligent algorithms: case of hydro in Turkey. Energy 35 (4), 1724–1729.
Coelho, V.N., Coelho, I.M., Coelho, B.N., Reis, A.J., Enayatifar, R., Souza, M.J., Guimarães, F.G., 2016. A self-adaptive evolutionary fuzzy model for load forecasting problems on smart grid environment. Appl. Energy 169, 567–584.
Collado, R., Creamer, G.G., 2016. Time series forecasting with a learning algorithm: an approximate dynamic programming approach. 22nd International Conference on Computational Statistics (COMPSTAT). pp. 111–122.
Conejo, A.J., Contreras, J., Espinola, R., Plazas, M.A., 2005. Forecasting electricity prices for a day-ahead pool-based electric energy market. Int. J. Forecast. 21 (3), 435–462.
Creamer, G.G., 2016. Trading network and systemic risk in the energy market. 2016 International Conference on Behavioral, Economic and Socio-cultural Computing (BESC). pp. 1–6.
Creamer, G.G., Creamer, B., 2018. Emissions abating technology adoption in a coal trading network. Soc. Netw. Anal. Min. 8 (1), 21.
Dagoumas, A.S., Koltsaklis, N.E., Panapakidis, I.P., 2017. An integrated model for risk management in electricity trade. Energy 124, 350–363.
Debnath, K.B., Mourshed, M., 2018. Forecasting methods in energy planning models. Renew. Sust. Energ. Rev. 88, 297–325.
Dehghanpour, K., Nehrir, M.H., Sheppard, J.W., Kelly, N.C., 2018. Agent-based modeling of retail electrical energy markets with demand response. IEEE Trans. Smart Grid 9 (4), 3465–3475.
Demšar, J., 2006. Statistical comparisons of classifiers over multiple data sets. J. Mach. Learn. Res. 7 (Jan), 1–30.
Diebold, F.X., Mariano, R.S., 2002. Comparing predictive accuracy. J. Bus. Econ. Stat. 20 (1), 134–144.
Dietterich, T.G., 2000. Ensemble methods in machine learning. International Workshop on Multiple Classifier Systems. Springer., pp. 1–15. https://doi.org/10.1007/3-540-45014-9_1.
Ding, Y., 2018. A novel decompose-ensemble methodology with AIC-ANN approach for crude oil forecasting. Energy 154, 328–336.
Dogah, K.E., Premaratne, G., 2018. Sectoral exposure of financial markets to oil risk factors in BRICS countries. Energy Econ. 76, 228–256.
Dudek, G., 2016. Multilayer perceptron for GEFCom2014 probabilistic electricity price forecasting. Int. J. Forecast. 32 (3), 1057–1060.
Ekonomou, L., 2010. Greek long-term energy consumption prediction using artificial neural networks. Energy 35 (2), 512–517.
Erev, I., Roth, A.E., 1998. Predicting how people play games: reinforcement learning in experimental games with unique, mixed strategy equilibria. Am. Econ. Rev. 88 (4), 848–881.
Ermis, K., Midilli, A., Dincer, I., Rosen, M., 2007. Artificial neural network analysis of world green energy use. Energy Policy 35 (3), 1731–1743.
Fan, X., Li, S., Tian, L., 2015. Chaotic characteristic identification for carbon price and an multi-layer perceptron network prediction model. Expert Syst. Appl. 42 (8), 3945–3952.
Fang, G., Tian, L., Fu, M., Sun, M., 2013. The impacts of carbon tax on energy intensity and economic growth — a dynamic evolution analysis on the case of China. Appl. Energy 110, 17–28.
Farajzadeh, Z., Nematollahi, M.A., 2018. Energy intensity and its components in Iran: determinants and trends. Energy Econ. 73, 161–177.
Forouzanfar, M., Doustmohammadi, A., Hasanzadeh, S., et al. 2012. Transport energy demand forecast using multi-level genetic programming. Appl. Energy 91 (1), 496–503.
Freund, Y., Schapire, R.E., 1997. A decision-theoretic generalization of on-line learning and an application to boosting. J. Comput. Syst. Sci. 55 (1), 119–139.
Fulford, D.S., Bowie, B., Berry, M.E., Bowen, B., Turk, D.W., et al. 2016. Machine learning as a reliable technology for evaluating time/rate performance of unconventional wells. SPE Econ. Manag. 8 (01), 23–39.
