International Journal of Electrical and Computer Engineering (IJECE)
Vol. 14, No. 3, June 2024, pp. 3415~3426
ISSN: 2088-8708, DOI: 10.11591/ijece.v14i3.pp3415-3426

Optimizing loss functions for improved energy demand prediction in smart power grids

Fariza Nussipova[1], Shynggys Rysbekov[1], Zukhra Abdiakhmetova[2], Amandyk Kartbayev[1]
[1]School of Information Technology and Engineering, Kazakh-British Technical University, Almaty, Kazakhstan
[2]Faculty of Information Technology, Al-Farabi Kazakh National University, Almaty, Kazakhstan

Article Info

Article history:

Received Dec 1, 2023
Revised Feb 7, 2024
Accepted Feb 20, 2024

Keywords:

Latent features
Neural networks
Power grid
Representation learning
Triplet loss

ABSTRACT

In this paper, our aim is to improve the accuracy and effectiveness of energy demand forecasting, particularly within modern electricity transmission systems and smart grid technology. To achieve this, we developed a hybrid approach that combines machine learning, representation learning, and other deep learning techniques. This approach is based on extracting essential features, including time-based attributes, identifiable trends, and optimal lags. The outcome of our investigation is the observation that triplet losses demonstrate remarkable accuracy, particularly when employed with a larger margin size and for longer prediction lengths. This finding signifies a substantial improvement in the precision and reliability of energy demand forecasting within modern electricity transmission systems. Our research not only improves predictive modeling in the power grid but also demonstrates the practical use of advanced analytics in addressing renewable energy integration challenges, refining energy demand forecasting for efficient management, system operation, and market analysis.

Corresponding Author:

Amandyk Kartbayev
School of Information Technology and Engineering, Kazakh-British Technical University
Tole Bi Street 59, Almaty, Kazakhstan
Email: a.kartbayev@gmail.com

1. INTRODUCTION

In recent years, the evolution of the electricity transmission landscape has been driven by the emergence of the smart grid, representing a substantial departure from traditional power systems. This next-generation power system stands out for its innovative grid infrastructure, primarily characterized by its connectivity capabilities. This connectivity is achieved through constant communication facilitated by the internet of things (IoT), a network of interconnected computing devices, and objects equipped with unique identifiers (UIDs). The IoT enables seamless data transfer across the network, eliminating the need for direct interaction. As a consequence, the smart grid generates a vast volume of data, necessitating advanced methodologies to extract valuable insights, make informed decisions, and optimize system operations.

In this context, our research addresses the critical challenge of renewable energy and electrical load forecasting within the smart grid framework. Accurate load forecasting is pivotal for effective energy management, informed market analysis, and prudent investment decisions. However, the integration of renewable energy sources introduces a significant degree of volatility and randomness into the power generation landscape. This volatility necessitates the maintenance of substantial reserve capacity within electric energy systems, leading to increased operational costs and reduced cost-effectiveness.

Moreover, the intermittent nature of renewable sources, such as wind and photovoltaic energies, poses additional challenges, particularly in low inertia power systems [1]. To address these complexities and advance the state of the art in energy forecasting, our research places a central focus on representation learning. This research domain aims to identify and extract latent or hidden features from data, enabling more accurate and informative representations of the underlying information. Representation learning is achieved through feature learning, a primary procedure that enhances the capability to capture meaningful patterns and relationships within the data.

The existing literature on energy demand forecasting encompasses a spectrum of methodologies, each offering unique approaches to tackle this complex task. These approaches can be broadly categorized into physical methods, statistical models, machine learning models, and deep learning models, with some hybrid methods that combine multiple techniques. Physical methods, as described in [2], utilize numerical weather prediction models to simulate atmospheric dynamics based on physical principles and boundary conditions. While effective in forecasting atmosphere dynamics, these methods demand significant computational resources and are not well-suited for short-term forecasting horizons. Statistical methods have been widely applied, including autoregressive moving average (ARMA) models [3], the Kalman filters [4], Markov chain models [5], the gray theory [6], support vector machines [7], feedback neural networks [8], various artificial neural networks [9], and ARIMA [10]. These methods aim to establish mathematical relationships within time series data to make informed predictions. The selection between these two methodologies is often contingent upon the specific forecasting horizon and the computational resources at hand. While physical methods excel in capturing the nuances of complex atmospheric dynamics, their computational demands may make them less practical for short-term predictions. In contrast, statistical methods offer a versatile toolbox for modeling and can be more suitable for shorter forecasting timeframes, as they rely on historical data patterns.

Machine learning models, such as support vector machines [11], fuzzy support vector machines [12], and linear regression [13], have also been employed for energy demand forecasting. These models leverage historical data and covariates to predict energy consumption patterns. Deep learning models, a subset of machine learning, introduce advanced neural network architectures for energy forecasting. Recurrent neural networks (RNNs) have been applied to measure environmental consumption levels for different regions [14]. Autoencoders have been used to extract building energy demand patterns and predict future consumption [15].

