VILNIUS UNIVERSITY
FACULTY OF MATHEMATICS AND INFORMATICS
MODELLING AND DATA ANALYSIS
MASTER'S STUDY PROGRAMME

FORECAST OF ENERGY PRODUCED BY SOLAR POWER PLANTS

SAULĖS ELEKTRINIŲ PAGAMINTOS ENERGIJOS PROGNOZĖ

Master's thesis

Author: Vaiva Narkutė
VU email address: vaiva.narkute@mif.stud.vu.lt
Supervisor: dr. Jolita Bernatavičienė

Vilnius

2022

Abstract

In this paper, based on materials from international scientific articles, forecasts of energy produced by solar power plants are analyzed. The main goal of this paper is to analyze the theoretical and practical part of forecasting energy produced by solar power from lighting and weather prognosis. 3 data sets of solar power plants in different locations in Lithuania are used for analysis. General LSTM model is compared with statistical Linear Regression and, after it, Vanilla LSTM, Stacked LSTM, and Bidirectional LSTM are compared with each other for forecasting energy produced by solar power plants. The best-fitting parameters of Stacked LSTM are adapted to forecasting the energy production of each solar power plant. Furthermore, minimum and maximum forecasts of produced energy are prepared for having intervals between the minimum and maximum possible value of the average energy production forecast.

Keywords: Forecasting, Time Series, Solar Power Plants, Vanilla LSTM, Stacked LSTM, Bidirectional LSTM, deep learning, neural networks

Santrauka

Šiame darbe, paremtame tarptautiniais moksliniais straipsniais, yra analizuojamas saulės elektrinių pagaminamos energijos prognozavimas. Šio darbo pagrindinis tikslas yra teoriškai ir praktiškai išanalizuoti saulės elektrinių pagaminamos energijos prognozavimą, įtraukiant oro ir apšvitos prognozes. Darbe yra naudojami trijų saulės elektrinių Lietuvoje duomenys. LSTM modelis yra lyginimas su klasikiniu statistiniu prognozavimo metodu - tiesine regresija. Vėliau trys LSTM modelio variacijos - „Vanilla LSTM", „Stacked LSTM" ir „Bidirectional LSTM" yra palyginti tarpusavyje prognozuojant energiją, pagamintą saulės elektrinių. Analizėje pritaikyti geriausiai tinkantys parametrai skirti „Stacked LSTM" modeliui kiekvienai saulės elektrinei atskirai. Taip pat paruoštos minimalių ir maksimalių intervalo reikšmių saulės elektrinių pagamintos energijos prognozės.

Raktiniai žodžiai: Prognozavimas, Laiko eilutės, Saulės elektrinės, Vanilla LSTM, Stacked LSTM, Bidirectional LSTM, gilusis mokymasis, neuroniniai tinklai

Notation

- **LR** - Linear Regression,
- **LSTM** - Long Short Term Memory networks,
- **MAE** - Mean Absolute Error,
- **MSE** - Mean Squared Error,
- R^2 **score** - Coefficient of Determination,
- **RMSE** - Root Mean Squared Error.

Contents

1 Introduction

Energy, produced by solar power plants, is one of the renewable energy sources. According to the statistics of Eurostat, the share of renewable energy more than doubled between 2004 and 2019 - the share of renewables in gross final energy consumption stood at 19.7 % in the 27 countries of EU in 2019, compared with 9.6 % in 2004 that is shown in Figure 1 [8]. Lithuania shows even better results. The share of renewables in gross final energy consumption stood at 17.2 % in 2004 - almost 2 times better than the average of EU and the percentage of it increased was 26 % in 2017. Nevertheless, the growth of the share of energy from renewable sources in Lithuania is stopped since 2017. That could be easily explained - the EU share of energy from renewable sources target of 2020 was 23 % for Lithuania. This goal was reached by 2014. Thus the expansion of renewable energy sources was slowed down. EU countries have already agreed on a new renewable energy target of at least 27 % of final energy consumption in the EU as a whole by 2030. Targets of the national level are not published yet but that should encourage all EU countries to continue the expansion of renewable energy sources.

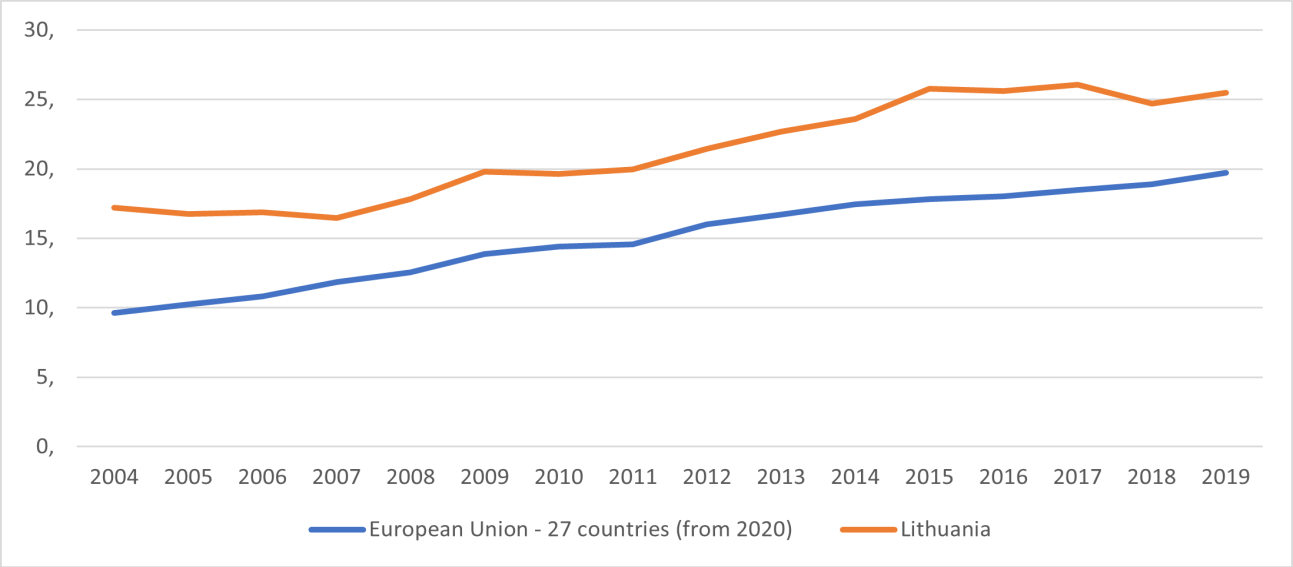

Figure 1: Share of energy from renewable sources, 2019 [8]

The efficient forecasting of time series plays a very important role in many fields. Renewable energy, or energy, produced by solar power plants, in particular, are only one of those fields. The main reason, why accurate forecasting is so important, is the ability to get information about the expected changes in the energy to be generated in the near future. It helps to facilitate participation in the energy market for providers and to have efficient resource planning. The necessity of efficient forecasting of solar energy is growing every year and is one of the main keys to persistence in the market.

Even if the topic of forecasting renewable energy is popular these days, we cannot find a lot of scientific works regarding it. At least, there are not as many works as we wish. In 2021, Google Scholar shows only 30 results with keywords "renewable energy forecasting". Naturally, the number of results decreases if the request is specified to "solar energy forecasting". Many related works analyze the forecasting problem by comparing different forecasting models. When the machine learning models have been proposed and applied for the time series forecasting, it was noticed that they can outperform tra-

ditional statistical models in most cases. Thus current researches are continued by comparing different machine learning models. In most cases, Long Short-term Memory (LSTM) networks outperform other machine learning models in renewable energy forecasting [14]. However, LSTM is always compared with other models and, to our knowledge, no prior studies have examined what effect the changes of variations of the LSTM model have on forecasting the energy produced by solar power plants including weather and lighting prognosis, therefore this is the main interest of this work.

Analysis of forecast of energy produced by solar power plants is performed in this work. After showing how simple univariate LSTM outperforms classical statistical models as Linear Regression (LR), different variations of the LSTM model are analyzed and compared in this work. The comparison includes 3 different variations of LSTM: Vanilla Long Short-Term Memory Model (Vanilla LSTM) [20], Stacked Long Short-Term Memory Networks (Stacked LSTM) [23], Bidirectional Long Short-Term Memory Model (Bidirectional LSTM) [29], and LSTM with other variations of parameters. Different variations of the LSTM model help customize a model to a solar energy forecasting model and reach the best possible results. Besides that, modification, which helps to avoid forecasting negative values, is made. This step is very important as there can never be negative values in solar energy production. Furthermore, many prior pieces of research analyses forecasting in daily level intervals or, if the forecast is prepared in hourly intervals, it is done for the short future period as 1-3 days [1, 6, 7, 12, 14]. We aim to prepare a longer 2-week forecast in hourly intervals, which could help for resource planning. The forecast is prepared for different periods of the year. That helps to identify if the variation of accuracy depends on the season for which forecast is prepared and, if necessary, make changes, which would lead to efficient forecast in any season. In addition, forecasts of minimum and maximum energy produced by solar power plants in hourly intervals are created and added to the average energy production forecast. To our knowledge, that is a new field of forecasting, which adds confidence intervals to the main energy forecast and gives more information for efficient resource planning. In this experiment, the forecasting performance of models is tested and compared using 3 data sets of real solar power plants in different locations of Lithuania. The exact locations of solar power plants will be unexposed regarding GDPR, and it will be called Solar Power Plant 1 (SPP1), Solar Power Plant 2 (SPP2), and Solar Power Plant 3 (SPP3) in further work.

The rest of the work is organized as follows. In Chapter 2, the literature review is presented, including the discussion of several various traditional statistical and machine learning models, their applications for the real world time series forecasting of renewable energy production with their strengths and weaknesses in making predictions. Additionally, the relevance of forecasting energy produced by solar power plants and how solar power plants work are reviewed. In Chapter 3 the main definitions and theoretical concepts, required for the experimental part of this work, are discussed, selected statistical (Linear Regression) and machine learning (Vanilla LSTM, Stacked LSTM, Bidirectional LSTM) models and forecasting performance metrics. In Chapter 4 simulation study and the real-world data application of the time series forecasting are explored using models described in Chapter 3. Final conclusions are given in Chapter 5.

2 Literature review

Solar power is the fastest-growing source: in 2008, it accounted for 1 % [8]. Growth in electricity from solar power has been dramatic, rising from just 7.4 TWh in 2008 to 125.7 TWh in 2019 [8]. As production and usage of solar power increases, the need for accurate forecasting for energy-producing raises. That helps to ensure a precise balance between electricity production and consumption at any moment for electricity operators by managing power system planning and operation. This paper more specifies on forecasting of energy produced by solar power plants, which are based on patterns of lighting and weather prognosis and considered as time series forecasting using artificial neural networks.

Accurate planning helps to produce enough electricity for clients on time. Wanting to achieve that, it is necessary to see and watch the operation of the system in real-time. Solar monitoring systems help to avoid possible failures and even major problems. The monitoring system is a software system that shares information about the performance, production values, consumption, and errors of the photovoltaic system. It not only offers information about energy consumption and generation but shows the ways of optimizing energy usage too. The installation of the monitor system is important for monitoring the condition of the solar system and for checking that the system is working properly. [9] The importance of monitoring was identified years ago with the first evaluations of the electrification programs. "Regular maintenance and monitoring" was identified by T. Urmee, D. Harries and A. Schlapfer [6] as one of the factors that contribute to rural electrification program success due to these actions increase the system lifetime and reduce the failure of operation of the solar power system and thus improve the confidence level of the user on the system. To this day, the monitoring of solar power systems continues to be identified as a crucial and necessary factor for achieving the success of electrification programs.

Solar monitoring systems operate through the solar system's inverter. Software that comes with inverters can monitor the process status. When solar inverter turns alternating current (AC) power to direct current (DC) power [7], it sends the data about system power level and energy production to a cloud-based tracking system. Then the information is sent to the software by this tracking system. Information can be accessed in several ways, including smartphones, smart home systems, computers. Aside from displaying energy consumption and generation data, monitoring systems offer many tools that help to understand solar energy setup. Using monitoring software problems and defects with panels can often be detected, and repairs to the setup can be recommended. Furthermore, the historical data of the system can be tracked. For instance, monitoring systems offer data on historical weather-based performance, so the impact of the weather on solar production in the past is visible. [10] This information can be used for future forecasting, which provides a way for grid operators to predict and balance energy generation and consumption.

Various forecasting techniques could be considered for solar power forecasting. The cloud imagery combined with physical models and the machine learning models is the two usual categories of forecasting. The choice of the forecasting method depends mainly on the prediction horizon and purpose. Different methods have different accuracy, which usually depends on the forecasting horizon used. The main classical techniques of forecasting solar radiation, such as naive methods, artificial neural network (ANN), autoregressive integrated moving average (ARIMA), support vector machine (SVM), regression

trees, random forest, were reviewed in the article of C. Voyant, G. Notton, S. Kalogirou, M. Nivet, C. Paoli, F. Motte and A. Fouilloy [1]. Main forecasting techniques using machine learning approaches that were used for prediction on the solar irradiance, which is essential for forecasting output power of solar systems, were reviewed in this article until the 2017 year. Also, main classical techniques, such as mean bias error (MBE), root mean square error (RMSE), mean absolute percentage error (MAPE), for evaluation of model accuracy were used. Many different forecasts are reviewed for comparison [1]. Review is combined from different exogenous and endogenous datasets of the worldwide locations. The horizon of forecasts differs from 1 minute to 1 day. Authors claim that the flexibility of ANN as a universal nonlinear approximation makes them more preferable than classical ARIMA. SVM and ANN give similar results but knowing that SVM is easier used than ANN, it could suggest using the SVM model if a choice is made between those 2 models. Generally, the quality of the training data is the main factor on which depends the accuracy of these methods. Regression trees and similar techniques are the methods, which are still not used wisely, but it looks promising to use them in the next years. If only forecasts of the longer period (1 day) are taken into consideration, the results of the best forecasting methods do not change.

Due to the non-linear dependence of the efficiency of solar power generation on meteorological variables such as irradiation and temperature [2], forecasting requires heavy preprocessing to refine input data. Thus, different researchers introduce the different hybrid forecasting methods which incorporate two or more techniques. Researches show that hybrid methods help to improve the accuracy of the forecast. In S. Nam's and J. Hur's [3] paper, a hybrid spatio-temporal forecasting model with Naïve Bayes Classifier method was applied. The authors forecasted the output of solar power a day ahead in South Korea using the weather data of the meteorological towers next to the solar farm in hourly intervals. This model let include the weather forecast in the prediction and showed improved results when compared to the deterministic persistence forecasting model. Normalized mean absolute error (nMAE) of the suggested model was about 4 % each month when nMAE of deterministic persistence forecasting model varied between 5-10 %.