Geem, Z.W., 2011. Transport energy demand modeling of South Korea using artificial neural network. Energy Policy 39 (8), 4644–4650.
Geem, Z.W., Roper, W.E., 2009. Energy demand estimation of South Korea using artificial neural network. Energy Policy 37 (10), 4049–4054.
Genc, T.S., 2017. OPEC and demand response to crude oil prices. Energy Econ. 66, 238–246.
Ghaffari, A., Zare, S., 2009. A novel algorithm for prediction of crude oil price variation based on soft computing. Energy Econ. 31 (4), 531–536.
Ghasemi, A., Shayeghi, H., Moradzadeh, M., Nooshyar, M., 2016. A novel hybrid algorithm for electricity price and load forecasting in smart grids with demand–side management. Appl. Energy 177, 40–59.
Gibbons, J.D., Chakraborti, S., 2011. Nonparametric statistical inference. International Encyclopedia of Statistical Science. Springer., pp. 977–979. https://doi.org/10.1007/978-3-642-04898-2_420.
Godarzi, A.A., Amiri, R.M., Talaei, A., Jamasb, T., 2014. Predicting oil price movements: a dynamic Artificial Neural Network approach. Energy Policy 68, 371–382.
Goodfellow, I., Bengio, Y., Courville, A., Bengio, Y., 2016. Deep Learning. 1. MIT Press Cambridge.
Granell, R., Axon, C.J., Wallom, D.C., 2014. Predicting winning and losing businesses when changing electricity tariffs. Appl. Energy 133, 298–307.
Grimmer, J., 2015. We are all social scientists now: how big data, machine learning, and causal inference work together. PS: Polit. Sci. Polit. 48 (1), 80–83.
Gu, S., Kelly, B.T., Xiu, D., 2018. Empirical Asset Pricing via Machine Learning.
Guerci, E., Rastegar, M.A., Cincotti, S., 2010. Agent-based modeling and simulation of competitive wholesale electricity markets. Handbook of Power Systems II. Springer., pp. 241–286. https://doi.org/10.1007/978-3-642-12686-4_9.
Günay, M.E., 2016. Forecasting annual gross electricity demand by artificial neural networks using predicted values of socio-economic indicators and climatic conditions: case of Turkey. Energy Policy 90, 92–101.
Gupta, S., Kambli, R., Wagh, S., Kazi, F., 2015. Support-vector-machine-based proactive cascade prediction in smart grid using probabilistic framework. IEEE Trans. Ind. Electron. 62 (4), 2478–2486.
Hajko, V., 2017. The failure of Energy-Economy Nexus: a meta-analysis of 104 studies. Energy 125, 771–787.
Hamzaçebi, C., 2007. Forecasting of Turkey's net electricity energy consumption on sectoral bases. Energy Policy 35 (3), 2009–2016.
Han, L., Lv, Q., Yin, L., 2017. Can investor attention predict oil prices? Energy Econ. 66, 547–558.
Hansen, P.R., 2005. A test for superior predictive ability. J. Bus. Econ. Stat. 23 (4), 365–380.
Harvey, D.S., Leybourne, S.J., Newbold, P., 1998. Tests for forecast encompassing. J. Bus. Econ. Stat. 16 (2), 254–259.
Hastie, T., Tibshirani, R., Friedman, J., 2008. The Elements of Statistical Learning: Data Mining, Inference and Prediction. 2nd ed., Springer-Verlag, New York.
Hatcher, W.G., Yu, W., 2018. A survey of deep learning: platforms, applications and emerging research trends. IEEE Access 6, 24411–24432.
He, K., Yu, L., Tang, L., 2015. Electricity price forecasting with a BED (Bivariate EMD Denoising) methodology. Energy 91, 601–609.
Heinermann, J., Kramer, O., 2016. Machine learning ensembles for wind power prediction. Renew. Energy 89, 671–679.
Hernández-García, A., König, P., 2018. Data Augmentation Instead of Explicit Regularization.arXiv preprint. arXiv: 1806.03852.
Hong, T., Jeong, K., Koo, C., 2018. An optimized gene expression programming model for forecasting the national CO_2 emissions in 2030 using the metaheuristic algorithms. Appl. Energy 228, 808–820.
