

## Journal of Sustainable Energy Systems

Home Page: <https://ses.ut.ac.ir>

## Research Paper

**Modeling short-term forecast of electricity consumption using artificial intelligence methods in Gilan province****Mobina Simaee<sup>1\*</sup>, Amirhosein Mirabadi<sup>2</sup>**<sup>1</sup> Master of Science, Department of Renewable Energies and Environment, Faculty of New Sciences and Technologies, University of Tehran, Iran<sup>2</sup> Science and Technology Vice-Presidency, Tehran, Iran**ARTICLE INFO****Article History:**

Received 30 March 2022

Revised 02 May 2022

Accepted 15 June 2022

**Keywords:**electricity consumption forecasting,  
short-term electricity forecasting (STLF)  
support vector regression (SVR)  
linear regression (LR)  
artificial neural network (ANN)**ABSTRACT**

Load forecasting is one of the most important aspects of the effective management of power systems and depends on various factors such as the means and sources of electricity production by each company, electric load demand, weather factors, economic factors and human activities.. The main goal of this article is the STLF in the regional electricity grid of Guilan province. In this research, after collecting data including wind speed, relative humidity, temperature, dew point, holidays, day length and the effect of corona disease, pre-processing is done on them and they are divided into five clusters using K\_Means clustering algorithm. In the following, the relationships between the independent and dependent variables of electricity consumption in Guilan province, using linear regression (LR) algorithms, artificial neural network (ANN) and support vector regression (SVR) along with the Grid search optimization method, were investigated and analyzed in Python software and modeled in the Google Colab environment. The results show that the SVR algorithm has higher accuracy and longer implementation time than the two algorithms of ANN and LR.

**Introduction**

Electricity as an energy carrier has different characteristics compared to other energy carriers, because it cannot be stored and must be produced and consumed according to the amount of demand. Control, monitoring and identification in large national and regional systems is one of the issues that require governments to find an optimal solution to achieve it. Load forecasting is one of the most important aspects of effective management of power systems. The forecast depends on various factors, including the means and sources of electricity production by each company, electric load demand, weather factors, economic reasons and human activities. Based on the time horizon, load forecasting can be classified into three main groups, short-term load forecasting (STLF), medium-term load forecasting (MTLF) and long-term load forecasting (LTLF). Electricity consumption forecasting methods are divided into

three categories, statistical models, artificial intelligence-based models, and hybrid models.

**Model description**

The present study proposes a method for hourly forecasting of electricity consumption. The studied area is Guilan province. Guilan province, with an area of 14,711 square kilometers, is one of the northern provinces of Iran, located at 36°33' to 38°27' north latitude and 48°32' to 58°36' east longitude from the Meridian and between Alborz and Talesh mountains are located. The electricity consumption of Guilan province has grown 48.96% over the past 12 years and the average annual growth rate is 0.026%. From February 2018 to August 2019, with the arrival of Corona and its peak, the social life cycle of people due to quarantine was accompanied by changes, and electricity consumption was focused on increasing demand in the domestic sector, and in contrast to the demand in other sectors, there was a relatively decreasing trend. In the short-term forecast, hourly electricity consumption data of the entire

\* Corresponding Author, Email: mobina.simaee@ut.ac.ir

province has been selected as the objective function. The hourly data of 6 years from 2016 to 2021 was prepared from the regional electricity of Gilan province.

### Results and discussions

The meteorological data used in this research was received from the country's meteorological system for every 3 hours of the day and night. The length of the day is one of the factors that has a significant effect on the amount of electricity consumption. In this research, after collecting data including wind speed, relative humidity, temperature, dew point, holidays, day length, the effect of Corona and performing pre-processing on them, the data is divided into five clusters using the K\_Means clustering algorithm. In the following, the relationships between independent and dependent variables of electricity consumption in Gilan province have been investigated and modeled using linear regression (LR) algorithms, artificial neural network (ANN) and support vector regression (SVR) along with Grid search (GS) optimization method in Python software and Google Colab environment. Error measurement criteria are used to

check the amount of prediction error. The amount of these criteria is different depending on the type of data and their quality and the algorithm used. The higher the prediction accuracy or the lower the error value of the prediction model, the lower the operating cost and the higher the profit. MAPE, MSE, MAE, RMSE values in linear regression algorithm are 0.251, 1689.29, 41.097, 24.473 and 0.982 respectively, in artificial neural network algorithm respectively 0.0259, 1673.669, 40.910, 25.330 and 0.982 and in support vector regression algorithm respectively 0.0214, 1288.567, 35.896, 20.874 and 0.986. The implementation time of linear regression, artificial neural network and support vector regression algorithms is different. In some researches, the goal is to achieve a high accuracy of prediction along with the minimum time to implement the algorithm training. The SVR+GS algorithm has the highest accuracy and the highest implementation time, the ANN algorithm has lower accuracy than the SVR model but with a shorter implementation time, and the LR algorithm has less accuracy and a much shorter implementation time.

## فصلنامه سیستم‌های انرژی پایدار

سایت نشریه: <https://ses.ut.ac.ir>

مقاله پژوهشی

### مدل‌سازی پیش‌بینی کوتاه‌مدت مصرف برق با استفاده از روش‌های هوش مصنوعی در استان گیلان

مبینا سیمائی<sup>۱</sup>، امیرحسین میرآبادی<sup>۲</sup>

<sup>۱</sup> کارشناسی ارشد، گروه مهندسی انرژی‌های نو و محیط زیست، دانشکده علوم و فنون نوین، دانشگاه تهران

<sup>۲</sup> معاونت علم و فناوری ریاست جمهوری

| اطلاعات مقاله                                                                                            | چکیده                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |
|----------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>تاریخ‌های مقاله:</b><br>تاریخ دریافت ۱۴۰۱/۰۱/۱۰<br>تاریخ بازنگری ۱۴۰۱/۰۲/۱۲<br>تاریخ تصویب ۱۴۰۱/۰۳/۲۵ | پیش‌بینی بار از مهم‌ترین جنبه‌های مدیریت مؤثر سیستم‌های قدرت است و به عوامل مختلفی از جمله ابزار و منابع تولید برق توسط هر شرکت، تقاضای بار الکتریکی، عوامل آب‌وهوایی، عوامل اقتصادی و فعالیت‌های انسانی بستگی دارد. براساس افق زمانی، پیش‌بینی بار را می‌توان به سه گروه عمده پیش‌بینی کوتاه‌مدت، میان‌مدت و بلندمدت دسته‌بندی کرد. هدف اصلی این مقاله، پیش‌بینی کوتاه‌مدت مصرف برق، در شبکه برق منطقه‌ای استان گیلان است. در این پژوهش، پس از جمع‌آوری داده‌ها شامل سرعت باد، رطوبت نسبی، دما، نقطه شبنم، تعطیلات، طول روز و تأثیر کرونا، پیش‌پردازش روی آن‌ها انجام شده و با استفاده از الگوریتم خوشه‌بندی K-Means به پنج خوشه تقسیم می‌شوند. در ادامه روابط بین متغیرهای مستقل و وابسته مصرف برق در استان گیلان با استفاده از الگوریتم‌های رگرسیون خطی <sup>۱</sup> ، شبکه عصبی مصنوعی <sup>۲</sup> و رگرسیون بردار پشتیبان <sup>۳</sup> به همراه روش بهینه‌سازی جستجوی شبکه <sup>۴</sup> ، مورد بررسی قرار گرفته و در نرم‌افزار Python و در محیط Google Colab، بررسی و مدل‌سازی شده است. در خوشه‌بندی، الگوریتم‌های یادشده، روی تمامی خوشه‌ها و مجموع آن‌ها اعمال می‌شود. نتایج تحقیق در این مقاله نشان می‌دهد الگوریتم رگرسیون بردار پشتیبان، دقت بالاتر و زمان اجرای بیشتری نسبت به دو الگوریتم شبکه عصبی مصنوعی و رگرسیون خطی دارد. الگوریتم شبکه عصبی مصنوعی نسبت به رگرسیون خطی دارای خطای کمتر و زمان اجرای بیشتری است. |

#### ۱. مقدمه<sup>۱</sup>

برق به عنوان یک منبع انرژی پاک و کارآمد، نقشی بی‌بدیل در زندگی روزمره ایفا می‌کند. الکتریسیته به عنوان یک حامل انرژی دارای ویژگی‌های متفاوتی در مقایسه با سایر حامل‌ها است، زیرا قابلیت ذخیره‌سازی

\* نویسنده مسئول

Email: Mobina.simaee@ut.ac.ir

1. Linear Regression Algorithm (LR)
2. Artificial Neural Networks (ANN)
3. Support Vector Regression (SVR)
4. Grid Search (GS)

نداشته و باید متناسب با مقدار تقاضا تولید و مصرف شود. در بهره‌برداری از سیستم برق، توانایی ارائه بار به مشترکان، چالش‌برانگیزترین جنبه است، به این معنا که باید همیشه نیازهای بار مشترکان به صورت لحظه‌ای و در هر زمان تأمین شود. ژنراتورها باید دارای بار اضافی باشند تا بتوانند در مواقع ضروری بخشی از برق شبکه را تأمین کنند. با توجه به نوسانات قابل توجه بار در هر روز و همچنین، مدیریت سمت عرضه برق، اپراتور سیستم باید بتواند تقاضای بار را برای چند ساعت و یا چند سال آینده

پیش‌بینی کند. به عنوان مثال، ژنراتورها به زمان قابل توجهی برای همگام‌سازی با شبکه نیاز دارند. از این‌رو، دانش قبلی در مورد نیازهای بار، اپراتور تأسیسات برق را قادر می‌سازد تا منابع تولید را به طور بهینه‌ای تخصیص دهند [۱]. سیستم قدرت الکتریکی یک سیستم تحویل انرژی در زمان واقعی است. این سیستم را یک سیستم لحظه‌ای می‌نامند، زیرا در لحظه‌ای که کلید برق را روشن می‌کنیم برق تولید، انتقال و تأمین می‌شود. شبکه برق از پیچیده‌ترین سازه‌های متشکل از تعداد بی‌شماری از ترانسفورماتورها، پست‌های برق، مجموعه ژنراتورها و خطوط انتقال نیرو است. پیکربندی معمولی یک سیستم قدرت در هر منطقه متفاوت خواهد بود. ساختار سیستم قدرت بسته به منطقه جغرافیایی، نفوذ منابع تجدیدپذیر و نیازهای بار در بخش‌های مختلف مصرف انرژی، تغییر خواهد کرد. نیروگاه‌های تجدیدپذیر برای تولید انرژی الکتریکی، از منابع تولید انرژی مانند گرما، خورشیدی، آب، باد استفاده کرده و نیروگاه‌های فسیلی، سوخت‌های فسیلی را به انرژی الکتریکی ولتاژ بالا تبدیل می‌کنند. برق با ولتاژ بالا از طریق خطوط به مراکز بار منتقل می‌شود. قبل از توزیع به مصرف‌کنندگان، ایستگاه‌های فرعی انرژی الکتریکی ولتاژ بالا را توسط ترانسفورماتورها به برق ولتاژ پایین‌تر تبدیل می‌کند و در ادامه با استفاده از خطوط توزیع بین مشتریان صنعتی، کشاورزی، خانگی و تجاری و دیگر مصارف توزیع خواهد شد [۲].

از طرف دیگر، کنترل، پایش و شناسایی در سیستم‌های بزرگ ملی و منطقه‌ای یکی از موضوعات مطرح است که دولت‌ها را ملزم به یافتن راه‌حلی بهینه برای دستیابی به آن می‌کند. برای رسیدن به این هدف داده‌های دقیق مخابره‌شده، نقش بسیار مهمی در کنترل شبکه‌ها دارند. با وجود رشد و توسعه کشور در زمینه مخابرات و هوشمندسازی بسترهای ثبت اطلاعات، همچنان تحلیل داده‌ها با توجه به الگوی حاکم بر مصرف و شناسایی بازه‌های کم‌باری، متوسط و اوج بار شبکه با مشکل روبه‌رو است. بازار برق در شرکت‌های برق منطقه‌ای و توزیع برق یکی از چالش‌های بزرگ پیش رو است و اطلاع از میزان مصرف برق در بازه‌های زمانی مختلف آینده می‌تواند موجب کنترل بهینه سیستم و به‌کارگیری مفید از ظرفیت نیروگاه‌ها شود [۳]. پیش‌بینی به عوامل مختلفی از

جمله ابزار و منابع تولید برق توسط هر شرکت، تقاضای بار الکتریکی، عوامل آب‌وهوایی، دلایل اقتصادی و فعالیت‌های انسانی بستگی دارد. عوامل آب‌وهوایی عواملی هستند که براساس ویژگی‌های هواشناسی مانند دما، رطوبت، سرعت باد و بارندگی روی مقدار تقاضای کوتاه‌مدت تأثیر بسزایی دارند. تأثیر فعالیت‌های انسانی بر مصرف بار الکتریکی را می‌توان در ویژگی‌های مختلفی تحلیل کرد. یکی از این ویژگی‌ها، ویژگی‌های تقویم، مانند روز هفته یا ماه سال است [۲]. سایر ویژگی‌ها به عوامل اقتصادی مانند فعالیت‌های اقتصادی در مناطق شهری یا معاملات اقتصادی مربوط می‌شود. نتیجه دقت‌های زیاد پیش‌بینی به صرفه‌جویی قابل توجهی در هزینه‌های عملیاتی، توسعه تأسیسات تولید، انتقال و توزیع و نگهداری و تصمیم‌گیری‌های صحیح در جهت بهینه‌سازی مصرف انرژی منجر می‌شود.