Another interesting hybrid forecasting model - Machine learning and Statistical Hybrid Model (MLSHM), that combines the machine-learning methods with Theta statistical method [4]. The authors used a real-data of two renewable energy parks in different locations - Shagaya in Kuwait and Cocoa, in Florida, USA, on five-minute resolution data. To compare the accuracy of the new proposed MLSHM, other machine learning models - gate recurrent unit (GRU), auto-encoder gate recurrent unit (Auto-GRU), long short-term memory (LSTM), and auto-encoder long short-term memory (Auto-LSTM) - were used in the experiment [4]. This hybrid model approved one more time that the accuracy of combining the prediction of machine learning models with statistical methods is higher comparing with machine learning models without statistical methods. nMAE of MLSHM was about 4 % of both datasets when other models performed weaker - nMAE of LSTM and Auto-LSTM was 7 %, nMAE of GRU and Auto-GRU was 8%.

It is important to notice that forecast effectiveness relies on how many steps ahead accurate forecast can be created. The forecast period from 5 minutes ahead to 3 hours ahead was increased and effectiveness was tried in one of the recent researches in 2020 [13]. Diversity of accuracy of different forecasting steps was tested on 4 machine learning algorithms - neural networks (NN), support vector regression

(SVR), random forest (RF), and linear regression (LR). The dataset of the largest rooftop photovoltaic (PV) system in Australia was given for each algorithm. Overall, in this experiment, RF showed the best results on the last steps of prediction with MRE 15 % compared with other models. These insights could be adapted to longer period forecasts when intervals are 1 hour instead of 5 minutes.

Talking specifically about improving current researches, it is very important to include lighting and weather prognosis in the solar energy forecast. Forecasts of produced energy of solar power with machine learning methods outperform forecasts with only statistical methods too [5]. The researches and methods of solar power forecasts including lighting and weather prognosis are still on their early stage. Most of the authors of these researches claims that their methods requires further developments. On this day Google Scholar shows around 17 000 results of forecasting, which includes weather prognosis, and only 5 000 results of forecasting, which includes lighting prognosis. Thus researches of forecasts, which includes weather prognosis, will be reviewed first and after it review will be done for researches of forecasts including lighting prognosis.

In 2020 [11] a weather scenario generation-based probabilistic forecasting model were developed. Dataset of 7 solar farms with different capacity, which ranges from 1.05 MW to 230 MW, in Texas were used for experiment. The weather data in these locations were collected from the National Solar Radiation Database. Solar power forecasts were done for 1 hour ahead. Experimental analysis showed that this developed model outperformed other weather scenario generation-based machine learning-based multi-models with the methods, such as fixed-date (FD), shifted-date (SD), and bootstrap (BS) in all solar farms forecasts. nMAE was around 3 % and nRMSE - 4 % when other methods performed with nMAE around 7 % and nRMSE - 11 %. One more recent research, which included wind speed, irradiation, temperature, and precipitation to the energy generation of renewable energy sources forecast, was done in Portugal in the same year [12]. For this forecast, a new novel ensemble algorithm based on kernel density estimation (KDE) was proposed. The method of this algorithm is similar to the weather scenario generation-based machine learning-based multi-model, which was reviewed before. A method takes an input of an already prepared weather forecast and gives a forecast of energy generation as a result. Datasets of historical weather forecasts and solar PV installations located in the vicinity of the city of Coimbra, Portugal were used for testing the KDE method on 3 hours-interval bases. The forecast was prepared for 1 day ahead and compared with the ANN forecast using the same data. The KDE method outperformed an ANN in most cases by having 32 times greater computational time and normalized root mean square deviation (nRMSD) was 9.5 % compared with 10.5 %.

Long short term memory (LSTM) networks are one more forecasting method, which could usually be used for 1 day-ahead forecasting for ideal weather conditions. M. Gao, J. Li, F. Hong and D. Long decided to check if LSTM could be used for large-scale photovoltaic plants based on non-ideal weather conditions too [14]. The solar power data from a 10 MW peak solar power plant located in Shandong province, China, in 1-hour intervals were used for the experiment. Results show that LSTM outperforms other methods as Back Propagation (BP) neural networks, Least Square Support Vector Machine (LSSVM), and Wavelet Neural Network (WNN) because nRMSE of LSTM was 5 % when other models performed with nRMSE around 17 %. Thus it can be considered to use the LSTM method even for non-ideal weather conditions. Another recent work was done by K. Wang, X. Qi, and H. Liu [33], where authors compared LSTM and CNN models and their hybrid models for solar energy forecasting.

The 1B DKASC, Alice Springs PV system data, and weather prognosis were used for experiments. The forecast was done for 5 days in 5 minutes intervals. What is interesting in this research, hybrid LSTM-CNN and CNN-LSTM models outperformed the simple LSTM model but CNN did not outperform it. The results of RMSE of LSTM were 0.709 compared with RMSE 0.822 of CNN, 0.693 of CNN-LSTM, and 0.621 of LSTM-CNN. However, it is found from the running time of the model that the running time of the hybrid model is higher than the running time of the single LSTM model. Furthermore, X. Qind and Y. Niu [34] used LSTM for solar irradiance prediction. The forecast was done for one day in hourly periods and compared with other forecasting methods as the linear regression (LR) and classical backpropagation algorithm (BPNN). Results were unquestionable - RMSE of LSTM was only 76.245 when the result of RMSE of LR was 230.9867 and the result of RMSE of BPNN was 133.5313. Once again LSTM outperformed other models. The comparison of mentioned previous experiments is shown in the Table 1. Thus it is decided to use different variations of LSTM in this work and find the best parameters of the model for forecasting energy produced by solar power plants in Lithuania.

References	Year	Horizon	Evaluation criteria	Results
[13]	2020	3 hours	MRE = 15% > MRE = 17-18%	RF is better than NN, LR, SVR.
[1]	2017	1 day	nMAE = 3% > MAPE = 4% > nRMSE = 9.4%	Hybrid of ANN-SVM gives the best results. ANN is slightly better than SVM.
[6]	2018	1 day	nMAE = 4 % > nMAE = 8%	Hybrid spatio-temporal model is better than deterministic persistence forecasting model.
[7]	2019	1 day	nMAE = 4% > nMAE = 7% > nMAE = 8%	MLSHM is better than LSTM, Auto-LSTM, GRU, Auto-GRU.
[11]	2020	1 h	nMAE = 3% > nMAE = 7%	Weather scenario generation-based probabilistic forecasting is better than fixed-date, shifted-date and bootstrap forecasting methods.
[12]	2020	1 day	nRMSD = 10% > nRMSD = 11%	KDE is better than ANN.
[14]	2019	1 day	nRMSE = 5% > nRMSE = 17%	LSTM is better than BP, LSSVM, WNN.
[33]	2019	5 min	RMSE = 0.621 > RMSE = 0.822	LSTM-CNN is better than CNN.
[34]	2018	1 h	RMSE = 76.25 > RMSE = 230.98	LSTM is better than LR, BPNN.

Table 1: The overview of forecasting methods. ">" in Evaluation criteria column shows which model between models in Results column performs better.

3 Methodology

3.1 Long-short time memory (LSTM) model

The LSTM setup most commonly used in the literature was originally described by Graves and Schmidhuber [19]. Inspired by classical recurrent neural networks (RNN), Long Short-Term Memory networks receive data samples sequentially and use the last prediction to predict the next data sample [18]. Classical RNNs have a feedback loop which brings back the latest outputs of the network in the input. This structure design leads to various problems such as exploding or vanishing gradient during the training of the network. To solve these problems, LSTM networks share an additional parameter, cell state, between sequences, which gives them the ability to remember/forget important/irrelevant features of data in any part of the sequence.

LSTM network can have different input and output structure layers. For instance, LSTM can receive all the input samples and return one output at the end of receiving all input samples, or it can provide an output for each input sample. In this work, as shown in Figure 2, the network has one output for each input value in the sequence.

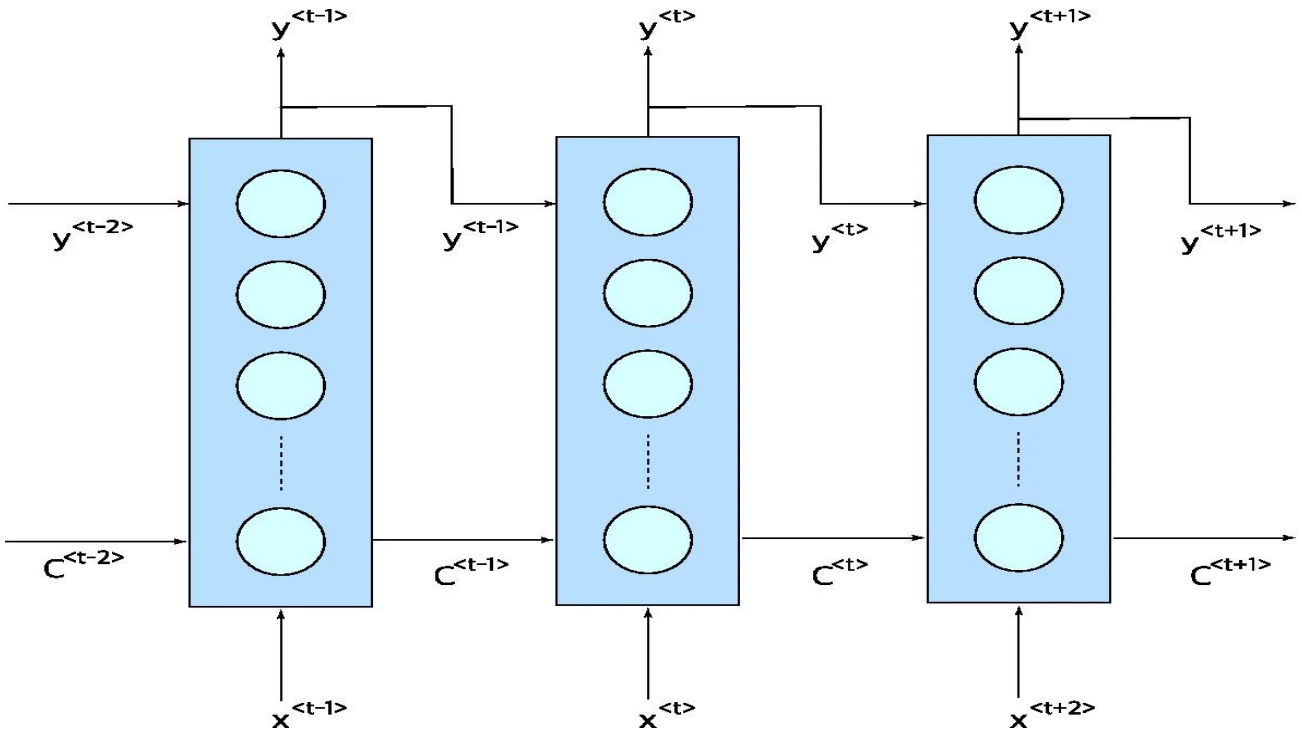

Figure 2: Unfolded LSTM network. Three time steps of LSTM are unfolded. In each step, LSTM receives a new input, last output and last carry and generates the next output and the next carry. [18]

The formulation for the LSTM network can be seen in Equations 1-6 [18].

$$\tilde{C}^{(t)} = \tanh(W_c[y^{(t-1)}, x^{(t)}]), \tag{1}$$

$$\Gamma_u = sigmoid(W_u[y^{(t-1)}, x^{(t)}]), \tag{2}$$

$$\Gamma_f = sigmoid(W_f[y^{(t-1)}, x^{(t)}]), \tag{3}$$

$$\Gamma_o = sigmoid(W_o[y^{(t-1)}, x^{(t)}]), \tag{4}$$

$$C^{(t)} = \Gamma_u \odot \tilde{C}^{(t)} + \Gamma_f \odot C^{(t-1)}, \tag{5}$$

$$y^{(t)} = \Gamma_o \odot C^{(t)}. \tag{6}$$

x and y are one value of a sequence of input and output data samples. W_c, W_u, W_f and W_o are carry, update, forgetting, and output weights, respectively, which are going to be learned during the training process and $\odot$ represents the element-wise product. The LSTM algorithm can be summarized in the following steps. First, using the current input sample and the previous output sample, a potential carry value, i.e., $^{(t)}\rangle$, is calculated using (1). Then again using the last output and the current input, the value of update, forget, and output gates are determined according to (2,3,4). These gates can have values between 0 and 1. For instance, in the extreme case where $\Gamma_u = 1$ and $\Gamma_f = 0$, the network completely forgets the previous values and updates the carry with the new carry potential value according to (5). Then, using the update and forget gates, the final carry value for the current step is calculated (6). Finally, the estimated output is calculated by the dot product of the current carry value and the output gate. In the original model, the bias values are considered in the Equations 1, 2, 3, and 4. Moreover, the output activation of Among different variants of LSTM structures is used. In this paper, three different LSTM structures are used for work comparison - Vanilla LSTM, Stacked LSTM and Bidirectional LSTM. The summary of the LSTM structure is shown in Figure 3.

3.2 Vanilla LSTM

The Vanilla LSTM includes changes from Gers [20] and Gers and Schmidhuber [21] to the original LSTM and uses full gradient training. The first paper [20] which proposed a modification to the LSTM architecture introduced the forget gate, enabling the LSTM to reset its own state. This modification enables learning of continuous tasks. Gers and Schmidhuber [21] argued that the cell must control the gates to learn precise timing. Until then, this was only possible through an open output gate. Peephole connections [connections from the cell to the gates Figure 4 (blue curves)] were added to the architecture to facilitate learning precise timing. In addition, the output activation function was omitted, as there was no evidence that it was essential for solving the problems on which the LSTM has been tested on so far. The final modification toward the Vanilla LSTM is BPTT training for LSTM networks using the architecture described in this section. The use of the full BPTT had the added benefit of allowing the LSTM gradients to be checked using finite differences, which made the practical implementations more reliable.