Huang, L., Wang, J., 2018. Global crude oil price prediction and synchronization based accuracy evaluation using random wavelet neural network. Energy 151, 875–888.
Jammazi, R., Aloui, C., 2012. Crude oil price forecasting: experimental evidence from wavelet decomposition and neural network modeling. Energy Econ. 34 (3), 828–841.
Jia, C.-l., Wu, H.-s., Gong, D.-w., 2007. Coal demand prediction based on a support vector machine mode. J. China Univ. Min. Technol. Chin. Ed. 36 (1), 107.
Ju, K., Su, B., Zhou, D., Wu, J., Liu, L., 2016a. Macroeconomic performance of oil price shocks: outlier evidence from nineteen major oil-related countries/regions. Energy Econ. 60, 325–332.
Ju, K., Su, B., Zhou, D., Zhang, Y., 2016b. An incentive-oriented early warning system for predicting the co-movements between oil price shocks and macroeconomy. Appl. Energy 163, 452–463.
Jurado, S., Nebot, À., Mugica, F., Avellana, N., 2015. Hybrid methodologies for electricity load forecasting: entropy-based feature selection with machine learning and soft computing techniques. Energy 86, 276–291.
Kaboli, S.H.A., Fallahpour, A., Selvaraj, J., Rahim, N., 2017. Long-term electrical energy consumption formulating and forecasting via optimized gene expression programming. Energy 126, 144–164.
Kaboli, S.H.A., Selvaraj, J., Rahim, N., 2016. Long-term electric energy consumption forecasting via artificial cooperative search algorithm. Energy 115, 857–871.
Kankal, M., Akpınar, A., Kömürcü, M.İ., Özşahin, T.Ş., 2011. Modeling and forecasting of Turkey's energy consumption using socio-economic and demographic variables. Appl. Energy 88 (5), 1927–1939.
Kavaklioglu, K., 2011. Modeling and prediction of Turkey's electricity consumption using Support Vector Regression. Appl. Energy 88 (1), 368–375.
Kawaguchi, K., Kaelbling, L.P., Bengio, Y., 2017. Generalization in Deep Learning.arXiv preprint. arXiv: 1710.05468.
Kaytez, F., Taplamacioglu, M.C., Cam, E., Hardalac, F., 2015. Forecasting electricity consumption: a comparison of regression analysis, neural networks and least squares support vector machines. Int. J. Electr. Power Energy Syst. 67, 431–438.
Keles, D., Scelle, J., Paraschiv, F., Fichtner, W., 2016. Extended forecast methods for day-ahead electricity spot prices applying artificial neural networks. Appl. Energy 162, 218–230.
Khosravi, A., Nahavandi, S., Creighton, D., 2013. Quantifying uncertainties of neural network-based electricity price forecasts. Appl. Energy 112, 120–129.
Kialashaki, A., Reisel, J.R., 2013. Modeling of the energy demand of the residential sector in the United States using regression models and artificial neural networks. Appl. Energy 108, 271–280.
Kotsiantis, S.B., Zaharakis, I., Pintelas, P., 2007. Supervised machine learning: a review of classification techniques. Emerging Artificial Intelligence Applications in Computer Engineering. 160. pp. 3–24.
Kourentzes, N., Barrow, D.K., Crone, S.F., 2014. Neural network ensemble operators for time series forecasting. Expert Syst. Appl. 41 (9), 4235–4244.
Koutroumanidis, T., Ioannou, K., Arabatzis, G., 2009. Predicting fuelwood prices in Greece with the use of ARIMA models, artificial neural networks and a hybrid ARIMA-ANN model. Energy Policy 37 (9), 3627–3634.
Lago, J., De Ridder, F., De Schutter, B., 2018a. Forecasting spot electricity prices: deep learning approaches and empirical comparison of traditional algorithms. Appl. Energy 221, 386–405.
Lago, J., De Ridder, F., Vrancx, P., De Schutter, B., 2018b. Forecasting day-ahead electricity prices in Europe: the importance of considering market integration. Appl. Energy 211, 890–903.