Furthermore, some research introduces covariates into feature learning to enhance adaptability to sample characteristics. In study [16], the architecture is based on long short-term memory (LSTM) networks, which incorporate structured covariates into the framework. The paper explores different methods of integrating covariates, including using them as inputs to fully connected layers, incorporating them into the gates, or combining them with LSTM outputs in the final fully connected layer. Additionally, Franceschi *et al.* [17] presents a novel approach to unsupervised loss training, employing a deep convolutional neural network with dilated convolutions. This technique outputs fixed-length vector representations of time series data, regardless of their input length. The paper's use of triplet losses adapted to time series for negative sampling is a notable tool adopted in this research, with modifications to fit within the GluonTS model framework.

This paper presents a novel approach to energy forecasting in smart grids, focusing on integrating renewable energy sources. It employs representation learning to enhance energy demand forecasting accuracy, aiding in better decision-making and system optimization for both energy providers and consumers. The methodology stands out in the existing literature by offering a unique solution to the complexities of modern energy systems, particularly addressing the volatility and intermittent of renewable energy sources. By combining traditional methods with advanced deep learning techniques, the paper aims to improve energy management strategies in low inertia power systems.

2. METHOD

Conducting our research project involved navigating various methodologies and experimental design choices. Focusing on improving energy demand forecasting in the context of the smart grid and renewable energy integration, we encountered multifaceted challenges. Working with irregular and unpredictable smart meter data required us to carefully select and implement methods. Our approach included representation learning, innovative loss training, and covariate integration. Additionally, dealing with the volatility of renewable energy sources added complexity. In this methods section, we comprehensively describe our strategies for addressing these challenges and ensuring the rigor and reliability of our research.

The datasets used in this study comprise two primary sources. The first dataset [18], obtained from the Pecan Street Institute, covers regions within the independent system operators of the United States, specifically California, New England, New York, and the Midcontinent. These datasets include month-long data with varying time intervals, either hourly or every 5 minutes, and encompass energy types such as system load resource, wind and solar forecasts, and hourly load. The additional dataset originates from the Australian smart grid smart city (SGSC) project, gathered from 10,037 customers in New South Wales over the period of 2011 to 2014. This dataset records electricity consumption in kWh at half-hourly intervals for each meter.

2.1. Data preprocessing

In the process of preparing the data for analysis, several essential preprocessing tasks were undertaken to ensure the quality and suitability of the datasets. As shown in Figure 1, these preprocessing tasks addressed issues such as missing values, data redundancy, feature selection, and the creation of informative features. One of the initial steps in data preprocessing involved dealing with missing values within the datasets [19]. To achieve this, columns with a high proportion of missing values, exceeding 95%, were identified and subsequently removed. This action not only reduced the computational complexity but also enhanced the overall data quality. In one dataset, the elimination of columns with predominantly NA values led to a significant reduction in the number of variables, reducing it from 72 to 35 of the original variables. Additionally, any remaining missing values were replaced with zeros to maintain data integrity.

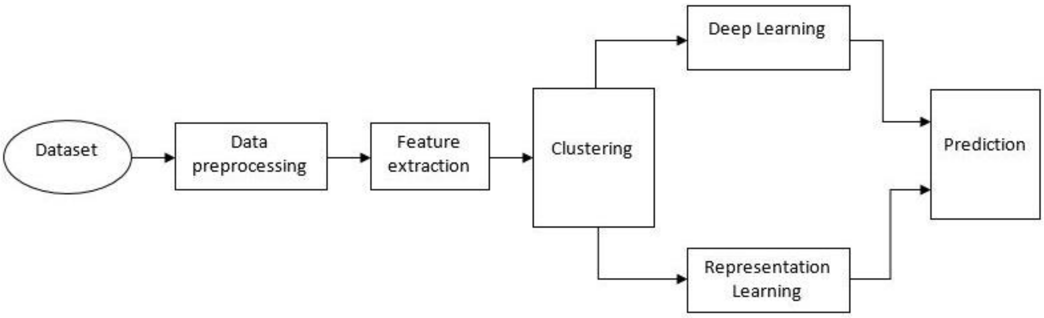
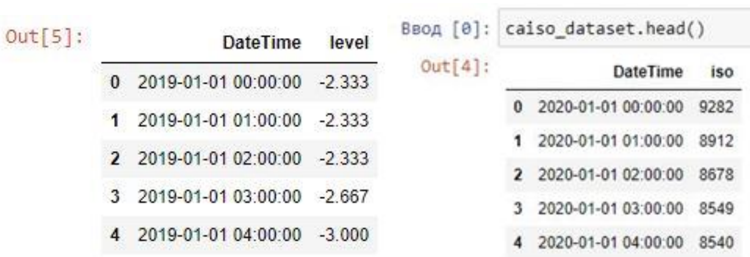


Figure 1. Data processing workflow

To ensure data quality and coherence, we conducted a thorough data examination process. This involved identifying rows with identical timestamps or those sharing the same timestamps occurring within the same day and hour. Duplicate entries with identical timestamps represent instances where the data was recorded multiple times, often inadvertently, and such redundancies can have adverse effects on the analysis and forecasting process. These duplicate entries, if left unaddressed, can introduce bias, distort analytical results, and undermine the precision of energy demand forecasts. By identifying and rectifying these issues through the removal or consolidation of duplicated data points, the overall quality and reliability of the dataset were significantly improved. This process ensures that each timestamp uniquely represents a specific moment in time, preventing data distortions that could potentially mislead the analytical models and compromise the validity of the study's findings as shown in Figure 2.