براساس افق زمانی، پیش‌بینی بار را می‌توان در سه گروه عمده پیش‌بینی کوتاه‌مدت<sup>۱</sup>، میان‌مدت<sup>۲</sup> و بلندمدت<sup>۳</sup> دسته‌بندی کرد [۴]. در پیش‌بینی کوتاه‌مدت با استفاده از عوامل فصلی، آب‌وهوایی و داده‌های تاریخی مصرف بار، میزان تقاضای بار الکتریکی از یک ساعت تا یک هفته آینده و تقاضای اوج بار روزانه را پیش‌بینی می‌کنند [۳]. در پیش‌بینی میان‌مدت بار، به منظور برنامه‌ریزی تعمیر و نگهداری، برنامه‌ریزی تأمین سوخت و تنظیمات جزئی زیرساخت‌های شبکه برق، بار الکتریکی از یک هفته تا یک سال پیش‌بینی خواهد شد [۵]. پیش‌بینی بلندمدت در یک دوره ۲۰ ساله به منظور ساخت نیروگاه‌های جدید، افزایش ظرفیت سیستم انتقال و به طور کلی برای برنامه‌ریزی توسعه تأسیسات برق انجام می‌شود [۶]. در پیش‌بینی، هیچ قطعیتی وجود ندارد و به طبع یک مسئله تصادفی است. به همین علت متغیر وابسته، به دلیل ماهیت تصادفی مدل پیش‌بینی‌شده، هرگز به طور ۱۰۰ درصد قابل دستیابی نخواهد بود. اطلاعات ناقص و کم‌دقت مربوط به عوامل مستقل تأثیرگذار بر پیش‌بینی مصرف برق و به‌کارگیری روش نامناسب برای مدل‌سازی، از عواملی هستند که ممکن است ما را به پیش‌بینی‌هایی با دقت بسیار کم سوق دهند.

1. Short term load forecasting (STLF)  
2. Mid term load forecasting (MTLF)  
3. Long term load forecasting (LTLF)

می توان به چهار دسته عوامل آب و هوایی، زمانی، اقتصادی و تصادفی تقسیم کرد. آب و هوا یک عنصر حیاتی برای عرضه و تقاضای برق باست و تغییرات آن، یک رفتار دوره ای را در پیش بینی بار از خود نشان می دهند. پیش بینی بار به شدت تحت تأثیر متغیرهایی مانند دما<sup>۱</sup>، رطوبت نسبی<sup>۲</sup>، پوشش ابر<sup>۳</sup>، سرعت باد<sup>۴</sup>، نقطه شبنم، بارندگی و... است [۸].

از عوامل زمانی می توان به آثار فصلی، هفتگی و روزانه اشاره کرد. به غیر از متغیرهای آب و هوایی، فعالیت انسان یکی دیگر از عوامل بسیار مؤثر است که بر مصرف بار الکتریکی تأثیر می گذارد. متغیر فعالیت انسانی در نظر گرفته شده به معنای تفکیک زمانی به روز هفته، ماه، سال و همچنین به اطلاعات مربوط به تعطیلات بستگی دارد. از دیگر فاکتورهای مؤثر می توان به تغییرات فصلی اشاره کرد، شکاف تقاضای انرژی در تابستان به دلیل افزایش دما و بارهای تهویه مطبوع به اوج خود می رسد [۸]. شاخص های اقتصادی مورد استفاده در پیش بینی های میان مدت و بلندمدت تأثیر زیادی بر تقاضای برق کشور دارد، به طوری که با افزایش فعالیت های اقتصادی، تقاضای برق می تواند سیر صعودی بگیرد و یا بر اثر بهینه شدن تجهیزات الکتریکی و صرفه جویی مصرف برق توسط مشترکان، روند نزولی داشته باشد. این شاخص ها شامل تولید ناخالص داخلی<sup>۵</sup>، رشد جمعیت، درآمد سرانه<sup>۶</sup>، کشش های قیمتی<sup>۷</sup>، قیمت انرژی و تعداد مشترکان و... هستند و مصرف برق آینده را مورد ارزیابی قرار می دهند [۹]. خاموشی ها، راه اندازی بارهای بزرگ مانند کارخانه فولاد و ذوب آهن، سایر رویدادهای غیرعادی مانند بیماری همه گیر کرونا که منجر به تکان های منحنی بار می شوند و رویدادهای ورزشی در دسته اختلالات تصادفی قرار می گیرند. از مشکلات پیش بینی می توان به بیش برآزش<sup>۸</sup>، اطلاعات ناقص برای متغیرهای مستقل، انتخاب ورودی های نامناسب و نویز<sup>۹</sup> یا همان داده های پرت نام برد [۱۰].

افزایش خطا در پیش بینی تقاضای بار الکتریکی، هزینه های عملیاتی را افزایش می دهد. اگر بار الکتریکی پیش بینی شده بیشتر از تقاضای واقعی باشد، هزینه عملکرد و مصرف سوخت به میزان قابل توجهی افزایش می یابد. از سوی دیگر، اگر بار الکتریکی پیش بینی شده کمتر از تقاضای واقعی باشد، باعث بروز خاموشی می شود که این رویداد می تواند هزینه زیادی به ویژه برای مشتریان بزرگ صنعتی به همراه داشته باشد. علاوه بر این، پیش بینی بار با دقت زیاد، میزان مصرف انرژی و آلودگی های محیط زیست را نیز کاهش خواهد داد [۱]. دلیل انتخاب این استان، شرایط محیطی و داشتن پتانسیل بالا در توسعه انرژی تجدیدپذیر در صورت داشتن برنامه و مدیریت صحیح است. استان گیلان به دلیل شرایط زیست محیطی و طبیعت بکر، منحنی تقاضای بار متفاوتی نسبت به یک استان گرم و خشک دارد. دلایلی مانند صنعتی نبودن و نداشتن محیط کسب و کار باعث تغییر روند زندگی می شود و همین امر مهاجرت به استان های مجاور را نیز افزایش می دهد. به دنبال این مسئله، ورود جمعیت زیادی در تعطیلات رسمی به خصوص عید نوروز، منحنی مصرف برق را تغییر خواهد داد. بنابراین، لازم است تمهیداتی اندیشیده شود تا خاموشی بخش مصرفی به ویژه در ساعات اوج مصرف به حداقل برسد. در این مقاله هدف اصلی پیش بینی کوتاه مدت مصرف برق، در برق منطقه ای استان گیلان است که در راستای دستیابی به آن، می توان به اهداف کوچک تری مانند مروری بر روش های مختلف مدل سازی و پیش بینی مصرف انرژی برق، بررسی عوامل مؤثر بر پیش بینی و همچنین طبقه بندی آن ها براساس افق های پیش بینی مختلف اشاره کرد.

تولید برق به تناسب تقاضای آن، مسئله حیاتی یک کشور یا منطقه است. اگر ظرفیت تولید برق یک کشور کمتر از تقاضای آن باشد، صنعت وابسته به برق، تحت تأثیر منفی قرار می گیرد و ممکن است خاموشی رخ دهد. از طرفی، تولید برق بیشتر از تقاضا، باعث می شود نیروگاه ها با ظرفیت غیرفعال کار کنند و در نتیجه اتلاف منابع اقتصادی را به دنبال خواهد داشت. نیاز به تولید و مدیریت برق در هر سال روند افزایشی دارد که نتیجه افزایش سریع جمعیت جهان و روند صعودی تعداد وسایل الکترونیکی به ازای هر فرد است [۷]. تقاضای برق توسط مجموعه ای از متغیرهای مختلف کنترل می شود که

1. Temperature
2. Humidity
3. Cloud Coverage
4. Wind Speed
5. GDP
6. Income Per Capita
7. Price Elasticities
8. Over Fitting
9. Anomalies

یادگیری ماشین و یادگیری عمیق زیرشاخه‌ای از هوش مصنوعی هستند که اساساً از یادگیری بیولوژیکی الهام گرفته و بر مفاهیمی از علوم کامپیوتر، آمار، مهندسی، تئوری بهینه‌سازی و بسیاری از رشته‌های ریاضیات و علوم بنا شده‌اند. الگوریتم‌های یادگیری ماشین شامل درخت تصمیم<sup>۱۶</sup>، الگوریتم ماشین بردار پشتیبان<sup>۱۷</sup>، الگوریتم K نزدیک‌ترین همسایه<sup>۱۸</sup>، الگوریتم جنگل تصادفی<sup>۱۹</sup>، الگوریتم ژنتیک<sup>۲۰</sup>، شبکه عصبی مصنوعی و بسیاری از روش‌های دیگر است [۱۳]. در جدول ۱ مقالات مختلف از جنبه‌های افق پیش‌بینی، مدل اجرایی، داده‌های مستقل به‌کاررفته، معیارهای خطا، نتایج و اهداف دسته‌بندی کرده است. طبق جدول ۱ می‌توان نتیجه گرفت که استفاده از مدل‌های مبتنی بر هوش مصنوعی دارای نتیجه بهتری هستند.

## ۲. مواد و روش‌ها

در این بخش، منطقه مورد مطالعه و مصرف برق آن از جنبه‌های مختلف مورد بررسی قرار می‌گیرد. سپس، پارامترهای مستقل در مصرف برق و معادلات مورد استفاده در الگوریتم‌های رگرسیون خطی، شبکه عصبی مصنوعی و الگوریتم ماشین بردار پشتیبان به تفصیل بیان می‌شود.

### ۲.۱. منطقه مورد مطالعه

استان گیلان، با مساحت ۱۴۷۱۱ کیلومتر مربع یکی از استان‌های شمالی کشور ایران است که در ۳۶ درجه و ۳۳ دقیقه تا ۳۸ درجه و ۲۷ دقیقه عرض شمالی و ۴۸ درجه و ۳۲ دقیقه تا ۵۰ درجه و ۳۶ دقیقه طول شرقی از نصف‌النهار و بین رشته‌کوه‌های البرز و تالش قرار دارد. این استان از غرب با استان اردبیل، از شرق با استان مازندران، از جنوب با استان زنجان و از شمال با دریای خزر همسایه است. شکل ۱ موقعیت جغرافیایی استان را نشان می‌دهد.

روش‌های پیش‌بینی اولیه به دو نوع اساسی کمی و کیفی طبقه‌بندی می‌شوند و انتخاب نوع مناسب بیشتر به داده‌های موجود بستگی دارد. در روش‌های پیش‌بینی کیفی، بار آینده به صورت ذهنی بر اساس نظرات کارشناسان پیش‌بینی می‌شود. در این روش به دور از حدس رویکردهای ساختاریافته‌ای برای به دست آوردن پیش‌بینی‌های مطلوب، بدون استفاده از داده‌های تاریخی مورد استفاده قرار می‌گیرد. از این‌رو، چنین روش‌هایی مانند برازش منحنی ذهنی<sup>۱</sup>، روش دلفی<sup>۲</sup> و مقایسه‌های تکنولوژیکی<sup>۳</sup>، زمانی مفید و کاربردی هستند که داده‌های تاریخی در دسترس یا کمیاب باشند. از سوی دیگر، روش‌های پیش‌بینی کمی مبتنی بر فرمول‌های ریاضی و آماری است و تحت دو شرط در دسترس بودن داده‌ها و تداوم الگوهای گذشته در آینده به کار گرفته می‌شوند [۱۱].