Lai, T., To, W.M., Lo, W., Choy, Y., 2008. Modeling of electricity consumption in the Asian gaming and tourism center — Macao SAR, People's Republic of China. Energy 33 (5), 679–688.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521 (7553), 436.

Li, J., Wang, R., Wang, J., Li, Y., 2018a. Analysis and forecasting of the oil consumption in China based on combination models optimized by artificial intelligence algorithms. Energy 144, 243–264.

Li, J., Xu, Z., Xu, H., Tang, L., Yu, L., 2017. Forecasting oil price trends with sentiment of online news articles. Asia-Pacific Journal of Operational Research 34 (02), 1740019.

Li, Y., Che, J., Yang, Y., 2018b. Subsampled support vector regression ensemble for short term electric load forecasting. Energy 164, 160–170.

Liang, Q., FAN, Y., WEI, Y.-m., 2005. A long-term trend forecasting approach for oil price based on wavelet analysis. Chin. J. Manag. Sci. 13 (1), 30–36.

Limanond, T., Jomnonkwao, S., Srikaew, A., 2011. Projection of future transport energy demand of Thailand. Energy Policy 39 (5), 2754–2763.

Lin, W.-M., Gow, H.-J., Tsai, M.-T., 2010. An enhanced radial basis function network for short-term electricity price forecasting. Appl. Energy 87 (10), 3226–3234.

Liu, N., Tang, Q., Zhang, J., Fan, W., Liu, J., 2014. A hybrid forecasting model with parameter optimization for short-term load forecasting of micro-grids. Appl. Energy 129, 336–345.

Liu, X., Moreno, B., García, A.S., 2016. A grey neural network and input-output combined forecasting model. Primary energy consumption forecasts in Spanish economic sectors. Energy 115, 1042–1054.

Mahmoud, M.A., Alajmi, A.F., 2010. Quantitative assessment of energy conservation due to public awareness campaigns using neural networks. Appl. Energy 87 (1), 220–228.

Marcjasz, G., Uniejewski, B., Weron, R., 2018. On the importance of the long-term seasonal component in day-ahead electricity price forecasting with NARX neural networks. Int. J. Forecast. https://doi.org/10.1016/j.ijforecast.2017.11.009.

Martínez, A., Schmuck, C., Pereverzyev, S., Jr, Pirker, C., Haltmeier, M., 2018. A machine learning framework for customer purchase prediction in the non-contractual setting. Eur. J. Oper. Res. https://doi.org/10.1016/j.ejor.2018.04.034.

Marugán, A.P., Márquez, F.P.G., Perez, J.M.P., Ruiz-Hernández, D., 2018. A survey of artificial neural network in wind energy systems. Appl. Energy 228, 1822–1836.

Masci, C., Johnes, G., Agasisti, T., 2018. Student and school performance across countries: a machine learning approach. Eur. J. Oper. Res. 269 (3), 1072–1085.

Mashhadi, A.R., Behdad, S., 2018. Discriminant effects of consumer electronics use-phase attributes on household energy prediction. Energy Policy 118, 346–355.

Matthews, B.W., 1975. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. Biochim. Biophys. Acta 405 (2), 442–451.

McCulloch, W., Pitts, W., 1943. A logical calculus of the ideas immanent in nervous activity. Bull. Math. Biol. 5 (number), 115–133.

Mirakyan, A., Meyer-Renschhausen, M., Koch, A., 2017. Composite forecasting approach, application for next-day electricity price forecasting. Energy Econ. 66, 228–237.

Mohan, N., Soman, K., Kumar, S.S., 2018. A data-driven strategy for short-term electric load forecasting using dynamic mode decomposition model. Appl. Energy 232, 229–244.

Moreno, J., 2009. Trading strategies modeling in Colombian power market using artificial intelligence techniques. Energy Policy 37 (3), 836–843.

Moshiri, S., Brown, L., 2004. Unemployment variation over the business cycles: a comparison of forecasting models. J. Forecast. 23 (7), 497–511.

Moshiri, S., Cameron, N., 2000. Neural network versus econometric models in forecasting inflation. J. Forecast. 19 (3), 201–217.

Moshiri, S., Cameron, N.E., Scuse, D., 1999. Static, dynamic, and hybrid neural networks in forecasting inflation. Comput. Econ. 14 (3), 219–235.

Moshiri, S., Foroutan, F., 2006. Forecasting nonlinear crude oil futures prices. Energy J. 27 (4), 81–95.

Movagharnejad, K., Mehdizadeh, B., Banihashemi, M., Kordkheili, M.S., 2011. Forecasting the differences between various commercial oil prices in the Persian Gulf region by neural network. Energy 36 (7), 3979–3984.