A schematic representation of the vanilla LSTM block can be seen in Figure 4. It has three gates (input, forget, and output), a block input, a single cell (the constant error carousel), an output activation function, and peephole connections [22]. The output of the block is recurrently connected back to the

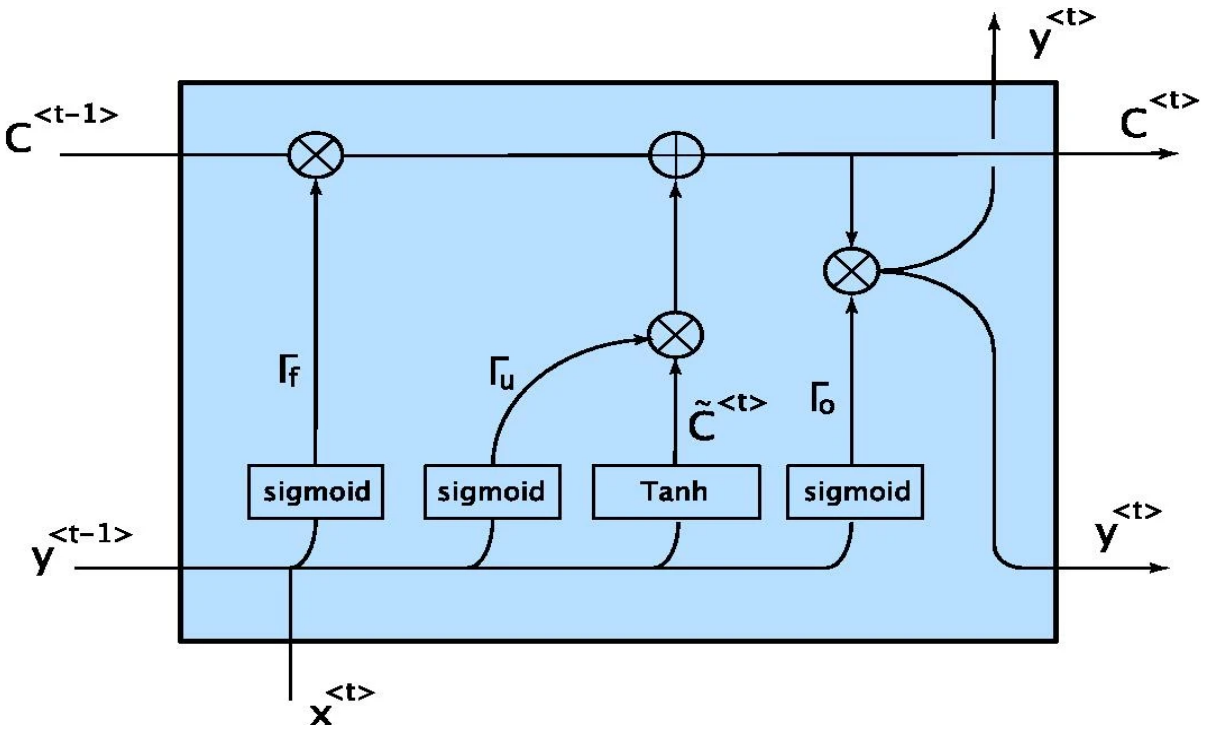

Figure 3: LSTM structure. Inputs, outputs and LSTM gates and their connections are illustrated. [18]

block input and all gates.

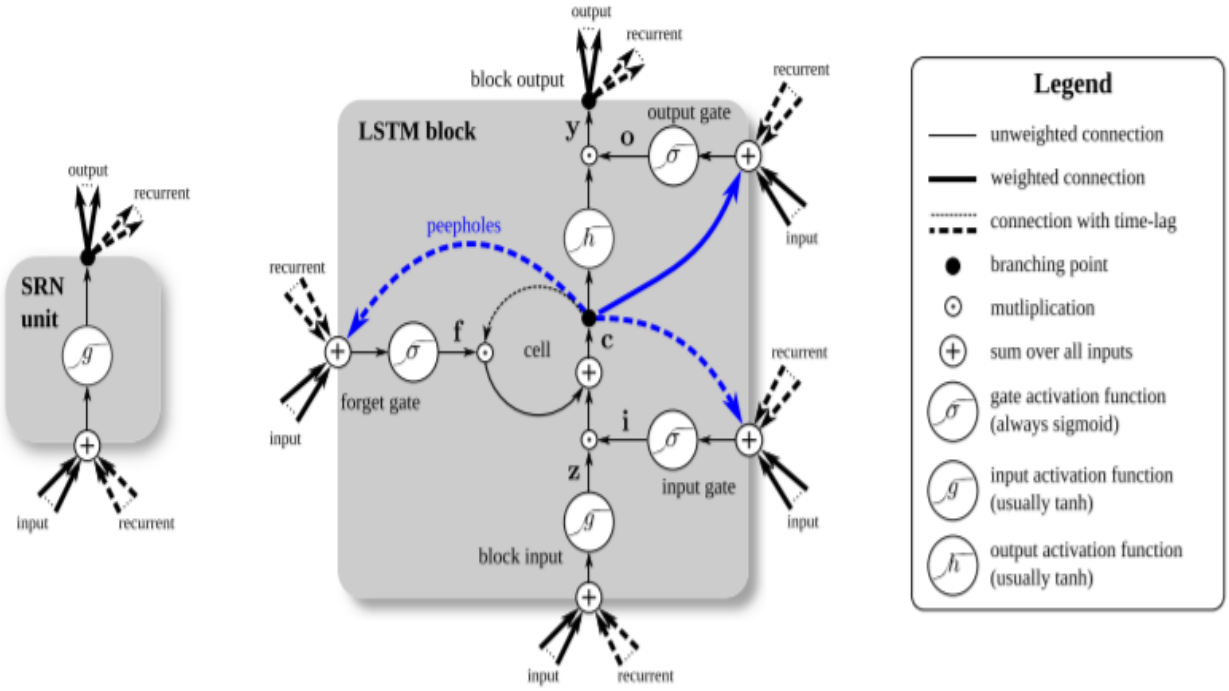

Figure 4: Vanilla LSTM structure. Inputs, outputs and LSTM gates and their connections are illustrated. [22]

3.2.1 Forward pass

Let x^t be the input vector at time t, N be the number of LSTM blocks, and M the number of inputs. Then, we get the following weights for an LSTM layer.

1) Input Weights: $W_z, W_s, W_f, W_o \in R^{N \times M}$.

2) Recurrent Weights: $R_z, R_s, R_f, R_o \in R^{N \times N}$.

3) Peephole Weights: $p_s, p_f, p_o \in R^N$.

4) Bias Weights: $b_z, b_s, b_f, b_o \in R^N$.

Then the vector formulas for a vanilla LSTM layer forward pass can be written as

$$\bar{z}^t = W_z x^t + R_z y^{t-1} + b_z, \tag{7}$$

$$z^t = g(\bar{z}^t) \text{ - block input}, \tag{8}$$

$$\bar{i}^t = W_i x^t + R_i y^{t-1} + p_i \odot c^{t-1} + b_i, \tag{9}$$

$$i^t = \sigma(\bar{i}^t) \text{ - input gate}, \tag{10}$$

$$\bar{f}^t = W_f x^t + R_f y^{t-1} + p_f \odot c^{t-1} + b_f, \tag{11}$$

$$f^t = \sigma(\bar{f}^t) \text{ - forget gate}, \tag{12}$$

$$c^t = z^t \odot i^t + c^{t-1} \odot f^t \text{ - cell}, \tag{13}$$

$$\bar{o}^t = W_o x^t + R_o y^{t-1} + p_o \odot c^t + b_o, \tag{14}$$

$$o^t = \sigma(\bar{o}^t) \text{ - output gate}, \tag{15}$$

$$y^t = h(c^t) \odot o^t \text{ - block output}. \tag{16}$$

where σ, g, and h are pointwise nonlinear activation functions. The logistic sigmoid $\sigma(x) = 1/(1 + e^{-x})$ is used as the gate activation function and the hyperbolic tangent $g(x) = h(x) = \tanh(x)$ is usually used as the block input and output activation function. Pointwise multiplication of two vectors is denoted by $\odot$.

3.2.2 Backpropagation Through Time

The deltas inside the LSTM block are then calculated as

$$\delta y^t = \Delta^t + R_z^t \delta z^{t+1} + R_i^t \delta i^{t+1} + R_f^t \delta f^{t+1} + R_o^t \delta o^{t+1}, \tag{17}$$

$$\delta o^t = \delta y^t \odot h(c^t) \odot \sigma'(\bar{o}^t), \tag{18}$$

$$\delta c^t = \delta y^t \odot o^t \odot h'(c^t) + p_o \odot \delta o^t + p_i \odot \delta i^{t+1} + p_f \odot \delta f^{t+1} + \delta c^{t+1} \odot f^{t+1}, \tag{19}$$

$$\delta f^t = \delta c^t \odot c^{t-1} \odot \sigma'(\bar{f}^t), \tag{20}$$

$$\delta i^t = \delta c^t \odot z^t \odot \sigma'(\bar{i}^t), \tag{21}$$

$$\delta z^t = \delta c^t \odot i^t \odot \sigma'(\bar{z}^t). \tag{22}$$

Here, Δ_t is the vector of the deltas passed down from the layer above. If E is the loss function, it formally corresponds $\delta E/\delta y^t$, but without including the recurrent dependencies. The deltas for the inputs are only needed if there is an underlying layer that needs to be trained, and can be computed as follows:

$$\delta x^t = W_z^t \delta z^t + W_i^t \delta i^t + W_f^t \delta f^t + W_o^t \delta o^t. \tag{23}$$

Finally, the gradients for the weights are calculated as follows, where $\star$ can be any of $\{z, i, f, o\}$, and $\{\star 1, \star 2\}$ denotes the outer product of two vectors:

$$\delta W_\star = \sum_{t=0}^{T} (\delta \star^t, x^t) \text{ and } \delta p_i = \sum_{t=0}^{T-1} c^t \odot \delta i^{t+1}, \tag{24}$$

$$\delta R_\star = \sum_{t=0}^{T-1} (\delta \star^{t+1}, y^t) \text{ and } \delta p_f = \sum_{t=0}^{T-1} c^t \odot \delta f^{t+1}, \tag{25}$$

$$\delta b_\star = \sum_{t=0}^{T} \delta \star^t \text{ and } \delta p_o = \sum_{t=0}^{T} c^t \odot \delta o^t. \tag{26}$$

3.3 Stacked LSTM

The stacked LSTM, also known as deep LSTM, was first formulated to address speech recognition problems by Graves [23]. The original LSTM model consists a single hidden LSTM layer followed by a standard feedforward output layer. The stacked LSTM model is an extension of that model and uses multiple LSTM layers that are stacked before being forwarded to a dropout layer and output layer at the final output. In a stacked LSTM, the first LSTM layer generates sequence vectors used as the input of the subsequent LSTM layer. In addition, the LSTM layer receives feedback from its previous time step, allowing it to capture data patterns. Moreover, the added dropout layer also excludes 10% of

neurons to avoid overfitting.

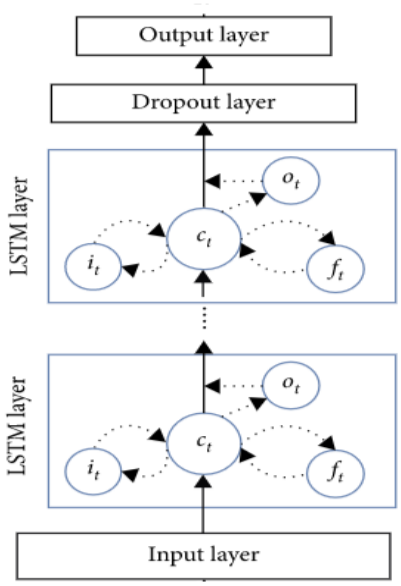

Figure 5: Stacked LSTM structure. Inputs, outputs and LSTM gates and their connections are illustrated. [24]

The basic structure of the LSTM, as shown in the LSTM layer in Figure [24], consists of an input gate i_t, an output gate o_t, a forget gate f_t, and a memory cell c_t as well as Vanilla LSTM. A single LSTM layer has a second-order RNN architecture that excels at storing sequential short-term memories and retrieving them at many time steps later [25]. An LSTM network is identical to a standard RNN, except the summation units in the hidden layer are replaced by memory blocks [26]. Equations 27-31 describe how the output values are updated at each step [27, 28, 20].

$$f_t = \sigma(W_f \cdot x_t + U_f \cdot h_{t-1} + b_f), \tag{27}$$

$$i_t = \sigma(W_i \cdot x_t + U_i \cdot h_{t-1} + b_i), \tag{28}$$

$$o_t = \sigma(W_o \cdot x_t + U_o \cdot h_{t-1} + b_o), \tag{29}$$

$$c_t = f_t \cdot c_{t-1} + i_t \cdot \sigma(W_c \cdot x_t + U_c \cdot h_{t-1} + b_c), \tag{30}$$

$$h_t = o_t + \sigma(c_t). \tag{31}$$

where x_t is the input vector; σ is the activation function; $W_f, W_i, W_o, W_c, U_f, U_i, U_o$, and U_c are the weight vector terms; b_f, b_i, b_o, and b_c are the corresponding bias terms; and h_t and h_{-1} are the current and previous hidden vectors, respectively.

3.4 Bidirectional LSTM

Since LSTM can only learn the above information of time series, Bidirectional LSTM makes is another improvement based on LSTM, which consists of a forward LSTM network and a backward LSTM network and incorporates the context information of time series. To overcome the shortcoming of a single LSTM cell that can only capture the previous context but cannot use the future context, Schuster and Paliwal [30] invented bidirectional recurrent neural networks (BRNN) that combine two separate hidden LSTM layers with opposite directions to the same output. The structure of the bidirectional LSTM neural network is shown in Figure 6 [29].

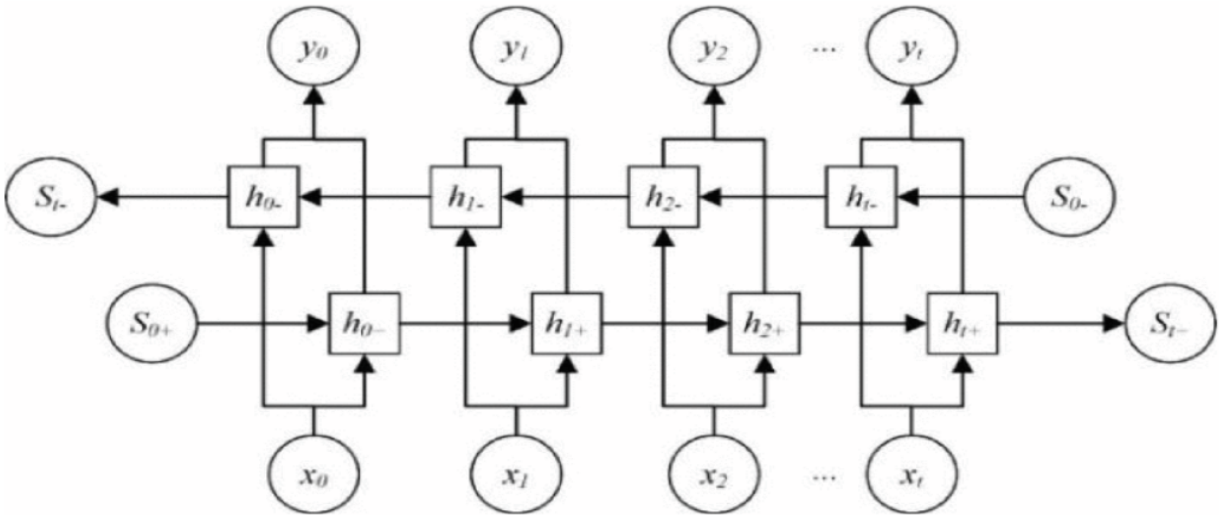

Figure 6: Bidirectional LSTM structure. Inputs, outputs and LSTM gates and their connections are illustrated. [29]

The calculation formulas are as follows:

$$\sigma i_t = (W_i \cdot [h_{t-1}, x_t] + b_i), \tag{32}$$

$$\sigma f_t = (W_f \cdot [h_{t-1}, x_t] + b_f), \tag{33}$$

$$\sigma o_t = (W_o \cdot [h_{t-1}, x_t] + b_o), \tag{34}$$

$$h_t = o_t \cdot \tanh(c_t), \tag{35}$$

$$c_t = f_t \cdot c_{t-1} + i_t \tilde{c}_t, \tag{36}$$

$$\tilde{c}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c), \tag{37}$$

$$\sigma(x) = \frac{1}{1+e^{-x}}, \tag{38}$$

$$\tanh(x) = \frac{e^{x}-e^{-x}}{e^{x}+e^{-x}}. \tag{39}$$

where, as in original LSTM, W_i, W_f and W_o respectively represent the weight of three gate connections, b is the offset, σ and tanh are the activation functions.