Out[5]:

	DateTime	level
0	2019-01-01 00:00:00	-2.333
1	2019-01-01 01:00:00	-2.333
2	2019-01-01 02:00:00	-2.333
3	2019-01-01 03:00:00	-2.667
4	2019-01-01 04:00:00	-3.000

Ввод [8]: caiso_dataset.head()

Out[4]:

	DateTime	iso
0	2020-01-01 00:00:00	9282
1	2020-01-01 01:00:00	8912
2	2020-01-01 02:00:00	8678
3	2020-01-01 03:00:00	8549
4	2020-01-01 04:00:00	8540

Figure 2. California dataset

Feature selection is a crucial step in refining data, focusing on variables strongly affecting energy demand forecasting accuracy by curating a more relevant subset [20]. This process prioritizes attributes with a substantial influence on forecasting models, offering significant advantages. Firstly, it streamlined the computational process, reducing complexity and enhancing computational efficiency. Secondly, it directed the models' attention toward the most salient and impactful data attributes, ensuring that the predictive models were well-informed by the most relevant information. These steps collectively contribute to the robustness of energy demand forecasting models, ensuring that they are primed to deliver accurate and reliable predictions that align with real-world energy consumption patterns. To gain insights into the relationships between different columns of data, correlation coefficients were calculated. This analysis helped identify patterns and dependencies between variables. Feature extraction was another vital aspect of data preprocessing, aiming to create new informative features that could enhance forecasting accuracy.

In the process of feature extraction, unique attributes were generated from the existing dataset to capture pertinent information that could enhance the forecasting process. One noteworthy feature introduced was the concept of "season," which was derived from the timestamp values. Beyond seasonality, additional date-time features were meticulously extracted, encompassing details such as the duration in minutes or hours within a day, whether a timestamp corresponded to typical business hours, whether it coincided with a weekend, the specific season of the year, and whether it aligned with a public holiday. These date-time features, depicted in Figure 3, offered valuable contextual information that could be leveraged to enrich the predictive models. By incorporating temporal characteristics, the models gained insights into the variations in energy demand patterns across different times of the day, week, season, and in response to holidays. This temporal context was invaluable in capturing the nuances of energy consumption behavior and enhancing the precision of forecasts [21].

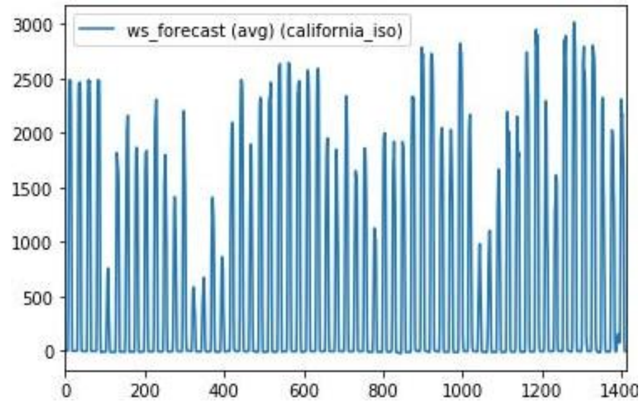


Figure 3. Average values for the California dataset by levels (y-axis) and windows (x-axis)

The introduction of lag features within the dataset allowed for the representation of values at previous time intervals. These lag features served as a bridge between historical data and future forecasts, enabling the analysis to consider the impact of past values on subsequent predictions at specific time intervals (referred to as k^{-th} lags). This temporal dimension added depth to the forecasting models, as it acknowledged the influence of historical trends on present and future energy demand. Window features were derived from the lag features, offering a condensed summary of values over fixed windows of prior time intervals. The windows were calculated using a rolling window approach, facilitating the extraction of essential statistics such as moving sums, moving means (within each window), moving variances, moving standard deviations, as well as moving minimum and maximum values. The window features provided a comprehensive view of temporal patterns and trends within the data, allowing for a more nuanced understanding of energy consumption dynamics.

2.2. Time series analysis

We leverage three key features for energy demand forecasting: the hour of the day, the trend, and the optimal lag. These features play a crucial role in our modeling process, contributing to the accuracy and reliability of our predictions. The hour feature is instrumental in creating distinct subseries within our dataset, based on the time of day. It categorizes each row into one of four categories: early morning, morning, afternoon, or evening. This classification allows our models to capture the variations in energy demand patterns that occur throughout the day. The optimal lag, another essential feature, is determined using the autocorrelation function. It iterates through different lag values, calculates the autocorrelation coefficient for each lag, and selects the lag that surpasses a predefined threshold, indicating its significance in capturing temporal dependencies within the data. The lag provides valuable information about the temporal dependencies within the data, helping our models account for historical trends and their influence on future demand as shown in Figure 4.