Mullainathan, S., Spiess, J., 2017. Machine learning: an applied econometric approach. J. Econ. Perspect. 31 (2), 87–106.

Murat, Y.S., Ceylan, H., 2006. Use of artificial neural networks for transport energy demand modeling. Energy Policy 34 (17), 3165–3172.

Nelwamondo, F.V., Mohamed, S., Marwala, T., 2007. Missing data: a comparison of neural network and expectation maximization techniques. Curr. Sci. 93 (11), 1514–1521.

Nguyen, H.T., Nabney, I.T., 2010. Short-term electricity demand and gas price forecasts using wavelet transforms and adaptive models. Energy 35 (9), 3674–3685.

Özmen, A., Yılmaz, Y., Weber, G.-W., 2018. Natural gas consumption forecast with MARS and CMARS models for residential users. Energy Econ. 70, 357–381.

Panapakidis, I.P., Dagoumas, A.S., 2016. Day-ahead electricity price forecasting via the application of artificial neural network based models. Appl. Energy 172, 132–151.

Panapakidis, I.P., Dagoumas, A.S., 2017. Day-ahead natural gas demand forecasting based on the combination of wavelet transform and ANFIS/genetic algorithm/neural network model. Energy 118, 231–245.

Pao, H.-T., 2006. Comparing linear and nonlinear forecasts for Taiwan's electricity consumption. Energy 31 (12), 2129–2141.

Papadimitriou, T., Gogas, P., Stathakis, E., 2014. Forecasting energy markets using support vector machines. Energy Econ. 44, 135–142.

Peng, L., Liu, S., Liu, R., Wang, L., 2018. Effective long short-term memory with differential evolution algorithm for electricity price prediction. Energy 162, 1301–1314.

Perera, K.S., Aung, Z., Woon, W.L., 2014. Machine learning techniques for supporting renewable energy generation and integration: a survey. International Workshop on Data Analytics for Renewable Energy Integration. Springer., pp. 81–96. https://doi.org/10.1007/978-3-319-13290-7_7.

Pinto, T., Sousa, T.M., Praça, I., Vale, Z., Morais, H., 2016. Support Vector Machines for decision support in electricity markets strategic bidding. Neurocomputing 172, 438–445.

Rode, D.C., Fischbeck, P.S., 2018. Reduced-form models for power market risk analysis. Appl. Energy 228, 1640–1655.

Rudin, C., Waltz, D., Anderson, R.N., Boulanger, A., Salleb-Aouissi, A., Chow, M., Dutta, H., Gross, P.N., Huang, B., Ierome, S., Isaac, D.F., Kressner, A., Passonneau, R.J., Radeva, A., Wu, L., 2012. Machine learning for the New York City power grid. IEEE Trans. Analy. Mach. Intell. 34 (2), 328–345. Feb.

Safari, A., Davallou, M., 2018. Oil price forecasting using a hybrid model. Energy 148, 49–58.

Shaikhina, T., Lowe, D., Daga, S., Briggs, D., Higgins, R., Khovanova, N., 2015. Machine learning for predictive modelling based on small data in biomedical engineering. IFAC-PapersOnLine 48 (20), 469–474.

Shambora, W.E., Rossiter, R., 2007. Are there exploitable inefficiencies in the futures market for oil? Energy Econ. 29 (1), 18–27.

Singh, N., Mohanty, S.R., Shukla, R.D., 2017. Short term electricity price forecast based on environmentally adapted generalized neuron. Energy 125, 127–139.

Skiba, M., Mrówczyńska, M., Bazan-Krzywoszańska, A., 2017. Modeling the economic dependence between town development policy and increasing energy effectiveness with neural networks. Case study: the town of Zielona Góra. Appl. Energy 188, 356–366.

Sözen, A., 2009. Future projection of the energy dependency of Turkey using artificial neural network. Energy Policy 37 (11), 4827–4833.

Sözen, A., Arcaklioglu, E., 2007. Prediction of net energy consumption based on economic indicators (GNP and GDP) in Turkey. Energy Policy 35 (10), 4981–4992.

Sözen, A., Gülseven, Z., Arcaklioğlu, E., 2007. Forecasting based on sectoral energy consumption of GHGs in Turkey and mitigation policies. Energy Policy 35 (12), 6491–6505.