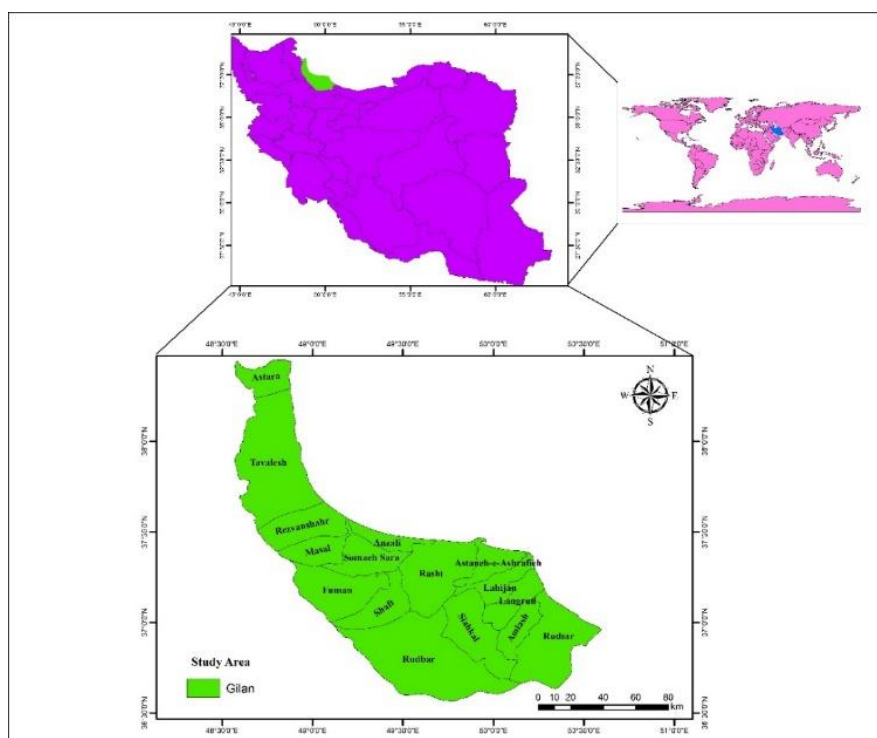
مدل آماری، نوعی مدل ریاضی است که مجموعه‌ای از مفروضات آماری مربوط به تولید داده‌های نمونه را در بر می‌گیرد. از انواع مدل‌های آماری می‌توان به مدل‌های پایه باکس جنکینز<sup>۴</sup>، الگوریتم فیلتر کالمن<sup>۵</sup>، مدل‌های خاکستری<sup>۶</sup> و هموارسازی نمایی<sup>۷</sup> نام برد. مدل‌های پایه باکس جنکینز شامل مدل خودرگرسیون<sup>۸</sup>، مدل میانگین متحرک<sup>۹</sup>، مدل میانگین متحرک خودرگرسیون<sup>۱۰</sup>، مدل میانگین متحرک یکپارچه اتورگرسیون<sup>۱۱</sup>، مدل میانگین متحرک یکپارچه اتورگرسیون فصلی<sup>۱۲</sup>، مدل میانگین متحرک یکپارچه اکتسابی<sup>۱۳</sup>، مدل خودرگرسیون میانگین متحرک با ورودی‌های خارجی<sup>۱۴</sup> و میانگین متحرک یکپارچه اتورگرسیون با ورودی‌های خارجی<sup>۱۵</sup> و مدل‌های رگرسیونی است [۱۲].

1. Subjective Curve Fitting
2. Delphi Method
3. Technological Comparisons
4. Box-Jenkins Basic Models
5. Kalman Filtering Algorithms
6. Grey Models
7. Exponential Smoothing
8. Autoregressive (AR)
9. Moving Average (MA)
10. Autoregressive Moving Average (ARMA)
11. Autoregressive Integrated Moving Average (ARIMA)
12. Seasonal Autoregressive Integrated Moving Average (SARIMA)
13. Autoregressive Fractionally Integrated Moving Average (ARFIMA)
14. Autoregressive moving average model with exogenous (ARMAX)
15. Auto Regressive Integrated Moving Average With Exogeneous (ARIMAX)

16. Decision Tree (DT)
17. Support Vector Machine (SVM)
18. K-Nearest Neighbour (KNN)
19. Random Forest (RF)
20. Genetic Algorithms (GA)

جدول ۱. مروری بر ادبیات

| شماره | افق پیش بینی | عوامل مستقل                                                                                                                                                                                                  | مدل پیش بینی                   | دستاورد                                                                                                                                                                                                                                                                | معیار خطا           | سال  |
|-------|--------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------|------|
| [۱۴]  | کوتاه مدت    | داده های بار هر ۱۵ دقیقه، دمای ساعتی<br>بیرون ساختمان، سرعت باد، رطوبت نسبی، تعطیلات<br>(داده های ساعتی ۲۰۱۳-۲۰۱۸)                                                                                           | LSTM, SVM                      | مدل LSTMs زمانی که داده های بار کافی باشد، دقت پیش بینی بیشتری دارد. با این حال، عملکرد کلی مدل SVM زمانی بهتر است که داده های بار مورد استفاده برای آموزش مدل کافی نباشد.                                                                                             | RMSE                | ۲۰۲۱ |
| [۱۵]  | کوتاه مدت    | مصرف برق ساعتی ساختمان، دمای هوا، رطوبت نسبی، روزهای هفته<br>(داده های ۲۰۱۲-۲۰۱۳)                                                                                                                            | Tree, SVR (ensemble ANN و KNN) | استفاده از یک مدل یادگیری جمعی بجای تک الگوریتم و رسیدن MAPE ۷/۵ و ۵۵/۸ درصد برای ساختمان های تجاری و مسکونی                                                                                                                                                           | MAPE                | ۲۰۱۵ |
| [۱۶]  | کوتاه مدت    | نوع روز (روزهای کاری شامل دوشنبه تا جمعه و آخر هفته شنبه و یکشنبه)، ساعت های مصرف برق، تهویه مطبوع، دما، بیرون، رطوبت نسبی، سرعت و جهت باد، میزان بارندگی و مصرف برق<br>(داده های ساعتی آگست و سپتامبر ۲۰۱۲) | ANN                            | پیش بینی وضوح ۱۵ دقیقه ای با RMSE ۱۰ درصد برای روزهای هفته و پیش بینی پیک تقاضای روزانه روزهای هفته را با APE ۵ درصد انجام شده است.                                                                                                                                    | RMSE<br>MBE<br>APE  | ۲۰۱۶ |
| [۱۷]  | میان مدت     | مصرف برق ماهانه<br>(داده های ژانویه ۱۹۹۰ تا دسامبر ۲۰۱۶)                                                                                                                                                     | AR, NPAR, STAR, ARMA           | تقسیم مصرف برق به دو دسته قطعی و تصادفی و تخمین مؤلفه قطعی با مدل های پارامتریک و ناپارامتریک و مؤلفه تصادفی با استفاده از چهار مدل سری زمانی تک متغیره از جمله مدل های خودرگرسیون پارامتریک، خودرگرسیو ناپارامتریک، خودبارگشتی انتقال هموار و میانگین متحرک اتورگرسیو | MAE<br>RMSE<br>MAPE | ۲۰۲۰ |
| [۱۸]  | میان مدت     | داده های ماهانه مصرف برق<br>(داده های ۱۹۹۸-۲۰۱۴ لهستان و ۱۹۹۱-۲۰۱۴)                                                                                                                                          | MLP                            | استفاده از یک شبکه عصبی پرسپترون چند لایه و تمرکز روی چرخه های فصلی سری های زمانی برای پیش بینی بار الکتریکی که نتایج بهتری را نسبت به شبکه عصبی فازی و مدل میانگین متحرک یکپارچه اتورگرسیو دارد.                                                                      | MAPE                | ۲۰۱۹ |
| [۵]   | میان مدت     | روز، ماه، مصرف برق خانگی، عمومی، تجاری و صنعتی<br>(داده های ساعتی ۲۰۰۹-۲۰۱۸)                                                                                                                                 | ARIMA, CNN-LSTM                | با تنظیم فرآیندهای مدل میانگین متحرک یکپارچه اتورگرسیو و مدل ترکیبی مقدار RMSE، برای مرحله آموزش ۱۴ درصد و تست ۲۰ درصد به دست آمده است.                                                                                                                                | RMSE<br>MSE         | ۲۰۲۱ |
| [۶]   | بلند مدت     | جمعیت، مصرف برق در بخش های خانگی، تجاری و صنعتی در مناطق شمال، شمال شرقی، جنوب شرقی، جنوب و جنوب غربی، GDP<br>(داده های ۱۹۷۹-۲۰۱۳)                                                                           | Fuzzy logic                    | در مدل فازی مقدار MAPE برای بخش خانگی، تجاری، صنعتی و سایر بخش ها به ترتیب ۱/۶۷، ۱/۴۲، ۲/۳۸ و ۰/۹۳ درصد به دست آمده است.                                                                                                                                               | MAPE                | ۲۰۱۶ |
| [۱۹]  | بلند مدت     | سطح توسعه اقتصادی، نرخ رشد شدت انرژی صنعت، نرخ رشد سهم صنعت ثالث، نرخ رشد جمعیت، متوسط دمای سالانه، واردات و صادرات انرژی، قیمت انرژی، نرخ رشد GDP<br>(داده های ۱۹۹۰-۲۰۱۳)                                   | Hybrid (PSO+GA+ RBF)           | نتایج پیش بینی نشان می دهد مدل PSO-GA-RBF دقت بیشتری نسبت به مدل شبکه عصبی مصنوعی دارد. مصرف برق ووهان با نرخ رشد متوسط سالانه حدود ۹/۷ تا ۱۱/۵ درصد به سرعت رشد خواهد کرد.                                                                                            | MAPE                | ۲۰۱۵ |



شکل ۱. موقعیت جغرافیایی استان گیلان<sup>۱</sup>

## ۲.۲. تحلیل مصرف برق استان گیلان<sup>۱</sup>

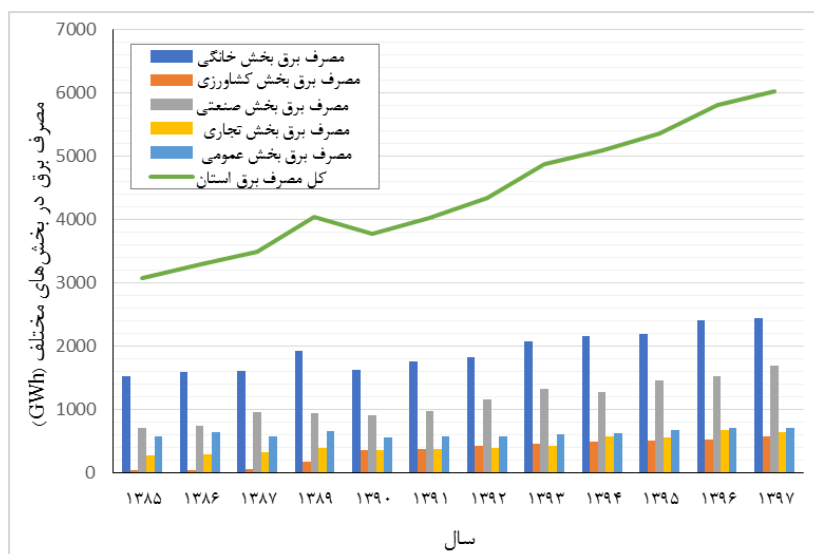
استان گیلان یکی از استان‌های غیرصنعتی در کشور است و بیشترین مصرف برق آن از طریق مصرف بخش خانگی است. هر ساله با افزایش جمعیت، پیشرفت تکنولوژی تجهیزات کشاورزی، ورود صنایع کوچک، رشد مناطق کم‌برخوردار و کابل‌کشی برق در این مناطق باعث افزایش مشترکان برق در این استان می‌شود. شکل ۲ مصرف برق در بخش‌های مختلف و همچنین، کل برق مصرفی را از سال ۱۳۸۵ تا ۱۳۹۷ برحسب گیگا وات ساعت (GWh) نشان می‌دهد.

مصرف برق استان گیلان رشد ۴۸/۹۶ درصد طی ۱۲ سال گذشته و نرخ رشد متوسط سالیانه ۰/۰۲۶ درصد را به خود اختصاص داده است. بار شبکه استان گیلان به دلیل شرایط اقلیمی تحت تأثیر شرایط فصلی است، به گونه‌ای که میزان حداکثر بار شبکه در سه ماه گرم سال تیر، مرداد و شهریور تقریباً دو برابر ماه‌های سرد سال یعنی مهر تا بهمن است. تعداد چاه‌های برق‌دار شده طبق آخرین آمار، ۲۰۶۲ عدد بوده که نسبت به سال ۱۳۹۶،

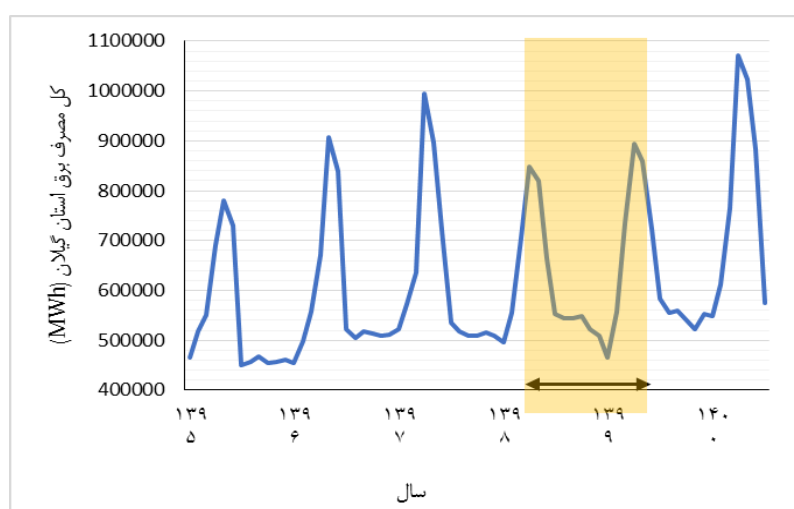
حدود ۱۷۹/۴ درصد رشد داشته است. در درازمدت تعداد مشترکان برق نیز روی مصرف برق مؤثر است. این شاخص نسبت به سال ۱۳۹۰ نیز ۳۰/۹۶ درصد افزایش داشته که بیشترین افزایش را بخش خانگی به خود اختصاص داده است. از طرفی، شیوع ویروس کرونا علاوه بر اینکه سلامت جوامع را به خطر می‌اندازد، فعالیت‌های اقتصادی و صنعتی را نیز تحت تأثیر قرار داده است، به گونه‌ای که ابعاد اپیدمی این بیماری قابل مقایسه با بحران‌های قبلی نیست. حوزه‌های مختلف مانند برق که زیرساخت شریان اقتصادی است، در صورت مشاهده تغییرات پایدار باید راهکارهایی برای مقابله با اثرات آن ارائه شود. از بهمن ۱۳۹۸ تا مرداد ۱۳۹۹ با ورود کرونا و اوج‌گیری آن، چرخه زندگی اجتماعی مردم به دلیل قرنطینه، با تغییراتی همراه بود و مصرف برق معطوف به افزایش تقاضا در بخش خانگی شد و در مقابل تقاضا در بخش‌های دیگر به نسبت روند کاهشی پیدا کرد، اما با تداوم شیوع کرونا و رفع برخی محدودیت‌ها، مصرف برق تقریباً به سطح قبل از کرونا بازگشته است. شکل ۳ تأثیر کرونا در این مدت را روی میزان تقاضا نشان می‌دهد.