In Bidirectional LSTM $x_1, x_2, \ldots, x_t$ are the input sequence, $\overrightarrow{h}_t$ and $\overleftarrow{h}_t$ are the forward and backward outputs computed at each time point, and then the forward and reverse outputs are calculated to obtain the final output y_t. Taking the forward output $\overrightarrow{h}_t$ at time t as an example, the forward and backward direction calculation formulas are consistent with the LSTM, that is, through equation (32) to equation (39), the forward and reverse temporary cell states $\overrightarrow{\tilde{c}}_t$ and $\overleftarrow{\tilde{c}}$, input gate $\overrightarrow{l}_t$ and $\overleftarrow{l}_t$, forget gate $\overrightarrow{f}_t$ and $\overleftarrow{f}_t$, output gate $\overrightarrow{o}_t$ and $\overleftarrow{o}_t$ can be calculated respectively.

The final output y_t at time t is:

$$y_t = [\overrightarrow{h}_t, \overleftarrow{h}_t]. \tag{40}$$

From the above equation, the output at each moment can be calculated, and then get the final output $Y = [h_0, h_1, \ldots, h_t]$.

3.5 Alternative method - Linear Regression

In addition, classical statistical forecasting method Linear Regression (LR) is used for comparison with LSTM models. Linear regressions are routinely used in chemometrics, econometrics, financial engineering, psychometrics and many other areas of applications to model the predictive relationships of multiple related responses on a set of predictors [32]. In general multivariate linear regression, we have n observations on q responses $y = (y_1, \ldots, y_q)$ and p explanatory variables $x = (x_1, \ldots, x_p)$, and [31]

$$Y = XB + E. \tag{41}$$

where $Y = (y_1, \ldots, y_n)$ is an $n \times q$ matrix, $X = (x_1, \ldots, x_n)$ is an $n \times p$ matrix, B is a $p \times q$ coefficient matrix, $E = (e_1, \ldots, e_n)$ is the regression noise and the e-s are independently sampled from $N(0, \Sigma)$.

3.6 Forecasting performance metrics

The main idea of the time series forecasting is to make predictions for the future observations by using the certain model. For the purpose of assessing the quality of forecasting models the most important step is to evaluate the accuracy of predictions [15]. Typically, different forecasting performance metrics are used to evaluate how close the predicted values are to the actual values of time series. The literature review shows that the most commonly used are these absolute and squared forecasting error measures: Mean Absolute Error (MAE), Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) which are calculated using the following formulas respectively [15, 16]:

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|x_t - \widehat{x}_t|, \tag{42}$$

$$MSE = \frac{1}{n}\sum_{t=1}^{n}(x_t - \widehat{x}_t)^2, \tag{43}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(x_t - \widehat{x}_t)^2} = \sqrt{MSE}. \tag{44}$$

Here x_t is the measured value of the time series at the time t, $\widehat{x}_t$ is the predicted value of x_t and n denotes the number of observations/predictions.

Also correlation metric is very useful for having an evaluation of accuracy between predicted and actual values in ratio, not only in units. R^2 score is one of them and it is used to evaluate the work of models in this work. In other words, R^2 score is the coefficient of determination, which is an improvement of the coefficient of correlation. The coefficient of correlation shows the overall resemblance between the observed and predicted values. [17] On the other hand, the coefficient of determination can show how much of the variance in the observed data exist in the predicted values. [18] The formulation of (r) and (R^2) are represented in the formulas below:

$$r = \frac{\sum_{t=1}^{n}(y_t - \tilde{y})(\widehat{y}_t - \tilde{\widehat{y}})}{\sqrt{\sum_{t=1}^{n}(y_t - \tilde{y})^2}\sqrt{\sum_{t=1}^{n}(\widehat{y}_t - \tilde{\widehat{y}})^2}}, \tag{45}$$

$$R^2 = 1 - \frac{\sum_{t=1}^{n}(y_t - \widehat{y}_t)^2}{\sum_{t=1}^{n}(y_t - \tilde{y})^2}. \tag{46}$$

In this formulation, y_t is the t-th sample of the target value and $\tilde{y}$ is the mean of the target value. In the same manner, $\widehat{y}_i$ is the t-th sample of predicted value and $\tilde{\widehat{y}}$ is the mean value of the predicted value.

In total, 4 metrics of errors are used for evaluating the accuracy of forecasting in this work. They are MAE, MSE, RMSE and R^2 score.

4 Experimental results

In the experiment, different variations of LSTM models are used for energy produced by solar power plants forecasting. Literature review shows that LSTM should suit the forecast of renewable energy the best. Thus it will be analyzed in experimental analysis. The main goal is to find the best modification of LSTM with the most accurate prediction of energy produced by solar power plants. To perform the analysis, historical data of exported energy by 3 solar power plants and data of the weather and lighting prognosis in different locations are used. The period of all 3 time series is the same - one and a half years from 2019-09 to 2021-03 in hourly intervals. The long time series will allow to experiment with the lengths of time series and ensure that lack of data will not affect the results of the analysis. The main goal is to prepare an hourly forecast of energy produced by solar power plants with the best accuracy. Most papers analyze the accuracy of the daily forecast for renewable energy but hourly forecast can ensure more accurate resource planning. Deeper analysis in this field is urgent.

This paper will be organized as follows. At first, linear regression will be compared with basic univariate LSTM to show how forecast accuracy varies between simple statistic models and neural networks. After this comparison, historical weather and lighting data and prognosis will be included in the forecast and only multivariate LSTM will be used for further experiments. The experiments with different lengths and model parameters of time series will be done using the basic LSTM model at first. After finding the best parameters, modifications of the LSTM model will be done to get better accuracy. 3 different time series should help to identify if the LSTM model works the same with any data of energy produced by solar power plants or some parameters should be specified for better prediction with different time series. The results will be evaluated using the four different metrics of errors - MSE, RMSE, MAE, and R2 regression score function.

Experimental analysis is performed using the Python program.

4.1 Data visualization and preparation

To analyze if different variations of the LSTM model fits energy produced by solar power plants data, real data sets of 3 solar power plants in different locations in Lithuania are chosen - SPP1, SPP2, and SPP3 (later in this paper it will be called SPP3). The period of the hourly data is from September 2019 to March 2021. The data sets contain more than 13000 rows. Data sets of all solar power plants contain the same metrics as a minimum, maximum and average exported energy, irradiation, and sensor cell temperature. Furthermore, weather and lighting historical data of the locations of these solar power plants was extracted. Metrics of daily sunny hours, hourly temperature, wind speed, cloud cover, and heat index are going to be used for prognosis. It is important to notice that this data was never used for LSTM analysis before. Also, it is not necessary to duplicate the same metrics, thus historical data of irradiation and sensor cell temperature of solar power plants will not be used. The main historical time series for the forecast of energy produced by solar power plants are average exported energy. Furthermore, minimum and maximum exported energy will be forecasted separately for having confidence intervals of average production of energy. 3 different time series help to analyze the accuracy of forecasting with different variations of LSTM.

The visualization of datasets is shown in Figure 7. Looking at it, it could be guessed that strong

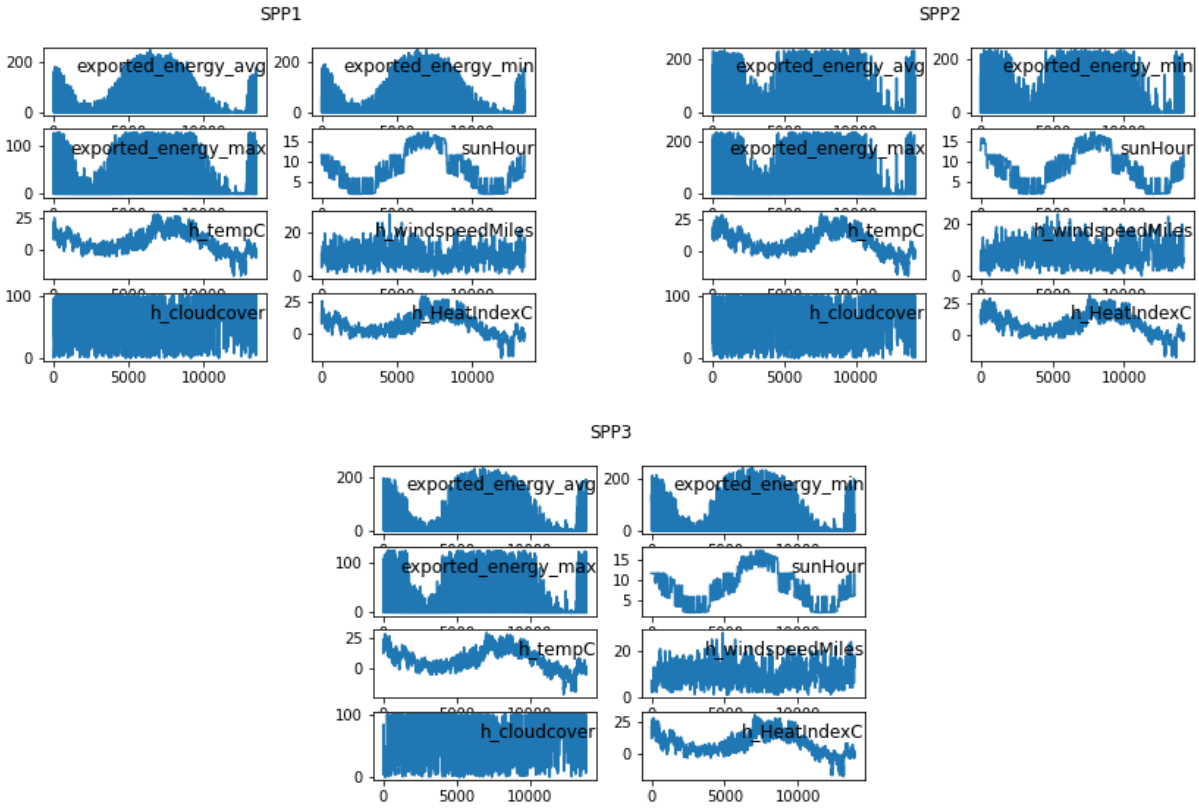

Figure 7: Time series of 3 solar power plants.

seasonality exists in all time series. This is logical knowing that Lithuania has 4 seasons and fewer sunny hours in the winter compared with other seasons. This also is visible from the sunny hours (sunHour) time series in the plots. This information is very useful for further analysis. It is important to check if LSTM models make a forecast with the same accuracy for different times of the year. Another important thing - a pattern of SPP2 exported energy is slightly different from SPP1 or SPP3. It seems that there are higher temperatures and more sunny hours in that location compared to others. Wind speed, cloud cover, and heat index could have an impact too. It is important information for further analysis.

After data visualization, correlation analysis is performed. It indicated a perfect positive correlation between hourly temperature and heat index in all time series. Thus heat index will not be used for further analysis and prognosis.

4.2 Linear regression (LR) and Long Short Term Memory networks (LSTM) comparison

Many scientific articles claim that using neural networks can create more accurate forecasts than using classical statistical methods. That is why neural networks are so popular for forecasting time series last years. To get a better view, a comparison between classical linear regression and long short term memory networks is made.

Data of all 3 different solar power plants is validated and divided into training and testing sets.

80% of values are chosen for training the models and the rest of the data is left for testing. It applies to both models.

For the linear regression forecast, all left weather metrics (sunny hours, temperature, wind speed, and cloud cover) are used for the produced average energy forecasting. Then for the same goal, we use the univariate LSTM model. Basic parameters are chosen for LSTM networks training that comparison between LR and LSTM would be more equal. It is 4 hidden layers, 1 dense layer, and the input shape of the real data. LSTM model is trained for 25 epochs with a batch size of 64. All parameters are the same for 3 different time series. The results of both LR and LSTM models are evaluated using the R^2 score parameter and they are shown in Table 2 below.

	LR	**LSTM**
SPP1	0.253856	0.873923
SPP2	0.112729	0.722089
SPP3	0.212743	0.857919

Table 2: Evaluation of the forecasting performance of Linear Regression and Long Short Term Memory networks models on the test data set using the R^2 score.

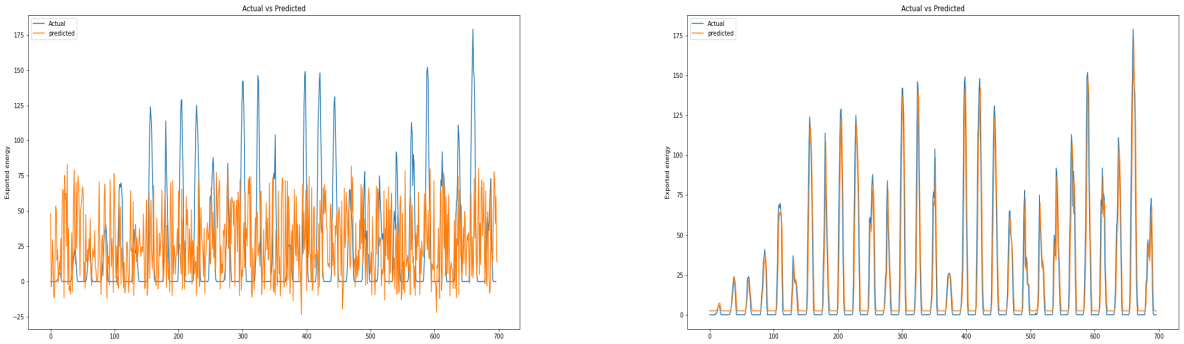

Figure 8: Actual and predicted testing data set of SPP1 - LR used on the left and LSTM used on the right.

It is obvious from the Table 2 that the LR forecast is so much weaker compared with LSTM. Furthermore, the Figures 8, 9 and 10 show that linear regression does not catch a forecasting pattern at all. LR forecast looks more like random walk when forecast, prepared using simplest LSTM, already catches the pattern. Surely, it is visible that LSTM fails with highest and lowest values forecasting, especially with the SPP2 time series. However, that does not change the fact that forecast accuracy using LSTM is already more than 3 times better than forecast using LR and forecast of highest and lowest values will be analyzed in further steps using deeper LSTM application.