Sun, G., Chen, T., Wei, Z., Sun, Y., Zang, H., Chen, S., 2016. A carbon price forecasting model based on variational mode decomposition and spiking neural networks. Energies 9 (1), 54.

Sun, M., Wang, X., Chen, Y., Tian, L., 2011. Energy resources demand-supply system analysis and empirical research based on non-linear approach. Energy 36 (9), 5460–5465.

Sun, S., Sun, Y., Wang, S., Wei, Y., 2018. Interval decomposition ensemble approach for crude oil price forecasting. Energy Econ. 76, 274–287.

Szoplik, J., 2015. Forecasting of natural gas consumption with artificial neural networks. Energy 85, 208–220.

Tang, L., Dai, W., Yu, L., Wang, S., 2015a. A novel CEEMD-based EELM ensemble learning paradigm for crude oil price forecasting. Int. J. Inf. Technol. Decis. Mak. 14 (01), 141–169.

Tang, L., Wang, Z., Li, X., Yu, L., Zhang, G., 2015b. A novel hybrid FA-Based LSSVR learning paradigm for hydropower consumption forecasting. J. Syst. Sci. Complex. 28 (5), 1080–1101.

Tang, L., Wu, Y., Yu, L., 2018a. A non-iterative decomposition-ensemble learning paradigm using RVFL network for crude oil price forecasting. Appl. Soft Comput. 70, 1097–1108.

Tang, L., Wu, Y., Yu, L., 2018b. A randomized-algorithm-based decomposition-ensemble learning methodology for energy price forecasting. Energy 157, 526–538.

Tang, L., Yu, L., He, K., 2014. A novel data-characteristic-driven modeling methodology for nuclear energy consumption forecasting. Appl. Energy 128, 1–14.

Tang, L., Yu, L., Wang, S., Li, J., Wang, S., 2012. A novel hybrid ensemble learning paradigm for nuclear energy consumption forecasting. Appl. Energy 93, 432–443.

Tang, L.-B., Tang, L.-X., Sheng, H.-Y., 2009. Forecasting volatility based on wavelet support vector machine. Expert Syst. Appl. 36 (2), 2901–2909.

Ünler, A., 2008. Improvement of energy demand forecasts using swarm intelligence: the case of Turkey with projections to 2025. Energy Policy 36 (6), 1937–1944.

Vapnik, V., 1995. The Nature of Statistical Learning Theory. Springer-Verlag, New York.

Voyant, C., Notton, G., Kalogirou, S., Nivet, M.-L., Paoli, C., Motte, F., Fouilloy, A., 2017. Machine learning methods for solar radiation forecasting: a review. Renew. Energy 105, 569–582.

Wang, D., Luo, H., Grunder, O., Lin, Y., Guo, H., 2017. Multi-step ahead electricity price forecasting using a hybrid model based on two-layer decomposition technique and BP neural network optimized by firefly algorithm. Appl. Energy 190, 390–407.

Wang, D., Wang, Y., Song, X., Liu, Y., 2018a. Coal overcapacity in China: multiscale analysis and prediction. Energy Econ. 70, 244–257.

Wang, H., Zheng, B., Yoon, S.W., Ko, H.S., 2018b. A support vector machine-based ensemble algorithm for breast cancer diagnosis. Eur. J. Oper. Res. 267 (2), 687–699.

Wang, J., Athanasopoulos, G., Hyndman, R.J., Wang, S., 2018c. Crude oil price forecasting based on internet concern using an extreme learning machine. Int. J. Forecast. 34 (4), 665–677.

Wang, J., Wang, J., 2016. Forecasting energy market indices with recurrent neural networks: case study of crude oil price fluctuations. Energy 102, 365–374.

Wang, J., Zhu, W., Zhang, W., Sun, D., 2009. A trend fixed on firstly and seasonal adjustment model combined with the ε-SVR for short-term forecasting of electricity demand. Energy Policy 37 (11), 4901–4909.

Wang, L., An, H., Liu, X., Huang, X., 2016. Selecting dynamic moving average trading rules in the crude oil futures market using a genetic approach. Appl. Energy 162, 1608–1618.

Wang, L., Lv, S.-X., Zeng, Y.-R., 2018d. Effective sparse adaboost method with ESN and FOA for industrial electricity consumption forecasting in China. Energy 155, 1013–1031.