## ۱. یافتہ محقق

سیمانی و میرآبادی: مدل سازی پیش بینی کوتاه مدت مصرف برق با استفاده از روش های هوش مصنوعی در استان گیلان



شکل ۲. نوع و ترکیب کل برق مصرفی استان گیلان طی دوره زمانی ۱۳۸۵ تا ۱۳۹۷ [۲۰]



شکل ۳. بررسی تأثیر کرونا در مصرف برق [۲۰]

### ۳.۲. پارامترهای مستقل تحقیق

در این مقاله الگوریتم های یادگیری ماشین، در محیط Python و با استفاده از Google Colab، پیاده سازی شدند. در پیش بینی کوتاه مدت، داده های ساعتی مصرف برق کل استان به عنوان تابع هدف انتخاب شده است. داده های ساعتی ۶ سال از سال ۱۳۹۵ تا سال ۱۴۰۰ از برق منطقه ای استان گیلان تهیه شد [۲۱]. داده های هواشناسی مورد استفاده در این تحقیق به ازای هر ۳ ساعت در شبانه روز از سامانه هواشناسی کشور دریافت شده است [۲۲]. طول روز یکی از عواملی است که در میزان مصرف برق تأثیر بسزایی دارد [۲۳]. این ویژگی به صورت روزانه در

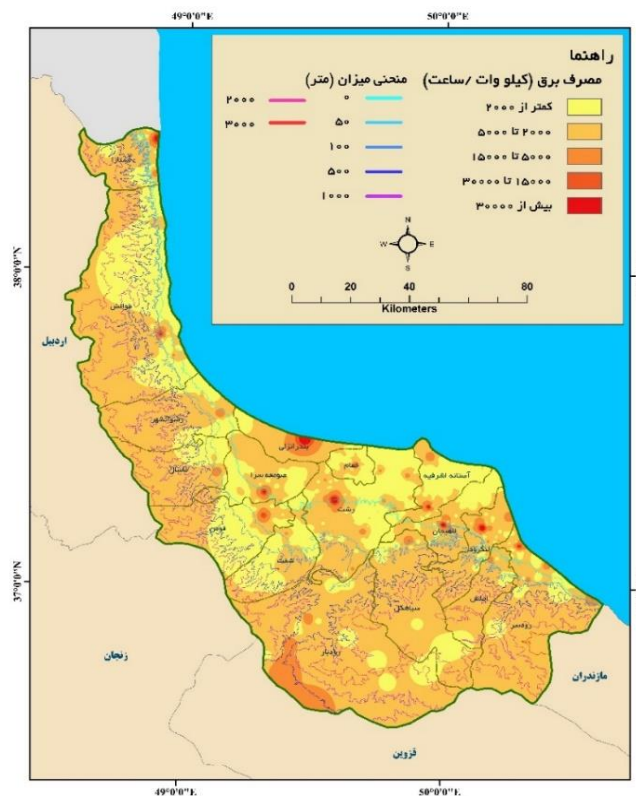
دسترس بوده و در مرحله پیش پردازش به داده ساعتی تبدیل خواهد شد. داده های مورد بررسی شامل متغیرهای مستقل، طول روز، تعطیلات، دما، رطوبت نسبی، سرعت باد، نقطه شبنم، میزان مصرف برق برای یک ساعت تا ۲۴ ساعت گذشته، میانگین میزان مصرف برق ۲۴ ساعت گذشته، میانگین مصرف یک هفته تا یک ماه گذشته است.

### ۴.۲. پیش پردازش داده ها

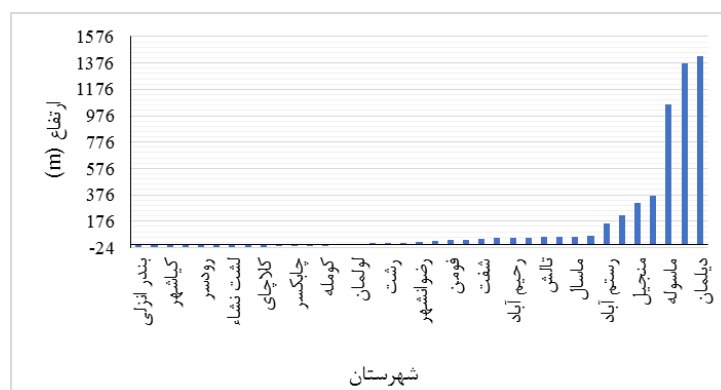
برای افزایش دقت مدل پیش بینی، پیش پردازش روی تمامی داده ها انجام می شود. داده های واقعی معمولاً کیفیت لازم جهت ورود به الگوریتم های یادگیری ماشین را ندارند و اغلب دارای داده های ازدست رفته و نویز بوده و باعث افزایش خطای

به همین منظور، استان گیلان به سه ناحیه با ارتفاع پایین، متوسط و بالا تقسیم شد. شکل ۵ شهرستان‌های استان گیلان را بر اساس افزایش ارتفاع آن‌ها نشان می‌دهد. شهر بندرانزلی با ارتفاع ۲۳- متر، شهرستانی با کمترین ارتفاع و به طور معمول گرم‌تر و شرجی‌تر و شهرستان دیلمان با ارتفاع ۱۴۲۵ متر، مرتفع‌ترین شهرستان استان گیلان است. شهر رشت نیز به دلیل جمعیت بالاتر و تمرکز بار صنعتی در آن به عنوان شهر با ارتفاع ۳ متر و متوسط در نظر گرفته می‌شود.

مدل پیش‌بینی خواهند شد. با توجه به کیفیت داده‌ها و مشکلات آن‌ها، از روش‌های مختلفی مانند نرمال‌سازی، استانداردسازی، انتخاب ویژگی و... استفاده می‌شود. دما یکی از مهم‌ترین متغیرهای مؤثر شرایط آب‌وهوایی بر مصرف برق در کوتاه‌مدت است. با افزایش دما، سهم استفاده از وسایل سرمایشی افزایش یافته و تقاضای بار شبکه نیز افزایش می‌یابد. همچنین، هرچه ارتفاع از سطح دریا افزایش یابد از مقدار دما کاسته می‌شود. شکل ۴ ارتباط بین مصرف برق و ارتفاع را به خوبی نشان می‌دهد.



شکل ۴. ارتباط مقدار مصرف برق هر منطقه با ارتفاع در استان [۲۰ و ۲۱]



شکل ۵. ترتیب شهرستان‌های استان گیلان بر حسب افزایش ارتفاع آن‌ها نسبت به سطح دریا

سیمانی و میرآبادی: مدل سازی پیش بینی کوتاه مدت مصرف برق با استفاده از روش های هوش مصنوعی در استان گیلان

اول حل مشکل مقادیر ازدست رفته در داده های موجود با جایگزینی آن ها با مقادیر قبلی است.

$$F_{1new} = \frac{2F_{1avr} + F_{2avr}}{3} \quad (2)$$

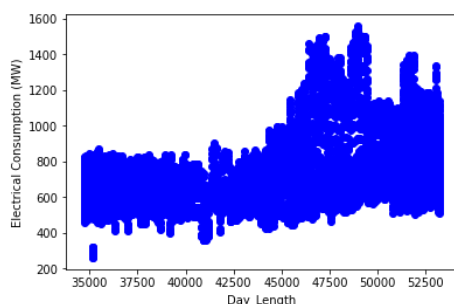
$$F_{2new} = \frac{F_{1new} + F_{2avr}}{2} \quad (3)$$

$F_{1avr}$  و  $F_{2avr}$  داده های هواشناسی اول و چهارم و  $F_{1new}$  و  $F_{2new}$  مقادیر جدید دوم و سوم داده هواشناسی است. شکل ۶ ارتباط متغیرهای مستقل در مدل پیش بینی با متغیر وابسته مصرف برق را نشان می دهد.

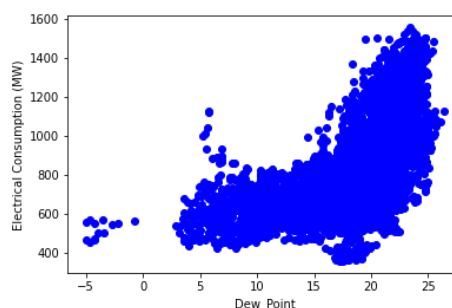
بنابراین، با توجه به میزان اهمیت سه شهرستان، در افزایش یا کاهش دما، میانگین وزنی به ازای هر ساعت برای دستیابی به میانگین دما برای کل استان در نظر گرفته شده است.

$$T_{ave} = \frac{0.1T_d + 0.6T_b + 0.3T_r}{3} \quad (1)$$

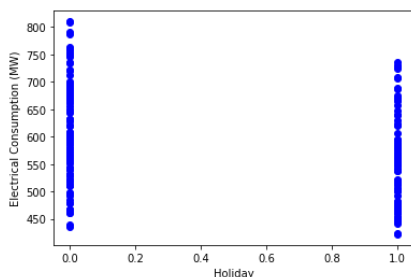
که در آن  $T_b$  نشان دهنده دمای شهر بندرانزلی با بالاترین ضریب وزنی،  $T_r$  دمای شهر رشت با ارتفاع متوسط و  $T_d$  دمای شهر دیلمان با کمترین ضریب وزنی است. گام بعدی پیش پردازش برای تمامی ویژگی های هواشناسی است. گام



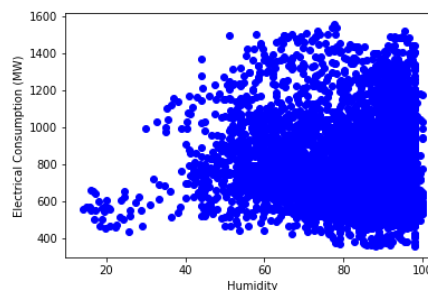
(ب) طول روز



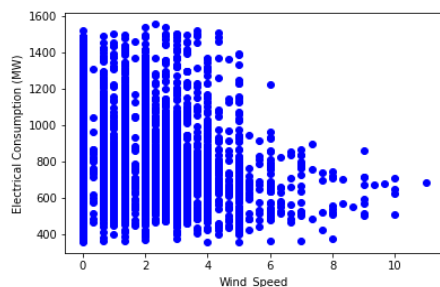
(الف) نقطه شبنم



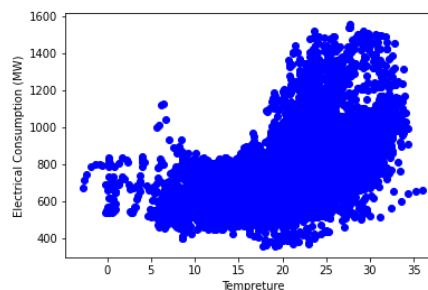
(ت) تعطیلات



(پ) رطوبت نسبی



(خ) سرعت باد



(ج) دما

شکل ۶. ارتباط متغیرهای مستقل با مصرف برق<sup>۱</sup>

داده‌های ورودی و پاسخ و معادله ۶ بیان‌کننده یک مدل رگرسیونی است [۲۴].

$$x = \begin{bmatrix} \dots & x_1 & \dots \\ \dots & \dots & \dots \\ \dots & x_n & \dots \end{bmatrix}, \quad Y = \begin{bmatrix} y_1 \\ \dots \\ y_n \end{bmatrix} \quad (6)$$

$$h(x) = \sum_{i=0}^d w_i x_i + e_i = W^T X \quad (7)$$

$w_i$  تعداد متغیرهای مستقل یا ورودی‌های مدل، وزن بایاس هر ورودی،  $x_i$  متغیرهای مستقل و  $e_i$  مقدار خطا است. هدف، یافتن وزن‌های بایاس بهینه است تا  $h(x)$  بهترین تقریب برای خروجی  $y$  یا همان متغیر وابسته باشد. در داده‌های آموزش، میانگین مربعات خطا<sup>۲</sup> به صورت معادله ۵ محاسبه می‌شود [۲۴]:

$$E_{in}(h) = E_{in}(W) = \frac{1}{n} \sum_{i=1}^n (h(x_i) - y_i)^2 = \frac{1}{n} \sum_{i=1}^n (W^T x_i - y_i)^2 = \frac{1}{n} \|X_w - y\|^2 \quad (8)$$

هدف از بررسی و اعمال یک مدل پیش‌بینی، کاهش خطای درون شبکه یا همان خطای داخلی مربوط به داده‌های آموزش است. در نتیجه، باید  $E_{in}(W)$  به حداقل برسد [۲۴]:

$$E_{in}(W) = \frac{1}{n} \|X_w - y\|^2 \rightarrow \nabla E_{in}(W) = \frac{2}{n} X^T (X_w - y) = 0 \quad (9)$$

حداقل بودن مقدار  $E_{in}(W)$  وابسته به مقدار بهینه برای وزن‌های ورودی است. اگر داده‌های ورودی چند بعدی  $X$ ، معکوس‌پذیر باشد با توجه به رابطه ۷، وزن‌های بهینه به دست می‌آیند [۲۴]:

$$w = X^{-1} y \quad (10)$$

در غیر این صورت، معکوس مجازی  $X$  به جای معکوس واقعی برای دستیابی به وزن بهینه استفاده می‌شود [۲۴]:

$$X^+ = (X^T X)^{-1} X^T \rightarrow w = X^+ y \quad (11)$$

$$X^+ (d+1) \times n = \left( X_{(d+1)n}^T X_{(d+1)n} \right)^{-1} X_{(d+1)n}^T \rightarrow \quad (12)$$

$$w_{(d+1)n \times 1} = X_{(d+1)n \times n}^+ y_{n \times 1}$$

$d$  و  $n$  بیان‌کننده تعداد ویژگی‌های به کار برده شده و تعداد داده‌های آموزش در مدل پیش‌بینی است.