In pursuit of more accurate energy demand forecasting, we adopt a hierarchical time series approach, a method that organizes time series data based on both time-related and zone-related properties. This hierarchical organization offers a robust framework for our forecasting models, allowing us to apply innovative techniques like triplet losses to subseries within the hierarchy. In the time-related hierarchy, we divide the time series into different subseries based on the hour of the day. This segmentation is instrumental in capturing the distinct patterns in energy demand that vary with the time of day. Importantly, this division aligns with the natural time structure of the data, obviating the need for additional aggregation or disaggregation of predictions. Essentially, predictions are already organized into the same subseries defined by this hierarchical division. However, this hierarchical organization proves beneficial for our modeling approach, enabling the application of triplet losses to subseries that should exhibit dissimilar behavior.

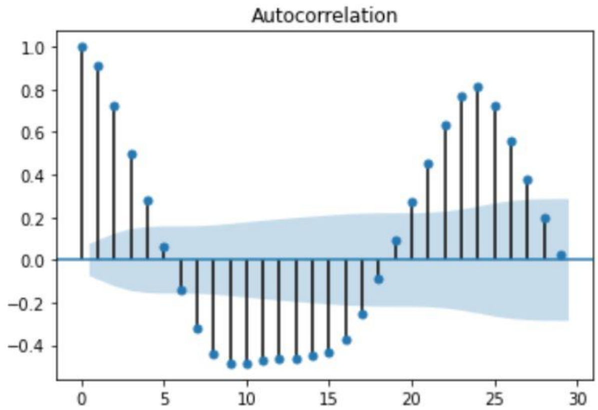


Figure 4. Autocorrelation plot of the time series by lags (x-axis)

Moving beyond time-related properties, we establish a hierarchical structure based on zone-related properties, allowing for both aggregation and disaggregation tasks. This hierarchy takes into account regional distinctions, an essential consideration in energy demand forecasting. In our forecasting strategy, we employ both top-down and bottom-up methods within the defined hierarchy, with each approach tailored to its specific context. In the case of bottom-up forecasting, we start by forecasting at the most granular and specific level within the hierarchy. This involves predicting energy demand for individual locations or regions. Subsequently, these forecasts are aggregated upwards to obtain estimates for higher-level regions or the total energy demand. This approach is particularly relevant when we aim to generate forecasts for specific locations and then consolidate them into regional or national energy demand predictions [22].

The centerpiece of our feature set is the trend, which is incorporated into the GluonTS models as the dynamic real features field. The trend is a critical component of our forecasting approach, and its calculation is detailed in Algorithm 1. The algorithm takes several features as input and produces a set of outputs. Size of the window (w): This is a numerical value that defines the size of the window for which the rolling mean is calculated. It represents the number of data points considered in each segment of the data for trend analysis. The intervals of the Trend (t): This is a numerical value that specifies the interval over which the trend is to be calculated. It determines the frequency at which the trend slope is recalculated and applied to the data. Trend's feature column (trend): The output of the algorithm is a new column added to the input data frame. This column contains the trend values calculated based on the rolling mean over the window period (w) and recalculated at every specified trend interval (t). The trend values are determined by the slope of the trend line fitted to the data points within each window, providing a dynamic representation of the trend over the entire dataset.

Actually, the top-down forecasting approach begins at the highest level of the hierarchical structure. Initial forecasts are made for the overall energy demand at the highest level, typically using historical proportions or other relevant factors. These forecasts are then distributed and disaggregated to obtain estimates for lower-level regions or locations. This method is suitable when the primary focus is on generating high-level forecasts, and subsequently breaking them down into more granular predictions. The choice between top-down and bottom-up forecasting within the hierarchical time series framework depends on the specific forecasting objectives, the granularity of predictions required, and the availability of data at different levels of the hierarchy. Leveraging both approaches strategically enables us to harness the full potential of hierarchical time series methods for energy demand forecasting.

Algorithm 1. For trend calculation

```
1. Rolling Mean Calculation: For each window of size w, calculate RollingMean(w).
2. Initialization:
   • a=[1,2,…,t]
   • trend[], tempTrend[]
   • counterTrend=0
3. For each row i in data:
   • If counterTrend=t:
      •  Reset counterTrend=1
      •  Calculate slope using a (x-values) and tempTrend (y-values)
      •  Extend trend with slope repeated t times
   • Else:
      •  Append rolling mean of row i to tempTrend
      •  Increment counterTrend
4. Finalize Trend Values: Add last set of trend values to trend
```

2.3. Neural network models

In the field of sequence data modeling for energy demand forecasting, the choice of architecture plays a pivotal role in capturing the temporal dependencies and patterns within the data. One of the primary methods implemented in this study is the LSTM model. The selection of this model over traditional RNNs stems from several critical considerations.

RNNs, while effective for certain tasks, suffer from inherent limitations, most notably their short memory. When confronted with lengthy sequences, it struggles to retain information from earlier time steps to later ones. This characteristic poses a substantial challenge for energy demand forecasting, where historical context and dependencies are vital. In practice, RNNs often end up omitting essential information from the initial portions of a sequence, leading to suboptimal predictions. Moreover, they are susceptible to the vanishing gradient problem, which hampers their ability to capture long-range dependencies effectively. As a result, when dealing with complex time series data with intricate patterns and dependencies, that may falter in providing accurate forecasts.