To sum everything up, forecasting with long short time memory networks leads to at least more than 3 times better results than forecasting with linear regression. That confirms current researches that using even simple neural networks can have better forecasting results than using classical statistical models. However, the accuracy of univariate LSTM is still not perfect and further research is required. Therefore the experimental analysis is continued using different variations of the LSTM model.

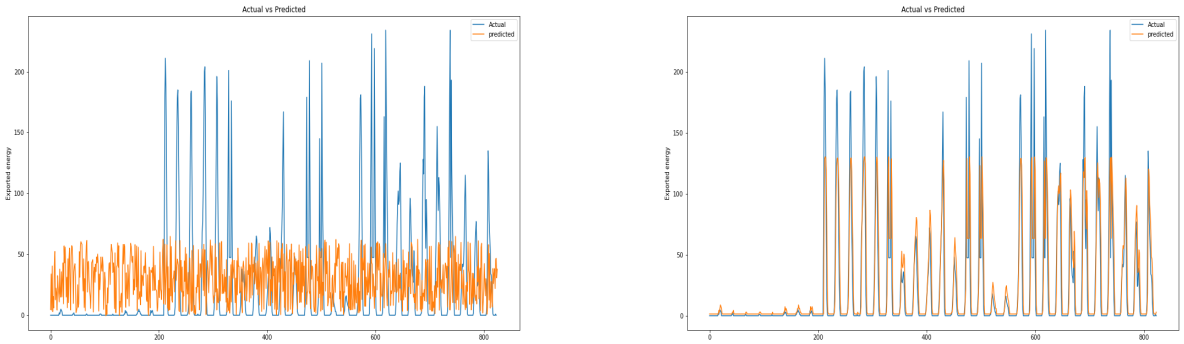

Figure 9: Actual and predicted testing data set of SPP2 - LR used on the left and LSTM used on the right.

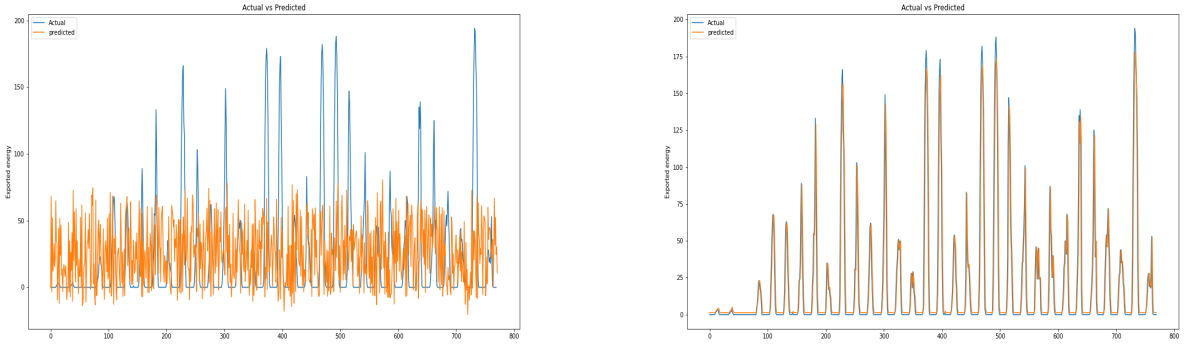

Figure 10: Actual and predicted testing data set of SPP3 - LR used on the left and LSTM used on the right.

4.3 Period of historical data

Different lengths of time series can increase or decrease accuracy. The too long period of historical data can be as bad as too short. Now it is more than 13000 values of the time series of each solar power plant. The main goal is to create an hourly forecast using LSTM. An assumption that time series with 13000 hourly values is too long for an hourly forecast is very logical. Thus the simulation with different lengths of time series using multivariate LSTM is done with data sets of each solar power plant.

One main change is done compared to the last chapter. From this point of the experimental part multivariate LSTM is used instead of univariate LSTM. This change is done wanting to include weather and lighting prognosis to forecast. In other words, specification of weather and lighting conditions can give better results than forecasting with only one time series of historical data. Additional metrics of sunny hours, hourly temperature, wind speed, and cloud cover are used for the forecasting. Comparison between forecasting using univariate and multivariate LSTM with all solar power plants and leaving the same parameters as they were in the previous chapter is visible in the Table 3 below. The difference between the accuracy of forecasting with univariate and multivariate LSTM is not significantly high for all time series. However, it allows using multivariate LSTM for further experiments.

	Uni-LSTM	Multi-LSTM
SPP1	0.873923	0.876157
SPP2	0.722089	0.719518
SPP3	0.857919	0.859472

Table 3: Evaluation of the forecasting performance of Univariate and Multivariate LSTM models on the test data set using the R^2 score.

To find out the necessary length of historical data for the best accuracy, the experiments of a period of historical data are started with a full length of data, which consists of more than 13000 values. During the experiments, time series is shortened until the optimal results are reached. The results are evaluated using the four different metrics of errors - MSE, RMSE, MAE, and R^2 regression score function and shown in the Table 4. The forecasting period - 2 full weeks or 336 values in hourly intervals.

Number of iterations	**SPP1**				**SPP2**				**SPP3**			
	MSE	**RMSE**	**MAE**	R^2	**MSE**	**RMSE**	**MAE**	R^2	**MSE**	**RMSE**	**MAE**	R^2
13000	57.3402	2.8341	7.5723	0.8761	228.840	4.8780	15.1274	0.7195	72.7532	3.5464	8.5295	0.8594
6500	396.879	12.9652	19.9218	0.8632	2114.312	25.4721	45.9816	0.4816	468.3857	13.9424	21.6422	0.8812
3700	224.777	10.9029	14.9925	0.8794	1886.355	24.4698	43.4321	0.5065	425.0157	12.7137	20.6159	0.8642
3000	198.1211	8.9451	14.0755	0.8793	1723.480	22.5518	41.5148	0.5407	407.5131	12.6587	20.1869	0.8754
2500	187.6588	9.4093	13.6988	0.8599	1526.716	21.3910	39.0732	0.5482	417.972	13.6228	20.4443	0.8782
2000	158.0455	7.9149	12.5716	0.8519	1488.465	20.8261	38.5806	0.5270	384.822	12.3596	19.6168	0.8888

Table 4: The comparison of values of the performance metrics calculated on the scaled validation data using the optimal parameters of LSTM and changing the number of iterations, required for the training of the network with 80% values for training set and 20% values for testing set.

At first, the time series is shortened to half of the full length (6500 values) and already shows better results, which is shown in the Table 3. As the length of time series is shortened more to 3700, 3000, 2500, or 2000 values, accuracy gets better on every step. That could make us think that the shorter the time series is, the better forecast gets. However, it is important not to forget that too short time series can cause a lack of data for training. Furthermore, line-plots in Figure 11 show that the first values are forecasted better than the last ones. When fewer iterations are taken for training the model, it also leads to less number of iterations for testing the model. Now it is splitted to 80 % for the training set and 20 % for the testing set. When the number of iterations is 2000 values, the forecast is prepared for 2 weeks in advance in hourly intervals. So far it shows the best results but it is important to check, what changes if we leave fixed 2 weeks of data for testing forecasting results and changing only the number of values for training the model. The results of these experiments are shown in the Table 5.

Number of iterations	**SPP1**				**SPP2**				**SPP3**			
	MSE	**RMSE**	**MAE**	R^2	**MSE**	**RMSE**	**MAE**	R^2	**MSE**	**RMSE**	**MAE**	R^2
2000	149.869	7.0107	12.2421	0.8351	1741.475	22.7350	41.7309	0.5134	412.775	13.4600	20.3168	0.8850
2500	160.7138	7.0476	12.6772	0.8231	1737.929	23.7350	41.6884	0.5144	434.495	13.8307	20.8445	0.8790
2200	159.8138	7.8659	12.6417	0.8241	1693.9650	23.4824	41.1578	0.5266	414.639	13.5660	20.3626	0.8845
1500	135.739	6.8145	11.6507	0.8506	1760.435	23.8293	41.9575	0.5081	432.594	13.6280	20.7989	0.8795
1800	144.849	7.5104	12.0353	0.8406	1790.274	23.5235	42.3116	0.4997	418.245	13.3639	20.4510	0.8835

Table 5: The comparison of values of the performance metrics calculated on the scaled validation data using the optimal parameters of LSTM and changing the number of iterations, required for the training of the network with fixed number of 336 values (2 weeks in hourly interval) for testing set.

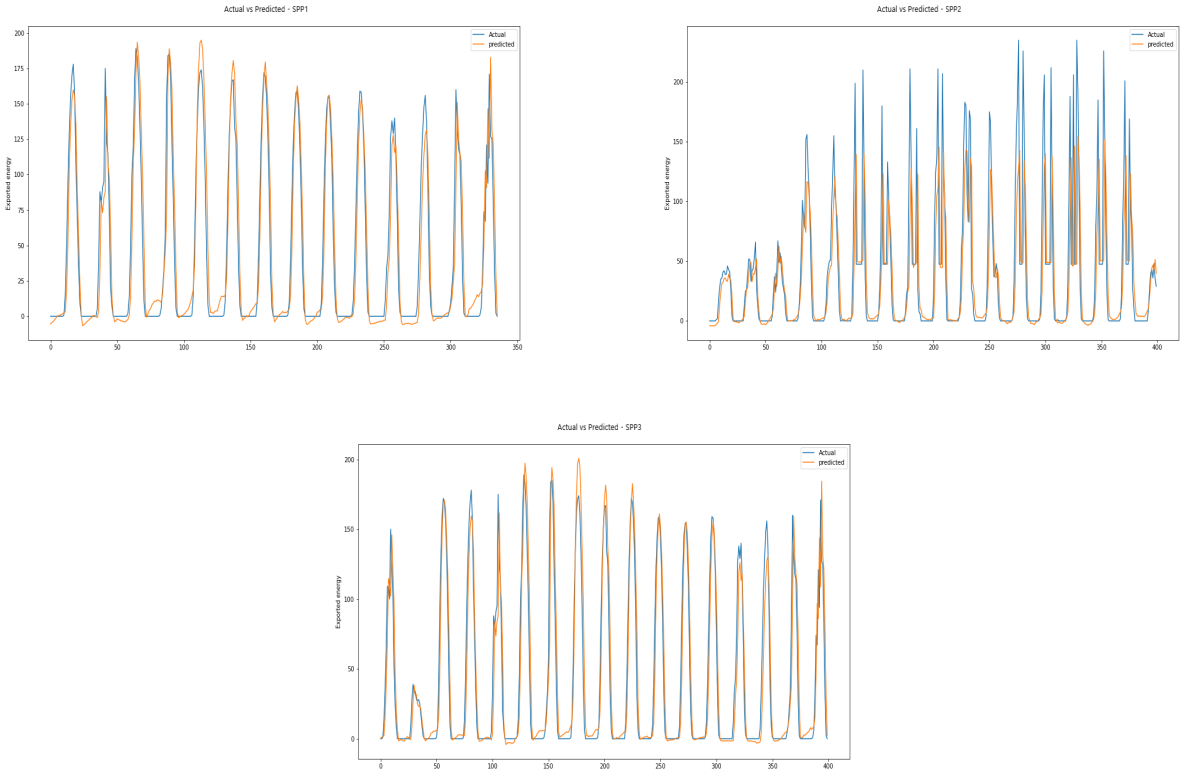

Figure 11: Actual and predicted testing data set with the 2000 iterations and 20% of testing set.

The experiments with different lengths of training data set show that too short time series for the training set decreases the accuracy of the forecast. The best example of it is shown with SPP3 data set. A number of iterations were changed from 2500 to 1500 in this experiment. When the number of iterations is 2500 in the forecast of SPP3, the R^2 score is 0.879. When the number of iterations is decreased to 2000, the R^2 score increases to 0.885. After it, the number of iterations is decreased even more and the R^2 score starts to decrease too. It is 0.8845 when 1800 iterations are chosen and 0.8795 when 1500 iterations are chosen. Similar results are noticed with SPP1 and SPP2 data sets too. Of course, different time series have different breakpoints for the result of forecasting starting to decrease. Each time series can be analyzed individually wanting to reach the best results. Especially it is necessary for SPP2 forecast because the pattern of energy produced by this solar power plant differs from others. However, at this point, the optimal result, which is 2000 iterations for 2 weeks forecast, is chosen for further analysis. R^2 score with 2000 iterations is 0.8850 of forecasting SPP3, is 0.8351 of forecasting SPP1, and 0.5134 of forecasting SPP2.

2000 values of the time series of each solar power plant are chosen as an optimal period of historical data for forecasting 336 values (2 weeks in hourly period). Experiments were done with a variation of a number of iterations from 1500 to 13000 values. Furthermore, the different proportions for the testing data set were analyzed. It was decided to prepare a forecast for 2 weeks in hourly intervals because a longer period decreases the accuracy of the forecast. Comparing univariate and multivariate LSTM forecasting gives similar results but including weather and lighting prognosis can help with

better accuracy in further experiments, thus multivariate LSTM is chosen for further analysis.

4.4 Hyperparameters tuning of Long Short Term Memory networks (LSTM)

Having all necessary input - different metrics of weather and lighting prognosis, length of historical data and purpose to prepare a forecast for 2 weeks, modifications of LSTM parameters can be made to achieve better accuracy.

At first, 3 different variations of LSTM networks for numerical time series are compared with each other using the data sets of 3 different solar power plants. Characteristics of Vanilla, Stacked, and Bidirectional LSTM are compared in detail in Chapter 3, hence practical application is shown in the experimental part. A Vanilla LSTM is an LSTM model that has a single hidden layer of LSTM units, and an output layer used to make a prediction. The second stacked-layer Rectified Linear Activation Function (or ReLU for short) is added to Vanilla LSTM, which is different from the previous multivariate LSTM model. That allows the neural network to learn nonlinear dependencies. The stacked LSTM model has referred to an LSTM model with multiple hidden LSTM layers that can be stacked one on top of another. The "return sequences" argument on the layer is set to the "True" in the LSTM model, which has the LSTM output a value for each time step in the input data, allowing us to use the three-dimensional output of the hidden LSTM layer as input to the next layer. Finally, on some prediction problems, it can be useful to use Bidirectional LSTM, which allows the LSTM model to learn the input sequence both forward and backward and concatenate both interpretations. It is implemented in Python by wrapping the first hidden layer in a wrapper layer called Bidirectional. These different variations of LSTM are applied to data of each solar power plant and the results are shown in the Table 6 below.