## ۲.۵. الگوریتم خوشه‌بندی K\_Means

الگوریتم خوشه‌بندی، یک الگوریتم غیرنظارتی در یادگیری ماشین است. خوشه‌بندی فرایند شناسایی گروه‌ها یا خوشه‌های طبیعی در داده‌های چندبعدی بر اساس برخی معیارهای شباهت است. مراحل فرایند خوشه‌بندی شامل استانداردسازی داده‌ها، انتخاب داده‌های مناسب، اعمال روش‌های خوشه‌بندی روی داده‌ها، اعتبارسنجی خوشه‌های تولیدشده و تفسیر نتایج است. فرایند خوشه‌بندی به روش K\_means، یک روش ترکیبی است. در این روش ابتدا، از قبل تعداد خوشه‌ها ( $K$ ) تعیین و سپس  $k$  نقطه به صورت تصادفی به عنوان مراکز خوشه انتخاب می‌شوند. محبوب‌ترین روش برای ارزیابی شباهت، استفاده از اندازه‌گیری فاصله است. فاصله اقلیدسی یکی از ابزارهای اندازه‌گیری فاصله است [۱۴].

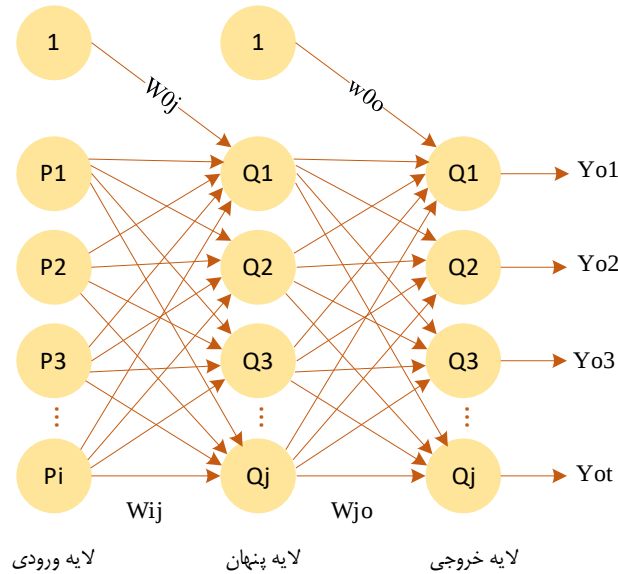
$$Dist(x, x') = \sqrt{\sum_{i=1}^m (x_i - x'_i)^2} \quad (4)$$

$m$  تعداد ویژگی‌ها،  $x_i$  و  $x'_i$  دو نقطه از داده در ویژگی‌ها است. بعد از خوشه‌بندی، میانگین نمونه‌های هر خوشه محاسبه می‌شود. این میانگین‌ها به عنوان مقادیر مرکزی جدید برای خوشه‌های مربوطه در نظر گرفته می‌شوند. در نهایت، کل فرایند با مراکز خوشه جدید تکرار می‌شود. تکرار ادامه می‌یابد تا زمانی که در دوره‌های متوالی نقاط مشابهی به هر خوشه اختصاص داده شود، در این مرحله مراکز خوشه تثبیت می‌شوند و برای همیشه ثابت خواهند ماند. معیار مجموع مربعات خطا<sup>۱</sup>، برای ارزیابی مناسب بودن هر خوشه مورد استفاده قرار می‌گیرد. در تکرارهای متوالی، هر خوشه‌ای که کوچک‌ترین مقدار SSE را داشته باشد به عنوان خوشه مناسب است [۱۴].

$$SSE = \sum_{i=1}^K \sum_{x \in C_i} [dist(m_i, x)]^2 \quad (5)$$

## ۲.۶. الگوریتم رگرسیون خطی

رگرسیون خطی، هم در مسائل طبقه‌بندی و هم رگرسیونی به کار برده می‌شود. در طبقه‌بندی، داده‌ها به دو دسته صفر و یک تقسیم می‌شوند. در مسائل رگرسیونی، خروجی یک عدد است.  $X$  و  $Y$  به ترتیب



شکل ۷. معماری شبکه عصبی پرسپترون چندلایه [۱۵]

## ۲.۷. الگوریتم شبکه عصبی مصنوعی

الگوریتم ANN، یک گروه بزرگ و مهم از مدل های یادگیری تحت نظارت و شبکه ای از واحدهای پردازش غیرخطی ریاضی هستند که مطابق با مغز انسان طراحی شده اند. مغز انسان از تعداد بسیار زیادی نورون<sup>۱</sup> تشکیل شده که به هم متصل هستند و با یکدیگر ارتباط دارند. این شبکه ها نیز مانند مغز، قابلیت آموزش، حفظ کردن و تحلیل داده ها و نتایج را دارند. ساختار کلی شبکه عصبی از سه بخش لایه ورودی<sup>۲</sup>، یک یا چند لایه وسط یا همان لایه پنهان<sup>۳</sup> و لایه خروجی<sup>۴</sup> در شکل ۷ آورده شده است. لایه ورودی همان سلول های دریافت کننده اطلاعات است، لایه پنهان وظیفه پردازش اطلاعات و لایه خروجی علاوه بر پردازش اطلاعات، وظیفه نمایش خروجی را نیز دارد. مراحل طراحی الگوریتم ANN شامل تقسیم داده ها، تعیین معماری شبکه، تعیین تابع فعال ساز، تعیین تابع بهینه ساز برای بروزرسانی وزن ها و تابع ضرر است [۳].

$$Y_t = W_0 + \sum_{j=1}^Q W_{0j} g \left( W_{0j} + \sum_{i=1}^P W_{ij} y_{t-1} \right) \quad (13)$$

$P$  تعداد گره های ورودی،  $Q$  تعداد گره های پنهان و  $g$  تابع فعال سازی است.  $\{W_{ij}, j=0, 1, 2, \dots, Q\}$  بردار وزن از

لایه پنهان به گره های خروجی،  $j=1, 2, \dots, P, j=1, 2, \dots, Q\}$  بردار وزن های ورودی به گره های پنهان و  $W_{0j}$  وزن های هر خروجی بین لایه ورودی و پنهان است. تابع فعال ساز در این مقاله، شامل خطی و سیگموئید است. توابع بهینه ساز هنگام آموزش شبکه مورد استفاده قرار می گیرند. وظیفه این توابع، تغییر ویژگی های شبکه عصبی مانند وزن و نرخ یادگیری به منظور کاهش تلفات است. از روش های بهینه سازی در این پژوهش، می توان به Adamax و Adam اشاره کرد. روش Adam اولین بار در سال ۲۰۱۴ معرفی شد. این روش با تکانه های اول و دوم سروکار دارد. استفاده از میانگین ها باعث می شود که الگوریتم با سرعت بیشتری به سمت حداقل ها همگرا شود [۲۶].

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t \quad (14)$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2$$

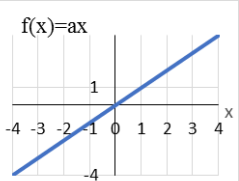
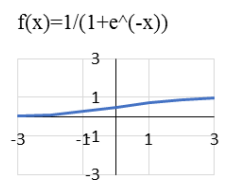
که در آن  $m_t$  میانگین یا تخمین تکانه اول،  $v_t$  تخمین تکانه دوم یا واریانس غیرمرکزی گرادیان ها و مقادیر  $\beta_1$  و  $\beta_2$  به ترتیب ۰/۹ و ۰/۹۹۹ است.  $m_t$  و  $v_t$  به روش زیر محاسبه شده و سپس برای به روزرسانی پارامترها استفاده می شود.

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t} \quad (15)$$

$$\hat{v}_t = \frac{v_t}{1 - \beta_2^t}$$

1. Neurons
2. Input layers
3. Hidden layers
4. Output layer

جدول ۲. توابع فعال‌ساز [۲۵]

| نام تابع فعال‌سازی | نوع تابع                      | شکل تابع                                                                          |
|--------------------|-------------------------------|-----------------------------------------------------------------------------------|
| تابع خطی           | $f(x) = x$                    |  |
| سیگموئید           | $f(x) = \frac{1}{1 + e^{-x}}$ |  |

$$\theta_{t+1} = \theta_t - \frac{\eta}{u_t} \widehat{m}_t \quad (20)$$

مقادیر رایج برای  $\eta$  و  $\beta_1$  و  $\beta_2$  به ترتیب ۰/۰۰۲، ۰/۹ و ۰/۹۹۹ است [۲].

#### ۲.۸. الگوریتم رگرسیون بردار پشتیبان

الگوریتم رگرسیون بردار پشتیبان، بر اساس مدل‌های طبقه‌بندی الگوریتم ماشین بردار پشتیبان ارائه می‌شود. هدف از این الگوریتم، به حداقل رساندن خطا، با شناسایی تابعی است که تعداد بیشتری از نقاط را در داخل لوله قرار می‌دهد. در SVR، داده‌های ورودی اصلی  $x$  به صورت غیرخطی در یک فضای ویژگی ابعاد بالاتر  $R^h$  با استفاده از یک تابع نگاشت غیرخطی  $\varphi: R^n \rightarrow R^h$  نگاشت می‌شود [۲۷].

$$M = \left\{ \left( (x_i, y_i) \mid i = 1, 2, \dots, N \right) \mid x_i \in R^N, y_i \in R \right\} \quad (21)$$

$$y_i = f(x) = \omega^T \cdot \varphi(X) + b \quad (22)$$

$x_i$  ورودی‌های مدل دارای  $N$  متغیر و  $y_i$  خروجی مدل است. در فرمول ۲۲،  $b$  یک مقدار ثابت،  $\omega$  بردار وزنی و  $\varphi(X)$  نشان‌دهنده تابع نگاشت در فضای ویژگی است. ضرایب  $b$  و  $\omega$  با به حداقل رساندن تابع ریسک منظم  $R$  تخمین زده می‌شود.

$$R = \text{Minimize} \quad \frac{1}{2} \|\omega\|^2 + C \frac{1}{N} \sum_{i=1}^N L_\epsilon(y_i - f(x_i), x_i) \quad (23)$$

به‌روزرسانی در Adam به شرح زیر است:

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{v_t} + \epsilon} \widehat{m}_t \quad (16)$$

مقدار  $\epsilon$ ،  $10^{-8}$  فرض می‌شود [۲۶].

الگوریتم AdaMax توسعه‌ای برای الگوریتم بهینه‌سازی Adam و براساس میانگین بی‌نهایت است. به طور کلی، AdaMax به طور خودکار نرخ یادگیری را برای هر پارامتر در مسئله بهینه‌سازی تطبیق می‌دهد. ضریب  $v_t$  در قانون به‌روزرسانی Adam، گرادینان را به طور معکوس با دو گرادینان گذشته ( $v_{t-1}$ ) و گرادینان فعلی ( $g_t^2$ ) مقیاس می‌کند.