In contrast, LSTM models offer a compelling solution to these challenges. They are specifically designed to address the short memory limitation of RNNs, because the model is equipped with mechanisms such as input gates, forget gates, and output gates that allow them to selectively store and retrieve information over longer sequences. This grants it the capability to capture essential temporal dependencies, making them well-suited for energy demand forecasting tasks. Additionally, LSTMs exhibit robustness in the face of noisy data, a common characteristic of real-world energy consumption patterns. They can effectively handle distributed representations and continuous values, accommodating the diverse nature of energy data. Moreover, in theory, they have the capacity to deal with an unlimited number of states, offering scalability for varying complexities of energy systems. These models are both spatially and temporally local, enabling them to focus on relevant patterns within the data efficiently.

While this study primarily focuses on LSTM-based models, it's worth noting that convolutional neural networks (CNNs) [23] also present a viable option for sequence data analysis, although they were not explored in this research. They excel in spatial feature extraction from grid-like data, making them particularly effective for image recognition tasks. In the context of energy demand forecasting, if the data were organized in a grid-like or spatially structured format, they could be a compelling choice for capturing spatial dependencies. However, since energy demand data is typically presenting as temporal sequences, the inherent strength of the models in spatial analysis may not be fully leveraged for this specific task. LSTMs overcome the short memory limitations, effectively handle noisy data, and exhibit adaptability to varying hyperparameters. While CNNs are a powerful tool in spatial data analysis, the temporal nature of energy demand data makes LSTM models a more suitable choice for this particular forecasting task.

Continuing our exploration of the models, we have implemented various configurations to assess their effectiveness in capturing intricate temporal dependencies within the data. These configurations include a vanilla model, stacked, and bidirectional LSTM models [24]. The vanilla LSTM model serves as our baseline, providing a foundation for understanding the core performance of the architecture. Stacked LSTMs, on the other hand, introduce a layered approach, wherein multiple neuron layers are sequentially stacked. This architecture enables the model to capture hierarchical patterns and dependencies within the time series data more comprehensively. Each subsequent layer refines the representation of the data, allowing for a more nuanced understanding of the underlying patterns. Incorporating bidirectional LSTMs, as shown in Figure 5, into our model further extends our capacity to capture temporal dependencies. They possess the unique ability to learn from the input sequence both forwards and backwards. This bidirectional learning approach enhances the model's ability to recognize and adapt to complex temporal patterns, thereby contributing to more accurate energy demand forecasts [25].

```
Epoch 8/10
988/988 [==============================] - 0s 73us/step - loss: 107348.6276 - val_loss: 108042.3497
Epoch 9/10
988/988 [==============================] - 0s 74us/step - loss: 93654.7557 - val_loss: 83137.7218
Epoch 10/10
988/988 [==============================] - 0s 73us/step - loss: 85627.2178 - val_loss: 76682.5028
[[80.12668]]
```

Figure 5. Training bidirectional LSTM

A notable innovation introduced in this study is the use of triplet losses adapted for time series data. This approach is inspired by the word2vec word representation learning method [26], where the goal is to linearly separate pairs of context and word instances. The underlying intuition is that the context of a word should be distinguishable from the context of a randomly selected word. In the context of time series data, triplet losses offer a valuable technique for performing negative sampling, thereby enhancing the model's ability to discriminate between relevant and irrelevant information within the temporal sequences. To augment the LSTM models, we introduced covariates into the modeling process, specifically focusing on lagged covariates. These covariates are derived from lag features extracted from the time series data. Lagged covariates play a crucial role in enhancing the model's explanatory power by providing additional time-related information. This supplementary information aids in unraveling complex relationships within the energy demand dataset, ultimately contributing to more refined and accurate forecasts.

By systematically integrating these advanced LSTM configurations and leveraging triplet losses, we aim to provide a comprehensive analysis of the model's performance and its capacity to capture nuanced temporal dependencies within the energy demand data. This multi-faceted approach empowers us to explore the full potential of LSTM-based modeling in energy demand forecasting, offering a deeper understanding of the intricacies involved in predicting energy consumption patterns as shown in Table 1.

Table 1. The performance of LSTM training

Epoch	Loss	Validation Loss
0	781457.4899	675370.7406
1	662117.2315	552152.1663
2	543855.9997	437261.3691
3	431340.4677	327326.3449
4	316356.7516	212985.2282
5	179729.5179	151361.4872
6	126581.8124	95367.7286
7	107348.6276	108042.3497
8	93654.7557	83137.7218
9	85627.2178	76682.5028