Wang, M., Tian, L., 2015. Regulating effect of the energy market — Theoretical and empirical analysis based on a novel energy prices-energy supply-economic growth dynamic system. Appl. Energy 155, 526–546.

Wang, M., Tian, L., Zhou, P., 2018e. A novel approach for oil price forecasting based on data fluctuation network. Energy Econ. 71, 201–212.

Wang, M., Zhao, L., Du, R., Wang, C., Chen, L., Tian, L., Stanley, H.E., 2018f. A novel hybrid method of forecasting crude oil prices using complex network science and artificial intelligence algorithms. Appl. Energy 220, 480–495.

Wang, S., Yu, L., Lai, K.K., 2005. A novel hybrid AI system framework for crude oil price forecasting. Data Mining and Knowledge Management. Springer., pp. 233–242. https://doi.org/10.1007/978-3-540-30537-8_26.

Wang, S., Yu, L., Tang, L., Wang, S., 2011. A novel seasonal decomposition based least squares support vector regression ensemble learning approach for hydropower consumption forecasting in China. Energy 36 (11), 6542–6554.

Wang, X., Luo, D., Zhao, X., Sun, Z., 2018g. Estimates of energy consumption in China using a self-adaptive multi-verse optimizer-based support vector machine with rolling cross-validation. Energy 152, 539–548.

Weidlich, A., Veit, D., 2008. A critical survey of agent-based wholesale electricity market models. Energy Econ. 30 (4), 1728–1759.

Weron, R., 2014. Electricity price forecasting: a review of the state-of-the-art with a look into the future. Int. J. Forecast. 30 (4), 1030–1081.

White, H., 2000. A reality check for data snooping. Econometrica 68 (5), 1097–1126.

Wickramasinghe, C.S., Marino, D.L., Amarasinghe, K., Manic, M., 2018. Generalization of Deep Learning for Cyber-Physical System Security: A Survey. IECON 2018-44th Annual Conference of the IEEE Industrial Electronics Society. IEEE., pp. 745–751. https://doi.org/10.1109/IECON.2018.8591773.

Xiao, J., Li, Y., Xie, L., Liu, D., Huang, J., 2018. A hybrid model based on selective ensemble for energy consumption forecasting in China. Energy 159, 534–546.

Xie, J., Song, Z., Li, Y., Zhang, Y., Yu, H., Zhan, J., Ma, Z., Qiao, Y., Zhang, J., Guo, J., 2018. A survey on machine learning-based mobile big data analysis: challenges and applications. Wirel. Commun. Mob. Comput. 2018, https://doi.org/10.1155/2018/8738613.

Xie, W., Yu, L., Xu, S., Wang, S., 2006. A new method for crude oil price forecasting based on support vector machines. International Conference on Computational Science. Springer., pp. 444–451. https://doi.org/10.1007/11758549_63.

Xiong, R., Nichols, E.P., Shen, Y., 2015. Deep Learning Stock Volatility with Google Domestic Trends. arXiv preprint. arXiv: 1512.04916.

Xuemian, H., Guohao, Z., 2008. Forecasting model of coal demand based on matlab bp neural network. Chin. J. Manag. Sci. 10 (16), 521–525.

Yang, J.H., Wu, C.Y., Zhang, G.M., 2014. Coal demand prediction in Shandong Province based on artificial firefly wavelet neural network. Advanced Materials Research. 962. pp. 1931–1935. Trans Tech Publ.

Yang, Z., Ce, L., Lian, L., 2017. Electricity price forecasting by a hybrid model, combining wavelet transform, ARMA and kernel-based extreme learning machine methods. Appl. Energy 190, 291–305.

Young, D., Poletti, S., Browne, O., 2014. Can agent-based models forecast spot prices in electricity markets? Evidence from the New Zealand electricity market. Energy Econ. 45, 419–434.

Yu, L., Dai, W., Tang, L., 2016a. A novel decomposition ensemble model with extended extreme learning machine for crude oil price forecasting. Eng. Appl. Artif. Intell. 47, 110–121.

Yu, L., Dai, W., Tang, L., Wu, J., 2016b. A hybrid grid-GA-based LSSVR learning paradigm for crude oil price forecasting. Neural Comput. Appl. 27 (8), 2193–2215.