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) |g_t|^2 \quad (17)$$

می‌توان به‌روزرسانی را به میانگین  $p$  تعمیم داد.

$$v_t = \beta_2^p v_{t-1} + (1 - \beta_2^p) |g_t|^p \quad (18)$$

میانگین‌های مقادیر  $p$  بزرگ، عموماً از نظر عددی ناپایدار و میانگین‌های  $\ell_1$  و  $\ell_2$  رایج هستند.  $\ell_\infty$  نیز به طور کلی رفتار پایدار نشان می‌دهد. به همین دلیل،  $v_t$  با  $\ell_\infty$  به مقدار پایدارتری همگرا می‌شود. برای جلوگیری از اشتباه گرفتن با روش Adam، از  $u_t$  برای نشان دادن میانگین بی‌نهایت محدود  $v_t$  استفاده می‌شود:

$$u_t = \beta_2^\infty v_{t-1} + (1 - \beta_2^\infty) |g_t|^\infty = \max(\beta_2 \cdot v_{t-1}, |g_t|) \quad (19)$$

با جایگزینی عبارت  $u_t$  در فرمول ۳۹ به جای  $\sqrt{v_t} + \epsilon$  فرمول ریاضی روش Adamax به دست می‌آید:

## ۲.۹. معیارهای اندازه گیری خطا

برای اندازه گیری مقدار خطای داخلی هر الگوریتم از معیارهای اندازه گیری استفاده می شود. توابع مختلفی برای ارزیابی مقدار خطا مورد استفاده قرار می گیرند. این معیارها برای ارزیابی عملکرد نتایج الگوریتم های یادگیری ماشین مورد استفاده قرار می گیرند. هرچه مقادیر معیارهای اندازه گیری خطا کمتر باشد، دقت مدل بیشتر و مدل قابل اعتمادتر است. در مقابل، معیار ضریب تعیین نیز برای ارزیابی دقت مدل مورد استفاده قرار می گیرد. معیارهای سنجش خطا و دقت مورد استفاده در این مقاله در معادلات (۳۰) تا (۳۴) آورده شده است [۲۹].

$$MAE = \sum_{t=1}^N \left| \frac{A_t - F_t}{N} \right| \quad (30)$$

$$MSE = \frac{1}{N} \sum_{t=1}^N (A_t - F_t)^2 \quad (31)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^N (A_t - F_t)^2} \quad (32)$$

$$MAPE = \frac{100}{N} \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right| \quad (33)$$

$$R^2 = 1 - \frac{\sum_{t=1}^N (A_t - F_t)^2}{\sum_{t=1}^N (A_t - \bar{A})^2} = 1 - \frac{RSS}{TSS} \quad (34)$$

که در آن  $N$ ، تعداد مقادیر پیش بینی شده،  $F_t$  مقدار پیش بینی شده در زمان  $t$ ،  $A_t$  مقدار واقعی و  $\bar{A}$  میانگین داده های واقعی است.

## ۳. یافته ها

هنگام توسعه مدل ها برای پیش بینی یا داده کاوی، پیش بینی کنندگان معمولاً داده ها را به سه بخش، آموزش، اعتبارسنجی و آزمایش تقسیم می کنند. از داده های آموزشی برای تخمین پارامترها استفاده می شود. از داده های اعتبارسنجی برای انتخاب مدل ها و از داده های تست برای تأیید عملکرد مدل استفاده می شود. در این مقاله، از داده های سال های ۱۳۹۵، ۱۳۹۶، ۱۳۹۷ برای مرحله آموزش، داده های سال ۱۳۹۸ برای مرحله اعتبارسنجی نمونه های آموزش دیده شده و از داده های سال های ۱۳۹۹ و ۱۴۰۰ برای مرحله آزمایش استفاده می شود. جدول ۳ تعداد داده های آموزش و تست را در ۵ خوشه نشان می دهد.

عبارت  $\frac{1}{2} ||w||^2$  هموار بودن تابع را اندازه گیری می کند.  $C$  و  $\varepsilon$  پارامترهای تجویز شده مدل هستند. پارامتر  $C$ ، خطای بین مقادیر پیش بینی شده و مقادیر واقعی در داده های آموزش را ارزیابی و  $\varepsilon$ ، پهنای لوله را در اطراف تابع تخمین<sup>۱</sup> تعیین می کند.  $L_\varepsilon$  تابع تلفات غیر حساس است.

$$L_\varepsilon(y - f(x), x) = \begin{cases} 0, & |y - f(x)| < \varepsilon \\ |y - f(x)| - \varepsilon, & \text{otherwise} \end{cases} \quad (24)$$

$\xi_i^*$  و  $\xi_i^*$  متغیرهای کمکی که میزان فاصله داده ها در بالا یا پایین لوله  $\varepsilon$  را کمی می کنند.

$$|y - f(x)| - \varepsilon = \xi_i \quad \text{نقاط بالای لوله} \quad (25)$$

$$|y - f(x)| - \varepsilon = \xi_i^* \quad \text{نقاط پایین لوله} \quad (26)$$

$$R = \frac{1}{2} ||w||^2 + C \sum_{i=1}^N (\xi_i + \xi_i^*) \quad (27)$$

$$y_i - (\omega^T \varphi(x_i) + b) \leq \varepsilon + \xi_i$$

$$(\omega^T \varphi(x_i) + b) - y_i \leq \varepsilon + \xi_i^* \quad (28)$$

$$\xi_i, \xi_i^* \geq 0; i = 1, 2, \dots, N$$

پارامترهای  $f(x)$  را می توان با استفاده از تابع لاگرانژ به دست آورد.

$$f(x) = \omega^T \varphi(x) + b = \sum_{i=1}^N (\alpha_i^* - \alpha_i) K(x_i, x) + b \quad (29)$$

$\alpha_i$  و  $\alpha_i^*$  ضرب کننده های لاگرانژی هستند. عبارت  $K(x_i, x_j)$  تابع هسته<sup>۲</sup> بوده و به صورت حاصل ضرب نقطه ای بین  $\varphi(x_i)^T$  و  $\varphi(x)$  است [۲۷].

روش هایی مانند جست و جوی شبکه و الگوریتم ژنتیک برای بهینه سازی فرایمترهای الگوریتم یادگیری ماشینی به کار برده می شوند. روش کار این الگوریتم به این صورت است که داده های آموزش را به دو دسته داده های آموزش و اعتبارسنجی تقسیم کرده و با استفاده از داده های آموزش و فرایمترهای از قبل تنظیم شده، مدل را آموزش می دهد. سپس، روی داده های اعتبارسنجی تست می کند. این عمل تا جایی تکرار می شود که بهترین فرایمترها انتخاب شوند [۲۸].

1. Hyperplane  
2. Kernel Trick

جدول ۳. تعداد داده‌های آموزش و تست در هر خوشه

| خوشه | تعداد داده‌های آموزش | تعداد داده‌های تست |
|------|----------------------|--------------------|
| ۱    | ۸۹۰۴                 | ۲۲۳۲               |
| ۲    | ۷۱۳۳                 | ۱۴۹۱               |
| ۳    | ۶۱۲۰                 | ۱۹۷۷               |
| ۴    | ۱۰۶۰۶                | ۲۱۴۵               |
| ۵    | ۵۹۴۴                 | ۱۸۳۲               |

## ۳.۱. نتایج پیاده‌سازی الگوریتم LR

متغیر مستقل شامل: داده‌های هواشناسی، داده تعطیلات، داده مصرف برق، داده طول روز و یک متغیر وابسته یا پاسخ شامل مصرف برق در این مقاله در نظر گرفته شده است. مدل رگرسیونی تحقیق به صورت رابطه ۳۵ است:

$$Y = \sum_{j=0}^{34} W_j x_j + e_i \quad (35)$$

که در آن پارامتر  $W_j$  زیانگر تعداد متغیرهای مستقل است. در این الگوریتم برای محاسبه مقدار خطای پیش‌بینی، از MAE، RMSE، MAPE و برای ارزیابی دقت آن از معیار  $R^2$  استفاده شده است. در جدول ۴ معیارهای اندازه‌گیری خطا در ۵ خوشه با استفاده از روش LR نشان داده شده است.

نمودار مقایسه دقت مقادیر واقعی و مقادیر پیش‌بینی شده بر حسب زمان به صورت شکل ۸ است. خطوط قرمز مقادیر پیش‌بینی شده و خطوط آبی مقادیر واقعی هستند. با استفاده از این شکل می‌توان توانایی مدل را در پیش‌بینی ساعت‌های کم‌باری، میان‌باری و اوج بار مشاهده کرد. همان‌طور که قابل مشاهده است، میزان اختلاف بین مقادیر واقعی و پیش‌بینی شده به‌جز در برخی

از ساعت‌های اوج بار با کمی اختلاف بسیار ناچیز است. به همین دلیل باید از مدل‌های دیگری برای بهبود نتایج و مقادیر پیش‌بینی در اوج بار استفاده کرد.

## ۳.۲. نتایج پیاده‌سازی الگوریتم ANN

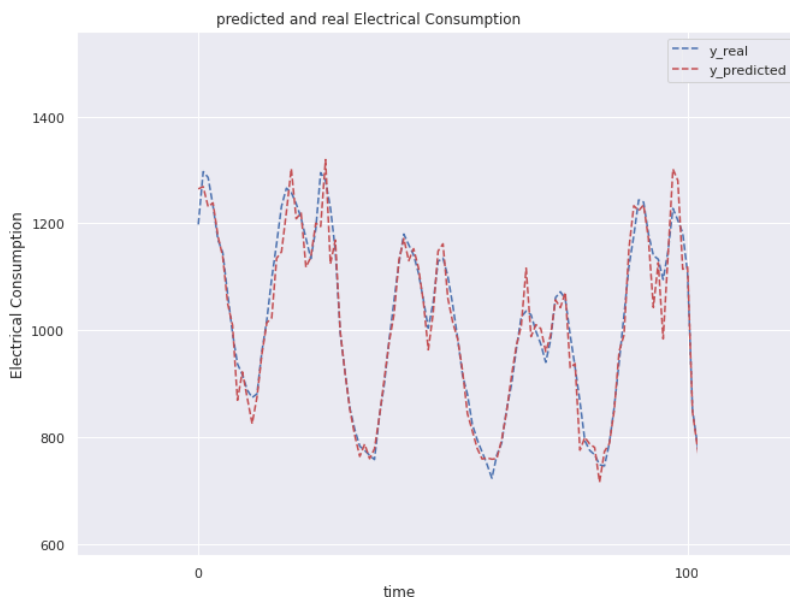
لایه ورودی شامل داده‌های مستقل ورودی اعم از داده‌های هواشناسی، روزهای تعطیل، داده‌های مربوط به طول روز و مصارف برق یک ساعت تا ۴ هفته قبل و میانگین آن‌ها است. در لایه پنهان، هیچ قانون کلی برای تعیین میزان نورون‌های این لایه‌ها وجود ندارد و می‌توان با اتکا به روش تجربی تعداد مطلوب را تعیین کرد. اگر تعداد لایه‌ها کم باشد، شبکه نمی‌تواند ارتباط پیچیده بین ورودی و خروجی را پیدا کند و با خطای بالایی آموزش می‌بیند. شکل ۹ معماری الگوریتم شبکه عصبی مصنوعی مقاله را نشان می‌دهد.

مطابق با جدول ۵ تعداد هر دوره از آموزش ۳۰۰ است و با استفاده از الگوریتم شبکه عصبی پیش‌خور در هر بار آموزش شبکه برای هر خوشه، مقادیر خطا در خروجی با استفاده از تابع خطای MSE محاسبه می‌شود. سپس، با استفاده از الگوریتم پسانتشار خطا، با کمک تابع بهینه‌ساز Adam و Adamax، مقادیر وزن و بایاس در ساختار لایه‌های شبکه عصبی به‌روزرسانی می‌شود.

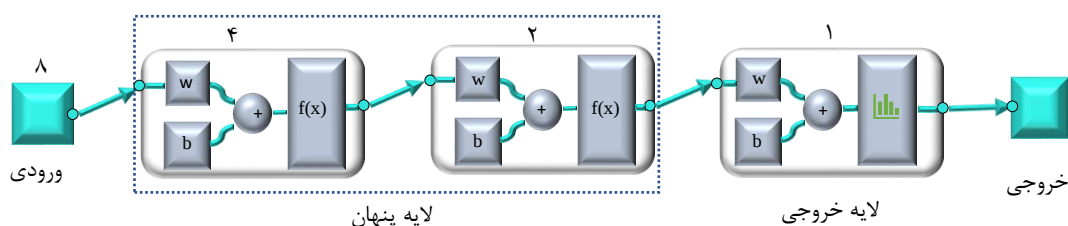
جدول ۴. مقادیر معیارهای اندازه‌گیری خطا در خوشه‌های مختلف در الگوریتم LR<sup>۱</sup>

| شماره خوشه    | MSE      | RMSE    | MAE    | MAPE   | R <sup>2</sup> |
|---------------|----------|---------|--------|--------|----------------|
| ۱             | ۴۰۶،۶۵۰  | ۲۰،۱۶۵  | ۱۴،۴۱۹ | ۰،۰۱۹۳ | ۰،۹۶۲          |
| ۲             | ۲۲۴۰،۸۴۱ | ۴۷،۳۳۷۵ | ۳۱،۱۴۸ | ۰،۰۲۶۴ | ۰،۹۸۶          |
| ۳             | ۴۴۱،۴۹۸  | ۲۱،۰۱۱  | ۱۵،۸۲۴ | ۰،۰۲۰۵ | ۰،۹۵۲          |
| ۴             | ۴۵۲۰،۹۷۲ | ۶۷،۲۳۸  | ۴۲،۲۲۲ | ۰،۰۳۵۸ | ۰،۹۶۳          |
| ۵             | ۸۳۲،۷۹۰  | ۲۸،۸۵۸  | ۱۹،۸۴۰ | ۰،۰۲۳۵ | ۰،۹۶۵          |
| مجموع خوشه‌ها | ۱۶۸۹،۰۲۹ | ۴۱،۰۹۷  | ۲۴،۴۷۳ | ۰،۰۲۵۱ | ۰،۹۸۲          |