In the approach, we're working with subseries (x^{ref}, x^{pos}, x^{neg}), which can be thought of as meaningful segments or subsets of our time series data (y_i, y_j). This subseries is crucial because they help us break down the complex time series into more manageable and interpretable parts. Essentially, subseries is like smaller pieces of the puzzle that, when put together, provide a comprehensive picture of the entire time series. The idea is that one type should exhibit some similarities to another type, while also displaying distinct differences from a third type. This approach allows us to capture the nuances and patterns within the time series more effectively [27]. To train our model, we use a mathematical formula that guides it in distinguishing between these different subseries. In our code, we have simplified this formula to make it more practical for implementation. The objective to minimize is

$$-\log\left(\sigma\left(\mathrm{f}(x^{ref},\theta)^{\top}\mathrm{f}(x^{pos},\theta)\right)\right)-\sum_{k=1}^{K}\log\left(\sigma\left(-\mathrm{f}(x^{ref},\theta)^{\top}\mathrm{f}\left(x_k^{neg},\theta\right)\right)\right) \tag{1}$$

where σ is a sigmoid function. Based on the loss, we use computed representations to distinguish x^{ref} and x^{neg} and to compare x^{ref} and x^{pos}. But, in the code we use the simplified triplet loss:

$$L=\sum_i \max\left(\| x_i^{\mathrm{pos}} - \mathrm{pred}_i\|_2^2 - \|x_i^{\mathrm{neg}} - \mathrm{pred}_i\|_2^2 + \mathrm{margin}, 0\right) \tag{2}$$

When it comes to selecting this subseries, we start with our time series data and randomly determine their sizes. One subseries (x^{ref}, x^{neg}) is then chosen randomly from the data, while another (x^{pos}) is selected from within the first subseries. We also involve a different set of data and select a subseries from it, making sure it's different from the first one. This process helps our model learn how to differentiate between these subseries effectively. In addition, we take the time of day into account, which is an essential temporal aspect of time series data. We divide our data into different time categories, such as morning, afternoon, or evening. Then, we choose one subseries from the same time category as our current data and another subseries from a different time category.

This approach helps our model recognize patterns related to different times of the day, which can be crucial for accurate forecasting. Also, in our triplet loss model, we have introduced a change. Instead of having just two inputs, we now have three. These inputs represent relationships between the subseries, including past data and future targets. For actually selecting this subseries during the training process, we follow a specific method. We randomly pick one subseries from another subseries, ensuring they have the same fixed length. We also randomly select a subseries from our data, excluding the first one. This approach fine-tunes our model's ability to work effectively with this subseries during training, helping it make more accurate predictions based on the complex time series data.

3. RESULTS AND DISCUSSION

In this study, a significant portion of the experimental work leveraged the capabilities of two prominent libraries: GluonTS [28] and MXNet, in addition to the use of the Keras library with a dependency on TensorFlow. These libraries provided the necessary tools and frameworks to develop and evaluate deep learning-based time series forecasting models. GluonTS, a crucial component of this research, is a specialized library designed for deep learning-based time series modeling. Its suitability for this work arises from its proficiency in developing deep learning forecasting models, particularly probabilistic models. GluonTS encompasses a comprehensive set of forecasting methods, categorizing them into generative and discriminative models.

One of the key advantages of GluonTS is its foundation on the MXNet library. MXNet, a robust deep learning library, plays a foundational role in enabling GluonTS to perform its tasks effectively. Within MXNet, the "Gluon" API proves invaluable for defining and training neural networks, providing a flexible and efficient platform for deep learning experimentation.

The experimental process involved several phases. First, we extensively explored the capabilities of GluonTS and MXNet, utilizing their built-in forecasting methods for both generative and discriminative models. We configured these models to operate on our time series data, incorporating various features and covariates to enhance their forecasting accuracy. Furthermore, we implemented LSTM-based models using the Keras library, taking advantage of TensorFlow as the underlying framework. These models were designed to accommodate the unique characteristics of our time series data and covariates. By experimenting with different architectures, including simple, stacked, and bidirectional LSTM, we assessed their suitability for our forecasting tasks as shown in Figure 6.

The lag value identified serves a dual purpose in our modeling approach. In the LSTM models, an analysis of mean squared error (MSE) revealed a normalized error close to 0.32, indicating its significant utility. Furthermore, this lag value finds application in GluonTS models as crucial lag features. Specifically, it contributes to the lags sequence input for the DeepAR estimator [29], signifying the indices of lagged target values to be incorporated into the RNN, thus amplifying its relevance and versatility in both modeling paradigms.

Further, we used GluonTS feedforward models, incorporating L1, L2, and Huber losses into the framework. Additionally, we employed the DeepAR Estimator, an auto-regressive model, using both LSTM and GRU cells for enhanced predictive capabilities. These models are designed to process input windows of a specified context length and generate subsequent windows of dimension prediction length, aligning with the output dimension. L1 loss, also known as mean absolute error, is robust to outliers and provides a more stable error measure for anomalies. L2 loss, or mean squared error, tends to be sensitive to outliers but is beneficial in emphasizing larger errors, which is crucial for certain time series applications. The Huber loss, a combination of L1 and L2, offers a balanced approach, reducing the sensitivity to outliers like L1 while maintaining the ability to penalize large errors like L2. The integration of these loss functions into GluonTS's feedforward models allows for more flexible and fine-tuned model training, catering to the specific needs of different time series datasets as shown in Figure 7.

To augment our analysis, we go beyond the default hybrid forward method of the training network, which typically returns the meaning of the L1 loss. Instead, we explored alternative loss functions, including L2, Huber, and triplet losses, to tailor our approach to specific objectives. Further enhancing our models' accuracy, we incorporated normalization and reverse normalization techniques, ensuring that our results remained both reliable and interpretable. Table 2 presents the average scores attained by these models.