Variation of LSTM	**SPP1**				**SPP2**				**SPP3**			
	MSE	**MAE**	**RMSE**	R^2	**MSE**	**MAE**	**RMSE**	R^2	**MSE**	**MAE**	**RMSE**	R^2
Vanilla	159.081	7.3286	12.6127	0.8249	1724.726	24.8495	41.5298	0.5181	452.015	14.6768	21.2606	0.8741
Stacked	154.868	8.7473	12.4446	0.8295	1656.851	24.2979	40.7044	0.5370	367.634	10.9276	19.1737	0.8976
Bidirectional	144.146	7.9716	12.0061	0.8413	1827.798	24.0495	42.7527	0.4893	458.062	14.0652	21.4023	0.8724

Table 6: The comparison of the values of the performance metrics calculated from the scaled validation data using the different LSTM types with the same input and forecasting the values for 2 weeks values at hourly intervals.

The results show that using different types of LSTM leads to different forecasting results. Stacked LSTM outperforms Vanilla LSTM in forecasting each time series of energy produced by solar power plants. MSE of the forecast using Stacked LSTM for SPP1 is 154.868 when it is 159.081 using Vanilla LSTM, Results of SPP2 and SPP3 are similar - MSE of the forecast using Stacked LSTM for SPP2 is 1656.851 and 1724.726 using Vanilla LSTM, MSE of the forecast using Stacked LSTM for SPP3 is 367.634 and 452.015 using Vanilla LSTM. Nevertheless, the results of accuracy of forecasts using Bidirectional LSTM varies. R^2 score of the forecast using Bidirectional LSTM is 4% lower of SPP2 and 2.5% lower of SPP3 than R^2 score of the forecast using Stacked LSTM. Otherwise, Bidirectional LSTM fits the best for forecasting energy produced by SPP1 with the result 0.8413 of R^2 score, which is more than 1% higher than using Stacked LSTM. Nonetheless, the increase of accuracy using Bidirectional LSTM for SPP1 is lower than the decrease of it for SPP2 and SPP3. Thus Stacked LSTM is used in

this work for further analysis wanting to find the best LSTM model for forecasting energy produced by solar power plants including weather and lighting prognosis.

4.4.1 Hyperparameters tuning of Stacked Long Short Term Memory networks (Stacked LSTM)

The literature review shows that the LSTM with a sufficient number of hidden nodes could give appropriate results in time series forecasting. Until now, 4 hidden layers were used in all experiments with LSTM models. The construction of the Stacked LSTM network architecture requires defining a number of hidden nodes and there was chosen to try the performance of the network with 2-24 hidden nodes. In addition to that, model performance is tested with different activation functions in the hidden layer. In previous Section 4.4 ReLU activation function was used but different combinations of activation functions and the number of hidden layers can lead to more accurate results. ReLU, Sigmoid, SeLU, Softplus, and Swish activation functions are chosen for testing the performance of the Stacked LSTM network.

Activation function	Number of hidden layers	SPP1				SPP2				SPP3			
		MSE	MAE	RMSE	R^2	MSE	MAE	RMSE	R^2	MSE	MAE	RMSE	R^2
ReLU	2	174.382	7.7369	13.2053	0.8081	1826.809	22.2368	42.7411	0.4895	588.672	14.7042	24.2625	0.8360
ReLU	4	154.868	8.7473	12.4446	0.8295	1656.851	24.2979	40.7044	0.5370	367.634	10.9276	19.1737	0.8976
ReLU	5	144.075	8.0394	12.0031	0.8414	1806.431	23.1687	42.5021	0.4952	493.284	15.2628	22.2100	0.8626
ReLU	6	204.940	8.6347	14.3157	0.7745	1655.381	22.5772	40.6863	0.5374	455.771	14.3414	21.3487	0.8730
ReLU	8	196.895	8.3873	14.0319	0.7833	1766.256	22.5222	42.0268	0.5064	394.193	11.4829	19.8543	0.8902
ReLU	10	267.973	9.4467	16.3698	0.7051	1672.820	22.4630	40.9001	0.5326	367.951	12.2292	19.1820	0.8975
ReLU	16	182.232	8.3204	13.4993	0.7994	1588.479	22.4657	39.8557	0.5561	349.825	10.9829	18.7036	0.9025
ReLU	24	198.480	8.1904	14.0883	0.7816	1627.325	22.2442	40.3401	0.5453	367.343	12.1099	19.1662	0.8977
ReLU	20	189.351	8.7576	13.7605	0.7916	1668.657	22.3210	40.8492	0.5337	347.984	10.5920	18.6543	0.9031
Sigmoid	4	181.702	9.7439	13.479	0.8000	1743.588	25.9834	41.7562	0.5128	460.310	14.7835	21.4548	0.8718
Sigmoid	5	170.577	9.3307	13.0605	0.8123	1755.876	26.1729	41.9031	0.5093	457.842	14.7302	21.3972	0.8725
Sigmoid	6	176.219	9.4726	13.2747	0.8061	1758.662	25.1833	41.9364	0.5086	456.145	14.5785	21.3575	0.8729
Sigmoid	16	154.761	8.6346	12.4403	0.8297	1796.678	24.4889	42.3872	0.4979	462.532	14.9586	21.5065	0.8712
Sigmoid	24	145.230	8.1954	12.0511	0.8402	1820.507	24.1989	42.6674	0.4913	466.989	15.1437	21.6099	0.8699
Sigmoid	20	147.645	8.2571	12.1509	0.8375	1806.052	24.4915	42.4976	0.4953	465.257	15.0704	21.5698	0.8704
Sigmoid	28	145.724	8.2742	12.0716	0.8396	1830.406	24.5060	42.7832	0.4885	480.963	15.8131	21.9308	0.8660
SeLU	4	186.478	7.7057	13.6557	0.7948	1689.021	23.0949	41.0977	0.5280	395.454	12.1284	19.8860	0.8898
SeLU	5	192.218	8.8324	13.8643	0.7885	1688.432	22.7785	41.0905	0.5282	371.236	11.6508	19.2674	0.8966
SeLU	6	177.750	8.0131	13.3323	0.8044	1664.844	22.2454	40.8025	0.5348	373.207	11.5717	19.3185	0.8960
SeLU	16	202.850	8.3794	14.2425	0.7768	1678.649	22.3271	40.9713	0.5309	366.148	11.1778	19.1349	0.8980
SeLU	24	197.198	7.6038	14.0427	0.7830	1688.386	23.1909	41.0899	0.5282	373.178	11.0625	19.3178	0.8960
SeLU	20	202.484	8.0957	14.2296	0.7772	1654.624	22.4193	40.6770	0.5376	371.331	11.5460	19.2699	0.8966
Softplus	5	155.417	8.6483	12.4666	0.8289	1774.155	23.8396	42.1207	0.5042	506.179	15.4706	22.4984	0.8590
Softplus	6	145.128	7.9202	12.0469	0.8403	1819.263	23.7464	42.6528	0.4916	496.752	14.8268	22.2879	0.8616
Softplus	16	170.887	8.7097	13.0723	0.8119	1807.330	24.6927	42.5127	0.4950	471.647	15.2417	21.7174	0.8686
Softplus	24	209.674	11.9572	14.4801	0.7692	1815.318	27.2849	42.6065	0.4927	490.404	16.6692	22.1450	0.8634
Softplus	20	146.762	8.7031	12.1145	0.8385	1796.673	26.6088	42.3871	0.4979	468.884	15.1626	21.6537	0.8694
Swish	5	178.472	9.4498	13.3593	0.8036	1731.400	24.7320	41.6100	0.5162	442.372	13.6999	21.0326	0.8768
Swish	6	169.644	8.5259	13.0247	0.8133	1890.646	24.1091	43.4815	0.4717	429.282	14.0501	20.7191	0.8804
Swish	16	159.000	7.2360	12.6095	0.8250	1731.530	23.2903	41.6116	0.5161	399.721	12.3744	19.9930	0.8887
Swish	24	156.500	7.6688	12.510	0.8278	1747.934	22.2458	41.8083	0.5116	400.121	13.1402	20.0030	0.8885
Swish	20	142.609	7.1811	11.9419	0.8430	1708.516	23.4364	41.3342	0.5226	397.294	12.6877	19.9322	0.8893

Table 7: The comparison of the values of the performance metrics calculated from the scaled validation data using the different activation functions and number of hidden layers of Stacked LSTM.

The results are shown in the Table 7. It is important to test different activation functions with a different number of layers because the Stacked LSTM model can predict the results better with fewer hidden layers with one activation function and worse with another when it would forecast better with more hidden layers. For instance, the MSE result of forecasting energy produced by SPP1 is the best

(MSE = 144.075) with 5 hidden layers and it is one of the highest with 24 hidden layers (MSE = 198.480) using ReLU function but using Sigmoid activation function MSE is 170.577 with 5 hidden layers and only 145.230 with 24 hidden layers, which is opposite to ReLU.

Overall, the R^2 score of forecasting with Stacked LSTM using different activation functions and hidden layers varies for 8-14% comparing the worst and the best prediction of each solar power plant. Variation of R^2 score of SPP1 is 0.7051-0.8430, SPP2 - 0.4717-0.5561 and SPP3 - 0.8360-0.9031. The worst and the best result of the R^2 score of SPP3 is reached even with the same activation function ReLU, only using a different number of hidden layers. That indicates how the choice of optimal parameters can improve the accuracy of time series forecasting.

Looking at the results of forecasting energy produced by 3 different solar power plants, it can be concluded that the most accurate forecast can get using the ReLU activation function with 16 hidden layers for SPP2 and SPP3 and Swish activation function with 20 hidden layers for SPP1, which is decided to use for further analysis. Stacked LSTM with ReLU activation function and 20 hidden layers shows a slightly better result for SPP3, but RMSE differs only by 0.5, which is not a big difference, thus it is decided to use the same parameters for both, SPP2 and SPP3, forecasting. The worst results are shown with the SeLU activation function for forecasting SPP1, the Sigmoid activation function for forecasting SPP2, and the Softplus activation function for forecasting SPP3. That lets us conclude that the activation function has to be adapted to each time series individually and can not be used as one optimal parameter for all forecasts of energy produced by solar power plants. Even if in this case Swish and ReLU activation functions seem better than Sigmoid, SeLU, and Softplus, different parameters will be used for further forecasting of SPP1 compared with SPP2 and SPP3. Thus parameters of each forecast with the Stacked LSTM model have to be chosen carefully.

To sum it up, parameters of each forecast with the Stacked LSTM model have to be chosen individually. Forecasting with different activation functions of Stacked LSTM requires a different number of hidden layers for the best result of accuracy. In our case, the R^2 score of forecasting with Stacked LSTM using different activation functions and hidden layers varies for 8-14% comparing the worst and the best prediction of each solar power plant. It is chosen to use Swish activation function with 20 hidden layers for SPP1 with R^2 score 0.8430 and ReLU activation function with 16 hidden layers for SPP2 with R^2 score 0.5561 and SPP3 with R^2 score 0.9025.

4.4.2 Application of the rectified linear unit activation function for forecasting non-negative values

Looking at the predicted values in Figure 11, we can face the problem that during the night periods, when the sun is not shining, our forecast sometimes predicts negative values. That does not only lower the final result of the accuracy but it is also a huge problem of logic. Solar energy production can never be lower than 0. The Stacked LSTM model needs a modification with allowance to forecast values, which are only in the interval between 0 and infinity. That is the place, where rectified linear unit activation function can help.

The Stacked LSTM model is modified by adding rectified linear unit activation function to the last layer. That restricts the output domain for predicting non-negative values from 0 to infinity. The difference of forecast with and without rectified linear unit activation function on the last layer is visible

in the Figures 12, 13 and 14. There are graphs of predicted and actual values with negative predictions on the left figures and graphs of predicted and actual values without negative predictions on the right figures of each solar power plant.

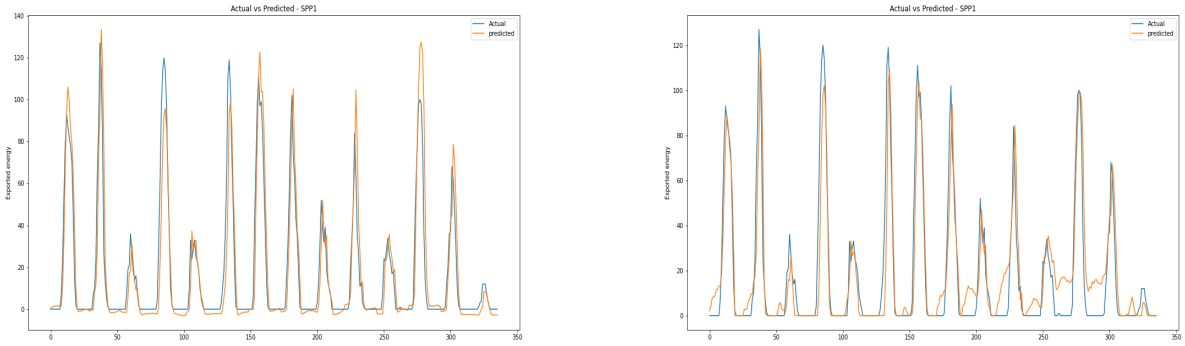

Figure 12: Actual and predicted testing data set of SPP1 - with negative values on the left and without negative values on the right.

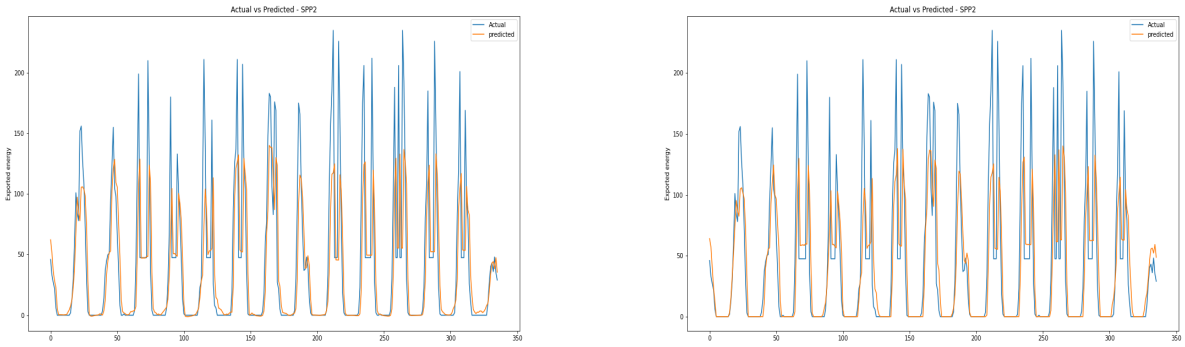

Figure 13: Actual and predicted testing data set of SPP2 - with negative values on the left and without negative values on the right.