سیمانی و میرآبادی: مدل سازی پیش بینی کوتاه مدت مصرف برق با استفاده از روش های هوش مصنوعی در استان گیلان



شکل ۸. مقایسه مقادیر واقعی و پیش بینی شده در الگوریتم LR



شکل ۹. معماری الگوریتم ANN

جدول ۵. پارامترهای اجرای الگوریتم ANN

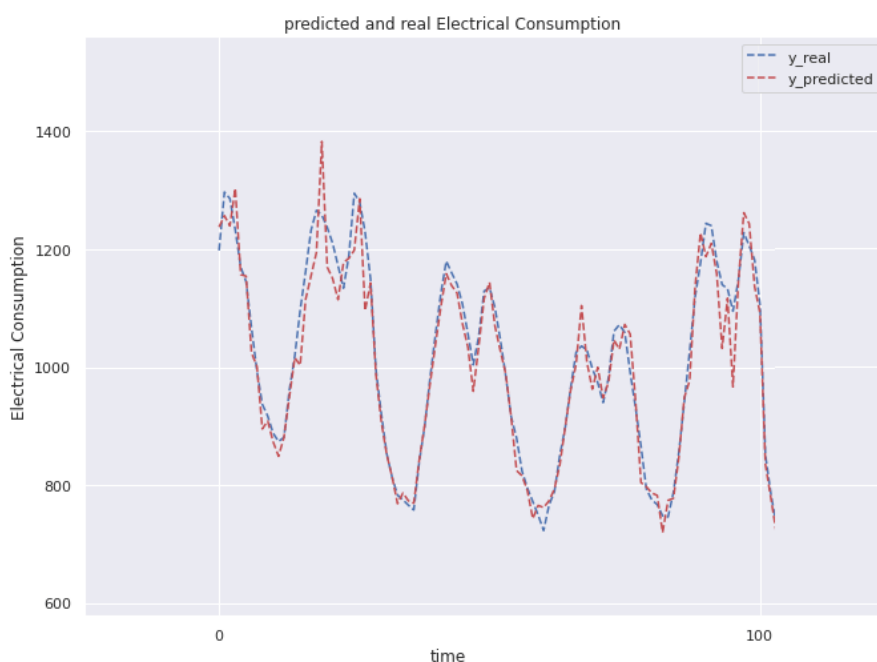
| شماره خوشه | epoch | تابع بهینه ساز | تابع خطا یا ضرر |
|------------|-------|----------------|-----------------|
| ۱          | ۳۰۰   | Adam           | MSE             |
| ۲          | ۳۰۰   | Adamax         | MSE             |
| ۳          | ۳۰۰   | Adamax         | MSE             |
| ۴          | ۳۰۰   | Adamax         | MSE             |
| ۵          | ۳۰۰   | Adamax         | MSE             |

ساختار هر الگوریتم به کار برده می شود، در خروجی آن ها تأثیرگذار است و انتخاب بهترین الگوریتم بسیار مهم است. نمودار مقایسه دقت مقادیر واقعی و مقادیر پیش بینی شده بر حسب زمان به صورت شکل ۱۰ است. مطابق شکل، الگوریتم ANN، عملکرد بهتری نسبت به الگوریتم رگرسیون خطی دارد. در ساعت های اوج بار نیز اختلاف بین مقادیر پیش بینی شده و مقادیر واقعی بسیار کمتر است.

نتایج حاصل از اجرای الگوریتم در ۵ خوشه و مجموع خوشه ها در جدول ۶ آورده شده است. نتایج به صورت میانگین هستند و در هر خوشه چندین بار الگوریتم اجرا می شود چون در هر بار اجرا نتایج متفاوت است، نتیجه نهایی به صورت میانگین به دست می آید. الگوریتم ANN در مقایسه با الگوریتم LR، میزان دقت را در تمامی خوشه ها به جز خوشه ۵ بهبود داده است. دلیل آن هم این است که براساس نوع داده ها، آموزش و پارامترهایی که در

جدول ۶. مقادیر معیارهای اندازه‌گیری خطای خوشه‌ها در الگوریتم ANN

| شماره خوشه    | MSE      | RMSE   | MAE    | MAPE   | $R^2$ |
|---------------|----------|--------|--------|--------|-------|
| ۱             | ۳۰۹,۷۸۹  | ۱۷,۶۰۰ | ۱۳,۲۴۴ | ۰,۰۱۷۸ | ۰,۹۷۱ |
| ۲             | ۲۲۶۶,۱۱۰ | ۴۷,۶۰۳ | ۳۱,۹۶۳ | ۰,۰۲۷۴ | ۰,۹۸۶ |
| ۳             | ۴۵۷,۰۸۵  | ۲۱,۳۷۹ | ۱۶,۰۷۷ | ۰,۰۲۰۸ | ۰,۹۵۱ |
| ۴             | ۴۳۸۹,۴۴۵ | ۶۶,۲۵۲ | ۴۴,۷۸۵ | ۰,۰۳۸۲ | ۰,۹۶۴ |
| ۵             | ۹۸۶,۲۸۰  | ۳۱,۴۰۵ | ۲۱,۸۶۳ | ۰,۰۲۵۶ | ۰,۹۵۹ |
| مجموع خوشه‌ها | ۱۶۷۳,۶۶۹ | ۴۰,۹۱۰ | ۲۵,۳۳۰ | ۰,۰۲۵۹ | ۰,۹۸۲ |



شکل ۱۰. مقایسه مقادیر واقعی و پیش‌بینی‌شده در الگوریتم ANN

استفاده نکردن از روش بهینه‌سازی برای  $C$  بزرگ، ناتوانی سیستم در بهینه کردن است. زیرا هرچه مقدار  $C$  بزرگ‌تر باشد، زمان اجرای مدل طولانی‌تر خواهد بود. جدول ۸. نتایج حاصل از اجرای مدل را با استفاده از معیارهای اندازه‌گیری خطا نشان می‌دهد. نمودار مقایسه دقت مقادیر واقعی و مقادیر پیش‌بینی‌شده بر حسب زمان به صورت شکل ۱۱ است. مطابق شکل، الگوریتم SVR، عملکرد بهتری نسبت به الگوریتم رگرسیون خطی و شبکه عصبی مصنوعی دارد. در ساعت‌های اوج بار نیز اختلاف بین مقادیر پیش‌بینی‌شده و مقادیر واقعی بسیار کمتر است.

### ۳.۳. نتایج پیاده‌سازی الگوریتم SVR و بهینه‌سازی توأم با روش جست‌وجوی شبکه

در این الگوریتم مقدار  $\epsilon$ ،  $\gamma$ ،  $C$  در یک بازه و تابع کرنل از بین چند تابع به صورت تصادفی انتخاب می‌شوند. الگوریتم جست‌وجوی شبکه، با این مقادیر آموزش دیده می‌شود و بهترین آن‌ها را پیدا می‌کند و سپس نتایج توسط الگوریتم SVR در هر خوشه و تمام خوشه‌ها ارائه می‌شود.

### ۳.۳.۱. فرآیند بهینه‌سازی SVR

جدول ۷ بهترین فرآیندها را توسط الگوریتم جست‌وجوی شبکه در ۵ خوشه را نشان می‌دهد. در روش آزمون خطا، بدون بهینه‌سازی، برای ثابت خطا عدد ۵۰۰ هزار به همراه بهترین مقادیر برای سایر پارامترها استفاده شده است. دلیل

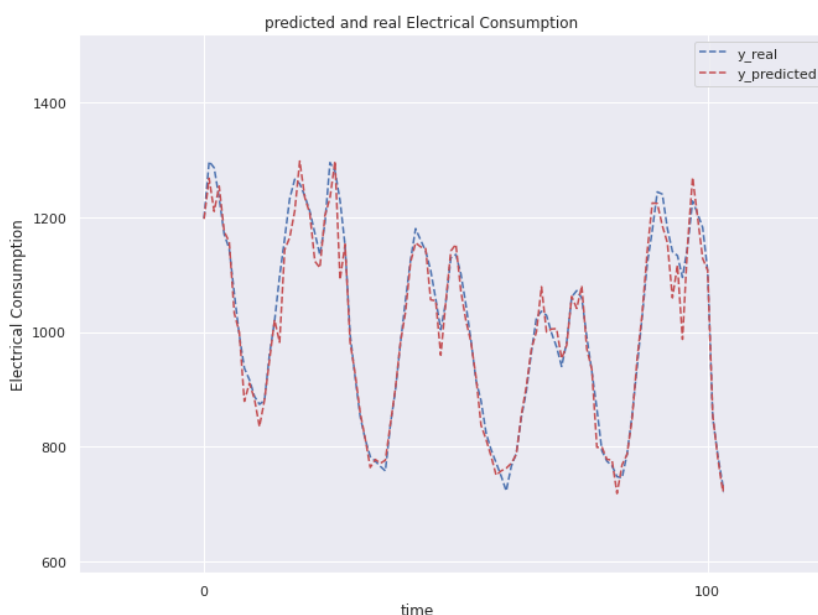
سیمانی و میرآبادی: مدل سازی پیش بینی کوتاه مدت مصرف برق با استفاده از روش های هوش مصنوعی در استان گیلان

جدول ۷. بهترین مقادیر فرارمترها در الگوریتم جست و جوی شبکه در ۵ خوشه

| خوشه          | $\epsilon$ | ضریب هسته $\gamma$ | تابع کرنل | ثابت خطا (C) |
|---------------|------------|--------------------|-----------|--------------|
| ۱             | ۳          | ۰,۰۱               | rbf       | ۱۵۰۰         |
| ۲             | ۵          | ۰,۰۱               | rbf       | ۱۰۰۰         |
| ۲ (آزمون خطا) | ۵          | ۰,۰۰۰۱             | rbf       | ۵۰۰۰۰        |
| ۳             | ۵          | ۰,۰۱               | rbf       | ۱۵۰۰         |
| ۴             | ۵          | ۰,۰۱               | rbf       | ۱۵۰۰         |
| ۴ (آزمون خطا) | ۵          | ۰,۰۰۰۱             | rbf       | ۵۰۰۰۰        |
| ۵             | ۰,۲        | ۰,۰۲               | rbf       | ۱۵۰۰         |

جدول ۸. مقادیر معیارهای اندازه گیری خطای خوشه ها در الگوریتم SVR

| شماره خوشه    | MSE      | RMSE   | MAE    | MAPE   | $R^2$  |
|---------------|----------|--------|--------|--------|--------|
| ۱             | ۲۶۸,۴۲۳  | ۱۶,۳۸۳ | ۱۱,۷۴۴ | ۰,۰۱۹۳ | ۰,۹۶۲  |
| ۲             | ۱۸۹۳,۲۶۲ | ۴۳,۵۱۱ | ۲۹,۳۵۳ | ۰,۰۲۶۴ | ۰,۹۸۶  |
| ۳             | ۳۵۳,۷۴۴  | ۱۸,۸۰۸ | ۱۳,۸۹۲ | ۰,۰۲۰۵ | ۰,۹۵۲  |
| ۴             | ۳۳۴۹,۸۳۹ | ۵۷,۸۷۷ | ۳۴,۳۵۱ | ۰,۰۳۵۸ | ۰,۹۶۳  |
| ۵             | ۶۳۴,۶۷   | ۲۵,۱۹۲ | ۱۶,۸۵۰ | ۰,۰۲۳۵ | ۰,۹۶۵  |
| مجموع خوشه ها | ۱۲۸۸,۵۶۷ | ۳۵,۸۹۶ | ۲۰,۸۷۴ | ۰,۰۲۱۴ | ۰,۰۹۸۶ |



شکل ۱۱. مقایسه مقادیر واقعی و پیش بینی شده در الگوریتم SVR

#### ۴. بحث و نتیجه گیری

استان گیلان به دلیل شرایط زیست محیطی و طبیعت بکر، منحنی تقاضای بار متفاوتی نسبت به یک استان گرم و خشک دارد. دلایلی از قبیل صنعتی نبودن و نداشتن محیط کسب و کار باعث تغییر روند زندگی می شود و همین

أمر مهاجرت به استان های مجاور را نیز افزایش می دهد. به دنبال این مسئله، ورود جمعیت زیادی در تعطیلات رسمی به خصوص عید نوروز، منحنی مصرف برق را تغییر خواهد داد. بنابراین، لازم است تمهیداتی اندیشیده شود تا خاموشی بخش مصرفی به ویژه در ساعات های اوج مصرف