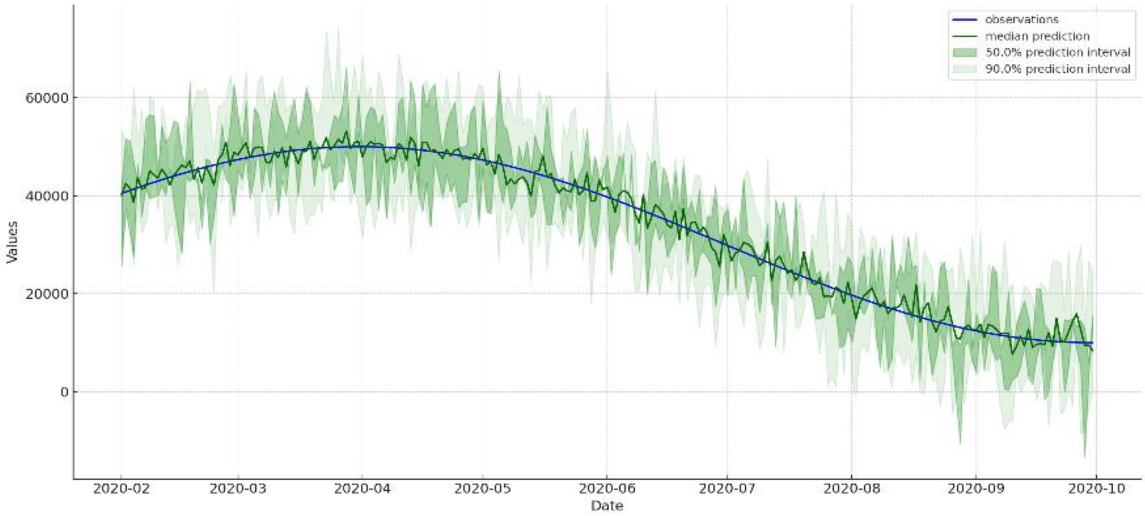


Figure 6. Initial energy demand prediction for California

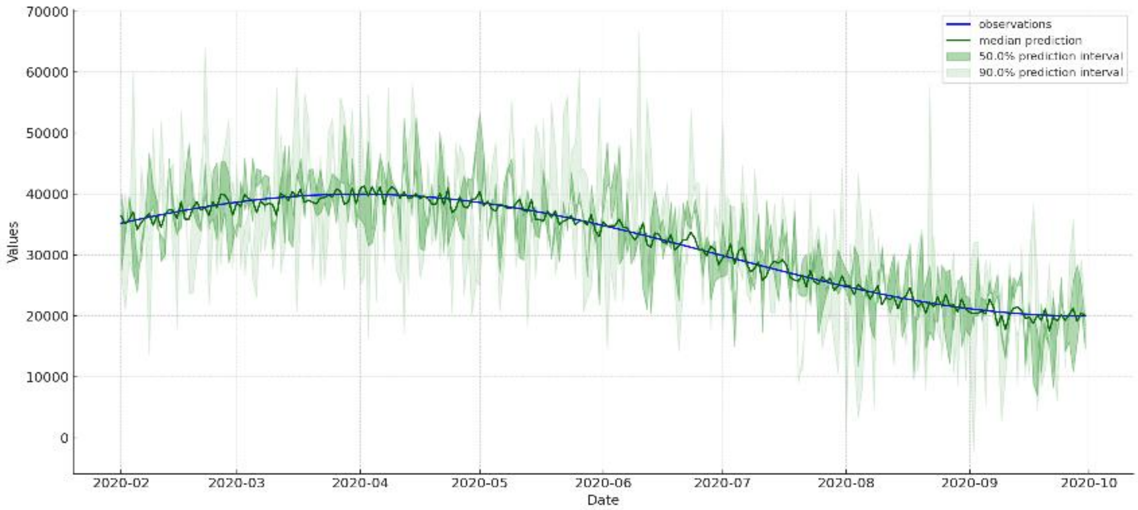


Figure 7. Energy demand prediction using GluonTS model

Table 2. The performance of GluonTS model

GluonTS model	MSE
Huber loss	0.171
L1 loss	0.238
DeepAR Estimator	0.249

During our analysis, we observed that the plot did not display the 50 or 90 percent confidence intervals of the triplet loss model, sparking our curiosity. We further investigated the model's outputs and examined accuracy across various prediction lengths using the DeepAR estimator framework. The results, as shown in the table, revealed a significant reduction in error when the prediction length was set at 10, while error levels remained stable between prediction lengths of 20 to 40. Surprisingly, the error increased as the prediction length reached 50. Table 3 offer insights into the model's performance and emphasize the importance of selecting an optimal prediction length to enhance forecasting accuracy.

Table 3. Absolute errors to prediction length

Prediction length	Absolute error
10	0.091
20	1.622
30	1.542
40	1.601
50	15.13

We implemented a straightforward feedforward architecture, replacing the loss in the hybrid forward approach with the triplet loss [30]. So, we conducted experiments to examine the impact of varying the margin (an input parameter for the triplet loss function) and prediction length on the model's performance. One notable approach involved designating the morning as the positive series and utilizing the test series from this subseries for predictions, limiting the prediction length to 6 to align with the six morning hours. Surprisingly, shorter prediction lengths yielded less accuracy, and it was observed that a margin closer to 0 resulted in lower accuracy compared to a margin around 10. The triplet loss is a fundamental component of the model, serving as a critical metric for evaluating its performance. Essentially, it measures the similarity between three input elements: an anchor, a positive sample, and a negative sample. In the context of this study, the anchor represents a specific subseries of time-series data, while the positive sample is another subseries from the same category (e.g., morning subseries). The negative sample, on the other hand, is a subseries from a different category (e.g., afternoon or evening subseries).