Metrics of errors do not show the improvement of accuracy for every solar power plant. RMSE decreases a bit for SPP3 - RMSE is 18.7036 with negative predicted values and it is 18.5439 with predicted values from 0 to infinity. RMSE of SPP2 increases from 39.8557 to 39.9906 with added rectified linear unit activation function on the last layer but the difference is low and can become different with any other simulation with the same parameters. Also, graphs show that it was fewer forecasted negative values for SPP2 even without added rectified linear unit activation function on the last layer, thus accuracy does not increase but the logical mistake is corrected. A totally different situation appears with SPP3. Even with a lot of negative values in the forecast before, added rectified linear unit activation function on the last layer increases RMSE from 11.9419 to 12.6498. That can be explained by looking at the graph on the right in the Figure 12. The accuracy of forecasted values gets worse in the forecast of the second week compared with the forecasted values with negative values on the graph on the left. The difference with forecasting SPP2, SPP3, and SPP1 is that SPP2 and

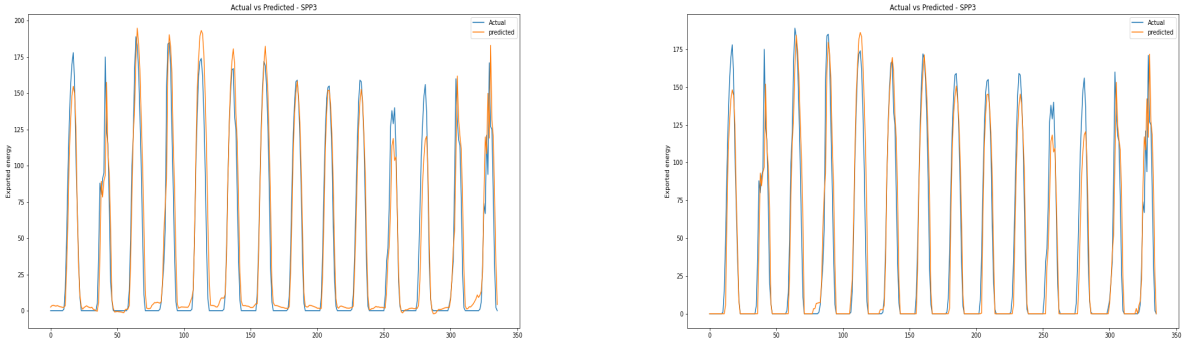

Figure 14: Actual and predicted testing data set of SPP3 - with negative values on the left and without negative values on the right.

SPP3 are forecasted with Stacked LSTM model with activation function ReLU and 16 hidden layers, and SPP1 is forecasted with Stacked LSTM model with activation function Swish and 20 hidden layers. That could be the reason for the worse forecast, thus it will be checked how the accuracy of forecasting SPP1 would look with the Stacked LSTM model with activation function ReLU and 16 hidden layers.

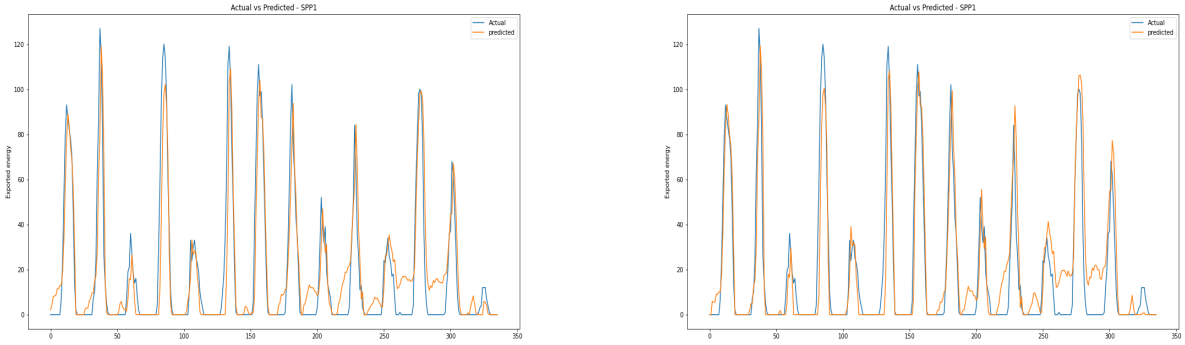

Figure 15: Actual and predicted testing data set of SPP1 - with activation function Swish and 20 hidden layers on the left and activation function ReLU and 16 hidden layers on the right.

Experiment shows that no matter which activation function is chosen, the accuracy of the forecast is going lower for the second week for SPP1 (Figure 15). RMSE with activation function Swish and 20 hidden layers on the left is 12.6498 and RMSE with activation function ReLU and 16 hidden layers is 13.1890. The result is even worse with activation function ReLU, thus the first parameters - activation function Swish and 20 hidden layers - will be left for further analysis keeping in mind that the forecast of SPP1 is less reliable for the second week than for the first week. Nevertheless, the logical error is fixed by adding rectified linear unit activation function to the last layer.

To sum it up, added rectified linear unit activation function to the last layer of the Stacked LSTM model corrects the logical error of forecasting negative values of energy produced by solar power plants. Forecasted values are limited to the interval from 0 to infinity. The accuracy increases negligibly for SPP2 and SPP3 and it even decreases for SPP1. Nevertheless, the experiment shows that parameters,

used in Chapter 4.4.1 can be kept for further analysis of time series, and predicted negative values can never be correct in energy production. Thus adding rectified linear unit activation function to the last layer of the Stacked LSTM model is necessary for forecasting solar power production.

4.5 Accuracy of forecasts for different seasonal periods

Accuracy can fluctuate at different periods of the year. So far all experiments are made for forecasting energy production for 2 weeks in October using the data from July. We have sufficiently good forecasting results for SPP1 and SPP3 on this period but they can differ at other times of the year. Thus 8 different forecasting periods of the year are tested for each solar power plant and results are compared. The same parameters, which showed the best results in the previous experiments - Stacked LSTM with activation function Swish and 20 hidden layers for forecasting SPP1 and Stacked LSTM with activation function ReLU and 16 hidden layers for forecasting SPP2 and SPP3 are used for forecasting.

Forecasting period	**SPP1**				**SPP2**				**SPP3**			
	MSE	**MAE**	**RMSE**	R^2	**MSE**	**MAE**	**RMSE**	R^2	**MSE**	**MAE**	**RMSE**	R^2
10-01 - 10-14	163.966	8.3321	12.8049	0.8195	1598.337	22.3347	39.9792	0.5534	330.266	9.8822	18.1732	0.9080
10-15 - 10-28	121.229	8.4736	11.0104	0.7675	2222.333	25.9029	47.1416	0.4550	281.221	8.6234	16.7696	0.7715
11-17 - 11-30	26.6992	2.8777	5.1671	0.7394	640.6099	13.9254	25.3102	0.6941	46.7189	2.9567	6.8351	0.7160
02-15 - 02-28	3687.927	36.8422	60.7283	0.0879	310.415	6.0522	17.6186	0.1843	33.2598	2.3205	5.7671	0.6502
05-16 - 05-29	819.097	18.5150	28.6198	0.8533	2053.234	24.9933	45.3126	0.4940	458.717	12.9660	21.4177	0.8978
04-16 - 04-29	3890.486	38.3214	62.3737	0.2274	2764.084	28.8590	52.5745	0.2751	960.199	17.9348	30.9870	0.7523
08-16 - 08-29	784.935	17.5222	28.0167	0.7814	2726.068	29.7659	52.2117	0.4153	395.438	12.9730	19.8856	0.9275
07-16 - 07-29	1110.472	22.3394	33.3237	0.8137	2293.240	27.2822	47.8877	0.4506	1148.963	22.5342	33.8963	0.7662

Table 8: The comparison of the values of the performance metrics calculated from the scaled validation data for different period forecasting.

Results in the Table 8 are compared using the same error metrics - MSE, MAE, RMSE, and R^2 score, as it is done in previous experiments too. Nevertheless, MSE, MAE, and RMSE evaluation of errors do not give a lot of information and can not be used for comparison of forecasting of different periods of time. The reason behind that is that different actual values exist in different periods of time and MSE, MAE and RMSE measure the difference between predicted and actual values in units. That leads to totally different values of metrics of errors, which cannot be compared. In spite of that R^2 score is the proportion of forecasted and actual values, thus it can be trusted for evaluation of models performance.

R^2 score in the Table 8 shows Stacked LSTM with activation function Swish and 20 hidden layers for forecasting SPP1 and Stacked LSTM with activation function ReLU and 16 hidden layers for forecasting SPP2 and SPP3 perform almost the best for firstly forecasted period from 1st October to 14th October, where R^2 score is 0.8195 for SPP1, 0.5534 for SPP2, and 0.9080 for SPP3. Despite that, there is at least one forecast of each solar power plant with higher accuracy than the forecast of the period from 1st October to 14th October. The energy produced of SPP1 from 16th May to 29th May is forecasted with R^2 score of 0.8533, the energy produced of SPP2 from 17th November to 30th November is forecasted with R^2 score of 0.6941, the energy produced of SPP3 from 16th August to 29th August is forecasted with R^2 score of 0.9275. All mentioned periods are different but they have one thing in common - it is the end of the season (autumn/spring/summer). It can be concluded that having the historical energy production and weather data of the whole season, the forecast gets better. Even so, forecasting

energy produced by each solar power plant on different periods, R^2 score is higher than forecasting the period from 1st October to 14th October only one time for each time series. Stacked LSTM performs the best with current parameters for the period from 1st October to 14th October. Thus it leads to the conclusion that parameters of Stacked LSTM have to be adapted separately having individual forecasting cases. The Stacked LSTM model should not be used as a universal model with the same parameters for different cases.

Another important result of this experiment is that the lowest value of R^2 score is obtained forecasting period from 15th to 28th February for each solar power plant. It is 0.0879 for SPP1, 0.1843 for SPP2, and 0.6502 for SPP3. The value of R^2 score of forecasting energy produced by SPP1 shows that predicted values are almost totally different than actual values. February is the winter month when the number of sunny days is the lowest during a whole year in Lithuania. During winter there are days when the sun is not shining per day at all. Hence Stacked LSTM model with current parameters does not catch the pattern of time series and can not predict energy production with high accuracy. The Stacked LSTM model with activation function Swish and 20 hidden layers for forecasting SPP1 and Stacked LSTM with activation function ReLU and 16 hidden layers for forecasting SPP2 and SPP3 are not suitable for energy production by solar power plants forecasting in the winter period. Further research with other parameters of Stacked LSTM or even another model is required for forecasting energy produced by solar power plants in the winter period. The stacked LSTM model is suitable for forecasting other seasons, only other parameters can be used for getting higher accuracy between predicted and actual values.

To summarize, the Stacked LSTM model forecasts energy production values more accurately at the end of the season, for instance, the end of August, when there are more historical data of similar weather prognosis. Furthermore, the parameters of Stacked LSTM, as activation function and a number of hidden layers, should be adapted to forecasting each period individually. Stacked LSTM is not so universal that it would forecast most accurately with the same parameters for each case. Eventually, further research is required for forecasting energy production by solar power plants during the winter period. The Stacked LSTM model with activation function Swish and 20 hidden layers for forecasting SPP1 and Stacked LSTM with activation function ReLU and 16 hidden layers for forecasting SPP2 and SPP3 are not suitable for this forecast with the current parameters of the model.

4.6 Confidence intervals of produced energy forecast by forecasting minimum and maximum energy production per interval

A confidence interval is a range of estimates for an unknown parameter, defined as an interval with a lower bound and an upper bound. In our case, the average energy produced by solar power plants is forecasted, and having its lower and upper bounds can help for resource planning. Then not only average energy production could be known but the minimum and maximum values of possible energy production of each interval would be known too. A good opportunity to remember that the real dataset of three solar power plants contains not only historical data of average exported energy per interval but historical data of minimum and maximum exported energy per interval too. This data is used for forecasting the minimum and maximum energy production of each solar power plant. Forecast of the minimum energy production is a lower bound and forecast of the maximum energy production

is an upper bound of average energy production forecast and the predicted values become confidence intervals for each forecasted value of average energy production forecast.

The final parameters of models, which were applied in Chapter 4.4.2 and gave the most accurate result for each solar power plant separately, are used for forecasting average, minimum and maximum energy produced by solar power plants. They are Stacked LSTM model with activation function Swish and 20 hidden layers for forecasting SPP1 and Stacked LSTM model with activation function ReLU and 16 hidden layers for forecasting SPP2 and SPP3. The forecast is made separately for average, minimum, and maximum energy production for the period from 14th October to 28th October, results are saved and merged. The patterns of historical data of average, minimum and maximum exported energy of SPP1, SPP2 and SPP3 are similar, thus the same parameters fit for all forecasts of the same solar power plant. The results of accuracy are similar for minimum and maximum produced energy forecasts as they are for average produced energy forecasts in Chapter 4.4.2. Visualisation of forecasted energy produced by each solar power plant is shown in the Figures 16, 17, and 18.

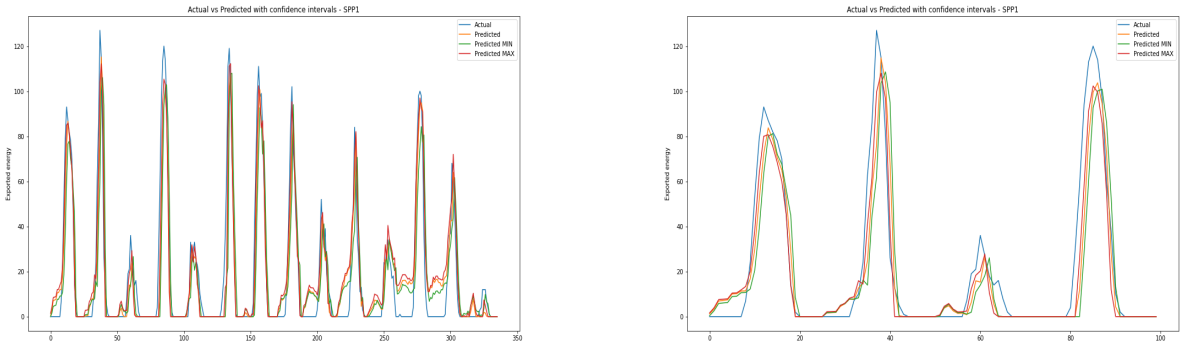

Figure 16: Actual and predicted testing data set with predicted minimum and maximum values of the interval of SPP1 - 2 weeks forecast on the left and first 100 forecasted values on the right.

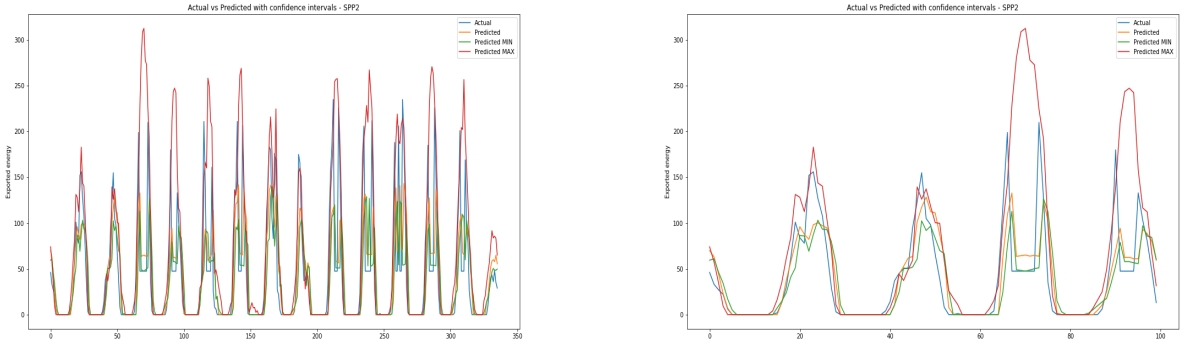

Figure 17: Actual and predicted testing data set with predicted minimum and maximum values of the interval of SPP2 - 2 weeks forecast on the left and first 100 forecasted values on the right.