به حداقل برسد. برای خریداران انرژی از بازار برق، نرخ فروش انرژی به تمام خریداران ثابت بوده و برابر با نرخ متوسط محاسبه شده بازار برق است. خریداران موظف هستند که در ازای دریافت انرژی، هزینه‌ای را بابت قدرت درخواستی، جریمه آزمون ناموفق مصرف و انرژی مصرفی بپردازند. هزینه در قبال قدرت درخواستی، مبلغی که خریدار به ازای هر مگاوات ساعت انرژی قابل عرضه می‌پردازد. این نرخ تحت تأثیر هزینه‌های تولید و انتقال و مدیریت شبکه برق است که در ساعت‌های مختلف شبانه‌روز و ماه‌های مختلف سال تغییر می‌کند. جریمه آزمون ناموفق عبارت است از: میزان خطای ناشی از پیش‌بینی بار در ساعت‌های کم‌باری، عادی و اوج بار که توسط خریدار پرداخت می‌شود و به دو دسته پیش‌بینی بیشتر و کمتر از مصرف واقعی تقسیم خواهد شد. حداکثر مقدار مجاز خطا برای ساعت‌های کم‌باری، عادی و اوج بار به ترتیب ۱۰، ۶ و ۲ درصد است. با توجه به شکل ۲، تقاضای برق در استان گیلان روند افزایشی خواهد داشت. این افزایش بر اثر عوامل مختلفی مانند افزایش جمعیت، افزایش بار صنعتی، افزایش دما و به دنبال آن، افزایش استفاده از وسایل سرمایشی پرمصرف، افزایش مشترکان برق، استفاده از انشعاب خانگی برای مصارف کشاورزی (برق عمومی برای روشنایی، موتور پمپ‌های چاه‌های آب و مراکز پرورش طیور و دام) و خاموش نکردن تجهیزات برقی بعد از استفاده است. مشکلات موجود بر سر راه افزایش ظرفیت تولید برق و از سوی دیگر، افزایش تقاضا، نیاز به مدیریت سمت تقاضا و بهینه‌سازی مصرف انرژی را پررنگ‌تر می‌کند. با استفاده از پیش‌بینی کوتاه‌مدت، می‌توان برنامه‌ریزی برای استفاده از نوع نیروگاه داشت. به این صورت تا حد ممکن از نیروگاه بخار برای تنظیم بار پایه و از نیروگاه برق‌آبی برای کنترل فرکانس و از نیروگاه گازی در زمان‌هایی که پیک مصرف بالا بوده و با مشکلات تولید مواجهیم، استفاده کرد. با استفاده از نتایج این پژوهش می‌توان روی توسعه هرچه بیشتر نیروگاه‌های برق‌آبی استان تمرکز کرد. افزایش پیک مصرف استان گیلان در تیر و مرداد و کاهش مصرف در دی و بهمن، نیاز به افزایش ذخیره پرخان را در استان افزایش می‌دهد. ذخیره چرخان، منبعی آماده برای تولید انرژی است که طی آن خاموشی را نیز در حد امکان کاهش خواهد داد. در این

پژوهش ابتدا با استفاده از روش‌های پیش‌پردازش، کیفیت داده‌ها جهت ورود به الگوریتم‌های یادگیری ماشین افزایش یافت. مرحله بعدی، داده‌ها با استفاده از الگوریتم خوشه‌بندی K-Means به پنج خوشه تقسیم شدند. تمامی الگوریتم‌ها برای کسب نتیجه بهتر، روی تمامی خوشه‌ها و سپس روی مجموع خوشه‌ها اعمال شد. با توجه به نتایج سه الگوریتم در جدول‌های ۴، ۶ و ۷، دقت نتایج بسیار زیاد است و می‌توان از نتایج این پژوهش در پیش‌بینی ساعتی استان گیلان استفاده کرد. مقادیر  $MAE$ ،  $MSE$ ،  $RMSE$  و  $R^2$  در الگوریتم رگرسیون خطی به ترتیب ۰/۲۵۱، ۰/۱۶۸۹/۲۹، ۴۱/۰۹۷، ۲۴/۴۷۳ و ۰/۹۸۲. در الگوریتم شبکه عصبی مصنوعی به ترتیب ۰/۰۲۵۹، ۱۶۷۳/۶۶۹، ۴۰/۹۱۰، ۲۵/۳۳۰ و ۰/۹۸۲. در الگوریتم رگرسیون بردار پشتیبان به ترتیب ۰/۰۲۱۴، ۱۲۸۸/۵۶۷، ۳۵/۸۹۶ و ۲۰/۸۷۴ و ۰/۹۸۶ است. از طرفی، هر الگوریتم دارای ۵ خوشه است که خطا در هر خوشه، در الگوریتم SVR کمتر از ANN و LR است. با توجه به اینکه مشکل اصلی در پیش‌بینی، پیش‌بینی ساعت‌های اوج مصرف است، الگوریتم SVR نسبت به دو الگوریتم ANN و LR بهتر عمل کرده است.

زمان اجرای الگوریتم‌های رگرسیون خطی، شبکه عصبی مصنوعی و رگرسیون بردار پشتیبان متفاوت است. در برخی از پژوهش‌ها، هدف، رسیدن به دقت زیاد پیش‌بینی به همراه کمترین زمان اجرای آموزش الگوریتم است. الگوریتم SVR+GS دارای بیشترین دقت و بیشترین زمان اجرا، الگوریتم ANN دقت سطح پایین‌تر نسبت به مدل SVR ولی با زمان اجرای کمتر نسبت به آن و الگوریتم LR دارای دقت کمتر و زمان اجرای بسیار کمتری است.

برای پژوهش‌های آتی، برای توسعه بخش نیروگاهی استان گیلان لازم است پیش‌بینی‌های بلندمدت نیز در پژوهش‌های آتی مورد توجه باشد. بنابراین، در نظر گرفتن شاخص‌های اقتصادی مانند جمعیت، تعداد مشترکان برق، شدت انرژی، یارانه، قیمت انرژی و GDP مورد ارزیابی قرار بگیرد. با تخمین میزان تقاضای کل مصرف برق استان، می‌توان برای توسعه زیرساخت‌های تولید برق در سال‌های آتی تمهیداتی اندیشید. همچنین، استفاده از ترکیب روش‌های مختلف شبکه‌های عصبی و میانگین آن‌ها برای

- [10]. S. R. Salkuti, "Short-term electrical load forecasting using radial basis function neural networks considering weather factors," *Electr. Eng.*, vol. 100, no. 3, pp. 1985–1995, 2018, doi: 10.1007/s00202-018-0678-8.
- [11]. S. N. Fallah, M. Ganjkhani, S. Shamshirband, and K. wing Chau, "Computational intelligence on short-term load forecasting: A methodological overview," *Energies*, vol. 12, no. 3, 2019, doi: 10.3390/en12030393.
- [12]. N. Elamin and M. Fukushige, "Modeling and forecasting hourly electricity demand by SARIMAX with interactions," *Energy*, vol. 165, pp. 257–268, 2018, doi: 10.1016/j.energy.2018.09.157.
- [13]. U. I. Akpan and A. Starkey, "Review of classification algorithms with changing inter-class distances," *Mach. Learn. with Appl.*, vol. 4, no. November 2020, p. 100031, 2021, doi: 10.1016/j.mlwa.2021.100031.
- [14]. H. Li et al., "Classification of electricity consumption behavior based on improved k-means and lstm," *Appl. Sci.*, vol. 11, no. 16, pp. 1–17, 2021, doi: 10.3390/app11167625.
- [15]. E. M. Burger and S. J. Moura, "Gated ensemble learning method for demand-side electricity load forecasting," *Energy Build.*, vol. 109, pp. 23–34, 2015, doi: 10.1016/j.enbuild.2015.10.019.
- [16]. Y. T. Chae, R. Horesh, Y. Hwang, and Y. M. Lee, "Artificial neural network model for forecasting sub-hourly electricity usage in commercial buildings," *Energy Build.*, vol. 111, pp. 184–194, 2016, doi: 10.1016/j.enbuild.2015.11.045.
- [17]. I. Shah, H. Iftikhar, and S. Ali, "Modeling and Forecasting Medium-Term Electricity Consumption Using Component Estimation Technique," *Forecasting*, vol. 2, no. 2, pp. 163–179, 2020, doi: 10.3390/forecast2020009.
- [18]. P. Pelka and G. Dudek, "Pattern-Based Forecasting Monthly Electricity Demand Using Multilayer Perceptron," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 11508 LNAI, pp. 663–672, 2019, doi: 10.1007/978-3-030-20912-4\_60.
- [19]. S. Yu, K. Wang, and Y. M. Wei, "A hybrid self-adaptive Particle Swarm Optimization-Genetic Algorithm-Radial Basis Function model for annual electricity demand prediction," *Energy Convers. Manag.*, vol. 91, pp. 176–185, 2015, doi: 10.1016/j.enconman.2014.11.059.

تخمین تقاضای بارالکتریکی، استفاده از الگوریتم‌های یادگیری ماشین و رسیدن به دقت زیاد و کاهش زمان اجرای برنامه باید مورد توجه باشد.

## منابع

- [1]. D. Lazos, A. B. Sproul, and M. Kay, "Optimisation of energy management in commercial buildings with weather forecasting inputs: A review," *Renew. Sustain. Energy Rev.*, vol. 39, pp. 587–603, 2014, doi: 10.1016/j.rser.2014.07.053.
- [2]. V. N. Sewdien, R. Preece, J. L. R. Torres, E. Rakhshani, and M. van der Meijden, "Assessment of critical parameters for artificial neural networks based short-term wind generation forecasting," *Renew. Energy*, vol. 161, pp. 878–892, 2020, doi: 10.1016/j.renene.2020.07.117.
- [3]. K. Chapagain, S. Kittipiyakul, and P. Kulthanavit, "Short-term electricity demand forecasting: Impact analysis of temperature for Thailand," *Energies*, vol. 13, no. 10, pp. 1–29, 2020, doi: 10.3390/en13102498.
- [4]. M. M. Rahman et al., "Prospective methodologies in hybrid renewable energy systems for energy prediction using artificial neural networks," *Sustain.*, vol. 13, no. 4, pp. 1–28, 2021, doi: 10.3390/su13042393.
- [5]. M. J. Gul, G. M. Urfa, A. Paul, J. Moon, S. Rho, and E. Hwang, "Mid-term electricity load prediction using CNN and Bi-LSTM," *J. Supercomput.*, vol. 77, no. 10, pp. 10942–10958, 2021, doi: 10.1007/s11227-021-03686-8.
- [6]. F. C. Torrini, R. C. Souza, F. L. Cyrino Oliveira, and J. F. Moreira Pessanha, "Long term electricity consumption forecast in Brazil: A fuzzy logic approach," *Socioecon. Plann. Sci.*, vol. 54, pp. 18–27, 2016, doi: 10.1016/j.seps.2015.12.002.
- [7]. S. Bissey, S. Jacques, and J. C. Le Bunetel, "The fuzzy logic method to efficiently optimize electricity consumption in individual housing," *Energies*, vol. 10, no. 11, 2017, doi: 10.3390/en10111701.
- [8]. E. A. Madrid and N. Antonio, "Short-term electricity load forecasting with machine learning," *Inf.*, vol. 12, no. 2, pp. 1–21, 2021, doi: 10.3390/info12020050.
- [9]. L. Suganthi and A. A. Samuel, "Energy models for demand forecasting - A review," *Renew. Sustain. Energy Rev.*, vol. 16, no. 2, pp. 1223–1240, 2012, doi: 10.1016/j.rser.2011.08.014.

- [20]. "Statistics and information network of Ministry of Energy." <https://isn.moe.gov.ir/> Accessed: 2023-02-0.
- [21]. "guilan regional electric company." <https://gilrec.co.ir/> Accessed: 2023-02-01.
- [22]. "National Meteorological Organization" <https://www.irimo.ir/> Accessed: 2023-02-01
- [23]. "Sunrise and sunset times in RashtGilan Province", Iran <https://sunrise-sunset.org/> Accessed: 2022-02-17.
- [24]. Maulud D, Abdulazeez AM. A Review on Linear Regression Comprehensive in Machine Learning. JASTT [Internet]. 2020Dec.31 [cited 2022Nov.3];1(4):140-7. Available from: <https://jastt.org/index.php/jasttpath/article/view/57>.
- [25]. J. Moon, S. Park, S. Rho, and E. Hwang, "A comparative analysis of artificial neural network architectures for building energy consumption forecasting," *Int. J. Distrib. Sens. Networks*, vol. 15, no. 9, 2019, doi: 10.1177/1550147719877616.
- [26]. R. K. Yadav and Anubhav, "PSO-GA based hybrid with Adam Optimization for ANN training with application in Medical Diagnosis," *Cogn. Syst. Res.*, vol. 64, pp. 191–199, 2020, doi: 10.1016/j.cogsys.2020.08.011.
- [27]. K. Mohammadi, S. Shamshirband, M. H. Anisi, K. Amjad Alam, and D. Petković, "Support vector regression based prediction of global solar radiation on a horizontal surface," *Energy Convers. Manag.*, vol. 91, pp. 433–441, 2015, doi: 10.1016/j.enconman.2014.12.015.
- [28]. D. Jap, M. Stöttinger, and S. Bhasin, "Support vector regression: Exploiting machine learning techniques for leakage modeling," *ACM Int. Conf. Proceeding Ser.*, vol. 14-June-20, 2015, doi: 10.1145/2768566.2768568.
- [29]. A. Román-Portabales, M. López-Nores, and J. J. Pazos-Arias, "Systematic review of electricity demand forecast using ann-based machine learning algorithms," *Sensors*, vol. 21, no. 13, pp. 1–23, 2021, doi: 10.3390/s21134544.