Yu, L., Wang, S., Lai, K.K., 2008. Forecasting crude oil price with an EMD-based neural network ensemble learning paradigm. Energy Econ. 30 (5), 2623–2635.

Yu, L., Wang, Z., Tang, L., 2015. A decomposition-ensemble model with data-characteristic-driven reconstruction for crude oil price forecasting. Appl. Energy 156, 251–267.

Yu, L., Xu, H., Tang, L., 2017a. LSSVR ensemble learning with uncertain parameters for crude oil price forecasting. Appl. Soft Comput. 56, 692–701.

Yu, L., Zhao, Y., Tang, L., 2014. A compressed sensing based AI learning paradigm for crude oil price forecasting. Energy Econ. 46, 236–245.

Yu, L., Zhao, Y., Tang, L., 2017b. Ensemble forecasting for complex time series using sparse representation and neural networks. J. Forecast. 36 (2), 122–138.

Yu, S., Wei, Y.-M., Wang, K., 2012. A PSO-GA optimal model to estimate primary energy demand of China. Energy Policy 42, 329–340.

Yun-cai, N., 2003. A superposition wavelet-neural network model of coal demand forecast [J]. J. China Coal Soc. 1, 021.

Zaidi, S.M.A., Chandola, V., Allen, M.R., Sanyal, J., Stewart, R.N., Bhaduri, B.L., McManamay, R.A., 2018. Machine learning for energy-water nexus: challenges and opportunities. Big Earth Data 2 (3), 228–267.

Zemene, F., Khedkar, V., 2017. Survey on machine learning based electric consumption forecasting using smart meter data. Int. J. Comput. Appl. 180 (6), 46–52. Dec.

Zendehboudi, S., Rezaei, N., Lohi, A., 2018. Applications of hybrid models in chemical, petroleum, and energy systems: a systematic review. Appl. Energy 228, 2539–2566.

Zeng, Y., Liu, J., Sun, K., Hu, L.-w., 2018. Machine learning based system performance prediction model for reactor control. Ann. Nucl. Energy 113, 270–278.

Zeng, Y.-R., Zeng, Y., Choi, B., Wang, L., 2017. Multifactor-influenced energy consumption forecasting using enhanced back-propagation neural network. Energy 127, 381–396.

Zhang, G.P., 2003. Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing 50, 159–175.

Zhang, J.-L., Zhang, Y.-J., Zhang, L., 2015. A novel hybrid method for crude oil price forecasting. Energy Econ. 49, 649–659.

Zhang, W., Robinson, C., Guhathakurta, S., Garikapati, V.M., Dilkina, B., Brown, M.A., Pendyala, R.M., 2018. Estimating residential energy consumption in metropolitan areas: a microsimulation approach. Energy 155, 162–173.

Zhang, W., Tian, L., Wang, M., Zhen, Z., Fang, G., 2016. The evolution model of electricity market on the stable development in China and its dynamic analysis. Energy 114, 344–359.

Zhang, Y., Ling, C., 2018. A strategy to apply machine learning to small datasets in materials science. npj Comput. Mater. 4 (1), 25.

Zhang, Z., Gao, G., Shi, Y., 2014. Credit risk evaluation using multi-criteria optimization classifier with kernel, fuzzification and penalty factors. Eur. J. Oper. Res. 237 (1), 335–348.

Zhao, L.-T., Wang, Y., Guo, S.-Q., Zeng, G.-R., 2018. A novel method based on numerical fitting for oil price trend forecasting. Appl. Energy 220, 154–163.

Zhao, Y., Li, J., Yu, L., 2017. A deep learning ensemble approach for crude oil price forecasting. Energy Econ. 66, 9–16.

Zhu, B., Han, D., Wang, P., Wu, Z., Zhang, T., Wei, Y.-M., 2017. Forecasting carbon price using empirical mode decomposition and evolutionary least squares support vector regression. Appl. Energy 191, 521–530.

Zhu, B., Shi, X., Chevallier, J., Wang, P., Wei, Y.-M., 2016. An adaptive multiscale ensemble learning paradigm for nonstationary and nonlinear energy price time series forecasting. J. Forecast. 35 (7), 633–651.

Zhu, B., Ye, S., Wang, P., He, K., Zhang, T., Wei, Y.-M., 2018. A novel multiscale nonlinear ensemble leaning paradigm for carbon price forecasting. Energy Econ. 70, 143–157.