Looking to the Figures 16, 17, and 18, it is visible that confidence intervals add more information to the forecast. Values of possible energy production become wider having minimum and maximum

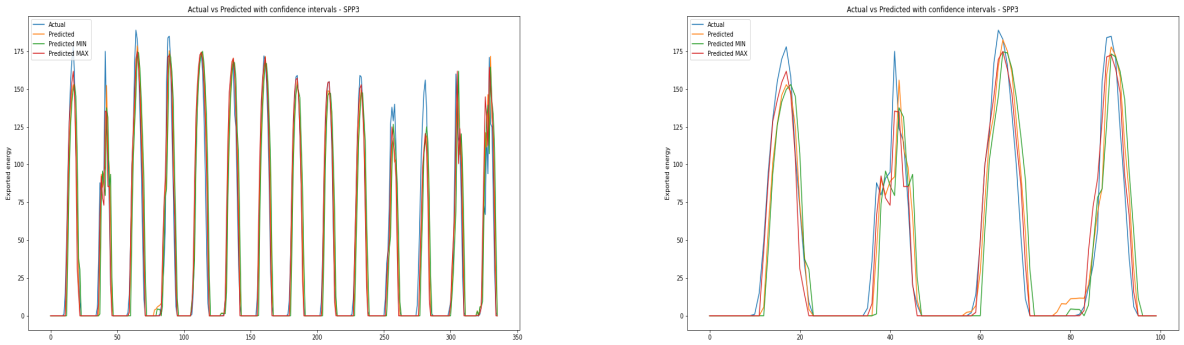

Figure 18: Actual and predicted testing data set with predicted minimum and maximum values of the interval of SPP3 - 2 weeks forecast on the left and first 100 forecasted values on the right.

possible values of produced energy of the interval. Also, it can be seen with the naked eye that when the forecast is more accurate, intervals between the minimum and maximum values are smaller as in Figure 18, where R^2 score is higher than 0.9. When the forecast is less accurate as in Figure 17, where the R^2 score is around 0.55, the intervals between the minimum and maximum values are bigger.

Intervals of minimum and maximum values are not always trustworthy. Some minimum values are higher than the average forecasted values, and some maximum values are lower than the average forecasted values. That is visible in Figures 16, 17, and 18 and it happens because every forecast - minimum, maximum or average,- is done separately. Thus predicted values can vary and pass each other. Despite that and minimum and maximum forecasted values, which are equal to average forecasted values (for instance, 0, when there is no sun), intervals between the minimum and maximum values vary from 0.0077 to 62.2662 of SPP1, from 0.3623 to 267.7448 of SPP2 and from 0.2260 to 64.7694 of SPP3. As mentioned before, these results confirm that the difference between the maximum and the minimum predicted value is lower if the forecast is more accurate. Furthermore, a lower difference between the minimum and maximum forecasted value or, in other words, the smaller confidence interval can help to think that the predicted value of average produced energy will be more accurate than the forecasted value with a bigger confidence interval.

To summarize, adding confidence intervals to the main energy production forecast by forecasting the minimum and maximum energy production values of each forecasted value gives additional information for resource planning. Confidence intervals of average energy production help to plan the scenarios if minimum or maximum energy is produced. Nevertheless, it would not always be reasonable to trust intervals of minimum and maximum values. Since minimum and maximum forecasts are done separately, minimum values sometimes are higher and maximum values sometimes are lower than the average energy production forecast. A lower difference between the minimum and maximum forecasted helps to recognize that the predicted value of average produced energy will be more accurate than the forecasted value with a higher difference.

5 Conclusions

The main idea of this work is to get the best accuracy forecasting energy produced by solar power plants including weather and lighting prognosis. The literature review shows that using machine learning models, Long Short-term memory networks, in particular, can lead to the best results. However, the impact of changes in different variations of LSTM for solar energy forecasting is not considered and this is the area of interest.

The first objective of this study was to ascertain if machine learning models could outperform traditional statistical models. Simulation results show that the neural networks models, as in this case LSTM, are able to outperform commonly used statistical Linear Regression when the real data sets of energy exported by 3 different solar power plants were used. Forecasting with long short time memory networks leads to at least more than 3 times better results in the accuracy of forecasts than forecasting with linear regression. R^2 score of forecast of SPP1 was only 0.2539 using Linear Regression when R^2 score of forecast of SPP1 using Long Short-term Memory model had a result 0.8739. R^2 score results of forecasts of other solar power plants improved even more - R^2 score of forecast of SPP2 using LR - 0.1127, using LSTM - 0.7220 and R^2 score of forecast of SPP3 using LR - 0.2127, using LSTM - 0.8579. Visualisations of predicted and actual values in the Figures 8, 9 and 10 showed that linear regression does not catch forecasting pattern at all and forecasting with classical traditional statistical model looks more like random walk than actual prediction. That confirms current researches that using even simple neural networks can have better forecasting results than using classical statistical models.

Other objectives of this work were to select the best period of historical data for 2 weeks forecasts in hourly intervals and to analyze LSTM accuracy for seasonality. Experiments showed that the best results are reached with approximately 1700 values of historical data in an hourly period, which is equal to 10 weeks, with RMSE 0.8579 of SPP1, 20.8261 of SPP2, 12.3596 of SPP3. Accuracy varied a bit between different solar power plants but the results were not statistically significant that it would be necessary to adapt values of historical data to each time series separately. Talking about seasonality, the accuracy of forecasting gets better in the period of the end of the season when there is enough historical data of similar weather. For instance, the energy produced of SPP3 from 16th August to 29th August is forecasted with a R^2 score of 0.927, which is the highest value compared to forecasts of other periods. On the contrary, at the beginning of the season, when we do have not enough historical data of it, the accuracy is the lowest - the worst forecasting results were get of the winter period of February when the sun does not shine every day and LSTM does not catch a pattern. LSTM model with current parameters is not adapted for forecasting solar power production in the wintertime. Further analysis is necessary for adapting the model for forecasting energy production of all seasons. Results were similar with data sets of all solar power plants as the weather in different parts of Lithuania does not vary significantly.

Comparison between univariate and multivariate LSTM shows that multivariate LSTM slightly outperforms univariate LSTM. When historical data of weather components, such as daily sunny hours, hourly temperature, cloud cover, and wind speed, were included, forecasting improvement was improved by only 0.3 % of R^2 score of different solar power plants on average. Nevertheless, this small improvement can help in the future prognosis. After it, an analysis between different variations of LSTM for

forecasting solar power energy with weather and lighting prognosis was made. Results show that for all time series of 3 different solar power plants the Stacked LSTM performs the best if it is compared with Vanilla LSTM and Bidirectional LSTM. All other created modifications show worse accuracy than the Stacked LSTM. Possibly, LSTM combinations with more difficult machine learning models may be able to get more accurate predictions for the energy produced by solar power plants including weather and lighting prognosis, and outperform Stacked LSTM. Moreover, forecasting accuracy was lower because the LSTM model was forecasting negative values of exported energy in the night hours. That was incorrect because exported energy can never be lower than 0. Improvement, which limits forecasting values from 0 to infinity, was made, and that improved the forecasting accuracy of RMSE by 0.15 of SPP2 and 0.7 of SPP3. RMSE of SPP1 slightly decreased by 0.15 because the forecast of the second week get less accurate. If the forecast is made for only one week, the result of accuracy could possibly be better. Nevertheless, the modification fixed the illogical error and it was used in other experiments.

Finally, forecasts of minimum and maximum energy produced by solar power plants in hourly intervals are created and added to the average energy production forecast. That adds confidence intervals to the main energy forecast. Having this feature, information is given not only about average planned energy production. Information about minimum and maximum energy production helps to plan the scenarios if minimum or maximum energy is produced.

Taking all results into account, it could be concluded that the LSTM model fits very well for forecasting energy produced by solar power plants including weather and lighting prognosis. Nevertheless, the accuracy of forecasting gets worse in the period of the new beginning of the season when the weather is changing and the accuracy is so much higher at the end of the season when there is enough historical data of similar weather and lighting prognosis. In addition, Stacked LSTM forecasts very weakly energy produced by solar power plants in the winter period and other variations of model or another model should be adapted for this period forecasting. Also, it can be concluded that more difficult modifications of LSTM, which expands the time of preparation of forecast and compiling results, is not necessarily lead to better accuracy. All modifications of LSTM, which were analyzed in this work, do not outperform Stacked LSTM. However, more difficult combinations with more difficult machine learning models could be performed in future analysis. Finally, forecasting minimum and maximum possible values next to the average forecast of solar energy production gives valuable insights and should be considered to use in the future.

References

[1] Voyant, Cyril and Notton, Gilles and Kalogirou, Soteris and Nivet, Marie-Laure and Paoli, Christophe and Motte, Fabrice and Fouilloy, Alexis. Machine learning methods for solar radiation forecasting: A review, *In: Renewable Energy*, 2017, p.p. 569-582.

[2] Chaturvedi, DK and Isha, I. Solar power forecasting: A review, *In: International Journal of Computer Applications*, 2016, p.p. 28-50.

[3] Nam, Seungbeom and Hur, Jin. Probabilistic forecasting model of solar power outputs based on the naive Bayes classifier and kriging models, *In: Energies*, 2018.

[4] AlKandari, Mariam and Ahmad, Imtiaz. Solar power generation forecasting using ensemble approach based on deep learning and statistical methods, *In: Applied Computing and Informatics*, 2020,

[5] Visser, Lennard and AlSkaif, Tarek and van Sark, Wilfried. *Benchmark analysis of day-ahead solar power forecasting techniques using weather predictions.* 2019, p.p. 2111-2116.

[6] Urmee, Tania and Harries, David and Schlapfer, August. Issues related to rural electrification using renewable energy in developing countries of Asia and Pacific, *In: Renewable Energy*, 2009, p.p. 354-357.

[7] Rodriguez, Jose and Pontt, Jorge and Correa, Pablo and Lezana, Pablo and Cortes, Patricio. Predictive power control of an AC/DC/AC converter, *In: Fourtieth IAS Annual Meeting. Conference Record of the 2005 Industry Applications Conference*, 2005, p.p. 934-935.

[8] Eurostat, *Renewable energy statistics*, 2020-12. https://ec.europa.eu/eurostat/statistics-explained/index.php/Renewable-energy-statistics.htm/

[9] Intec Energy, *Solar Monitoring System*, 2019-10. https://www.in-tecenergy.com/solar-monitoring-system/

[10] EnergySage, *Solar Monitoring System*, 2020-07. https://www.energysage.com/solar/solar-operations-and-maintenance/solar-monitoring-systems/

[11] Sun, Mucun and Feng, Cong and Zhang, Jie. Probabilistic solar power forecasting based on weather scenario generation, *In: Applied Energy*, 2020.

[12] Lotfi, Mohamed and Javadi, Mohammad and Osório, Gerardo J and Monteiro, Cláudio and Catalão, João PS. A novel ensemble algorithm for solar power forecasting based on kernel density estimation, *In: Energies*, 2020.

[13] Rana, Mashud and Rahman, Ashfaqur. Multiple steps ahead solar photovoltaic power forecasting based on univariate machine learning models and data re-sampling, *In: Sustainable Energy, Grids and Networks*, 2020.

[14] Gao, Mingming and Li, Jianjing and Hong, Feng and Long, Dongteng. Day-ahead power forecasting in a large-scale photovoltaic plant based on weather classification using LSTM, *In: Energy*, 2019.

[15] Shcherbakov, Brebels, Shcherbakova, Tyukov, Janovsky and Kamaev. A Survey of Forecast Error Measures, *In: World Applied Sciences Journal*, 2013.

[16] Hyndman. Another look at forecast-accuracy metrics for intermittent demand, *In: Foresight: The International Journal of Applied Forecasting*, 2006.

[17] Mukaka. Statistics Corner: A guide to appropriate use of Correlation coefficient in medical research, *In: Malawi Medical Journa*, 2012.

[18] Kashefi, M., Daliri, M.R.. A stack LSTM structure for decoding continuous force from local field potential signal of primary motor cortex (M1), *In: BMC Bioinformatics*, 2021.

[19] Graves and Schmidhuber. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *In: Neural Netw., vol. 18, nos. 5–6, pp. 602–610*, 2005.

[20] Gers, Schmidhuber, and Cummins. Learning to forget: Continual prediction with LSTM. *In: Proc. 9th ICANN, vol. 2*, 1999.

[21] Gers and Schmidhuber. Recurrent nets that time and count. *In: Proc. IEEE-INNS-ENNS IJCNN, vol. 3*, 2000.

[22] Greff, Srivastava, Koutník, Steunebrink, and Schmidhuber. LSTM: A Search Space Odyssey, *In: IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, VOL. 28, NO. 10*, 2017.

[23] Graves. Generating Sequences With Recurrent Neural Networks, 2014.

[24] Wei. Development of Stacked Long Short-Term Memory Neural Networks with Numerical Solutions for Wind Velocity Predictions, *In: Hindawi*, 2020.

[25] Weninger, Geiger, Wöllmer, Schuller, and Rigoll. Feature enhancement by deep LSTM networks for ASR in reverberant multisource environments, *In: Computer Speech Language*, 2014.

[26] Graves. Supervised Sequence Labelling with Recurrent Neural Networks. 2012.

[27] Hochreiter and Schmidhuber. Long short-term memory. *In: Neural Computation*, 1997.

[28] Liang, Nguyen, and Jin. A multi-variable stacked long-short term memory network for wind speed forecasting. *In: Proceedings of the 2018 IEEE International Conference on Big Data*, 2018.

[29] Zhang, Tan, and Wu. Ship Motion Attitude Prediction Based on an Adaptive Dynamic Particle Swarm Optimization Algorithm and Bidirectional LSTM Neural Network. 2020.

[30] Schuster and Paliwal. Bidirectional recurrent neural networks. *In: IEEE Transactions on Signal Processing*, 1997.

[31] Diskin. Definition and Uses of the Linear Regression Model. 1970.

[32] Yuan, Ekici, Lu, Monteiro. Dimension reduction and coefficient estimation in multivariate linear regression. 2007.

[33] Wang, Qi, Liu. Photovoltaic power forecasting based LSTM-Convolutional Network. 2019.

[34] Qing, Niu. Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM. 2018.