Figure 8 presents a time series from 13:00 to 18:00, where the median prediction (green line) deviates from a straight line, indicating a non-linear forecast that decreases initially, then slightly increases towards the end of the period. The triplet loss aims to minimize the distance between the anchor and the positive sample while maximizing the distance between the anchor and the negative sample. This objective encourages the model to distinguish between different subseries accurately, which is crucial for robust energy demand forecasting. By optimizing the margin and prediction length parameters in the triplet loss, we can enhance the model's ability to capture temporal patterns and improve its overall accuracy in making predictions. Table 4 highlight the intricate relationship between prediction length, margin values, and forecasting accuracy.

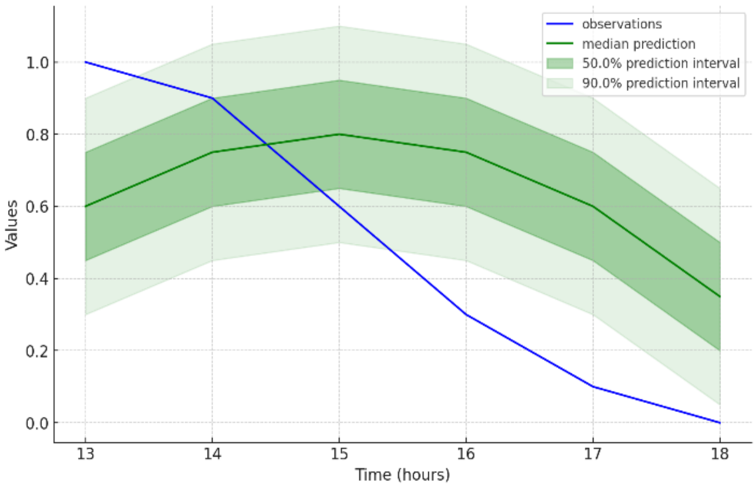


Figure 8. Applying triplet loss on afternoon time series

Table 4. The optimization results

x^{pos}	Margin	Prediction length	MSE	Absolute error
evening(t)	10	20	0.211	8.338
evening(t)	10	30	0.181	11.423
morning(t)	10	6	0.524	3.846
afternoon(t)	10	6	0.296	2.688
afternoon(t)	10	20	0.218	7.819
early morning(t)	10	20	0.271	10.436
early morning(t)	1	20	0.463	13.447
early morning(t)	3	10	0.385	6.129

4. CONCLUSION

In this study, our focus has been on leveraging representation learning to enhance the accuracy of energy demand forecasting. We accomplished this by employing feature extraction techniques and incorporating both time-related and region-related features into our models. Extensive experimentation with the GluonTS library allowed us to fine-tune our models and adapt them to adhere to fundamental theoretical principles. Notably, we observed that triplet losses yielded impressive results, especially when using larger margin sizes and extended prediction lengths, signifying their accuracy and potential for application in real-world scenarios.

Looking ahead, our research opens the door to several promising directions. One key direction involves extending the implementation of probabilistic models beyond synthetic datasets to real-world datasets, providing a more comprehensive evaluation of their performance. Additionally, we plan to explore the use of alternative time-related hierarchies, such as distinguishing between business hours, weekends, and different seasons, to create positive and negative subseries samples for the triplet loss model. Furthermore, we aim to integrate additional representation learning and sequence-based learning techniques, such as autoencoders and manifold learning, into the realm of time series forecasting. These techniques can provide a deeper understanding of time series data, enabling the extraction of more complex features and relationships, which in turn could significantly enhance the forecasting accuracy and contribute to the advancement of time series analysis.

REFERENCES

[1] W. Wu and M. Peng, "A data mining approach combining K-means clustering with bagging neural network for short-term wind power forecasting," *IEEE Internet of Things Journal*, vol. 4, no. 4, pp. 979–986, Aug. 2017, doi: 10.1109/JIOT.2017.2677578.

[2] X. Zhao, J. Liu, D. Yu, and J. Chang, "One-day-ahead probabilistic wind speed forecast based on optimized numerical weather prediction data," *Energy Conversion and Management*, vol. 164, pp. 560–569, May 2018, doi: 10.1016/j.enconman.2018.03.030.

[3] Aasim, S. N. Singh, and A. Mohapatra, "Repeated wavelet transform based ARIMA model for very short-term wind speed forecasting," *Renewable Energy*, vol. 136, pp. 758–768, Jun. 2019, doi: 10.1016/j.renene.2019.01.031.

[4] D. Yang, "On post-processing day-ahead NWP forecasts using Kalman filtering," *Solar Energy*, vol. 182, pp. 179–181, Apr. 2019, doi: 10.1016/j.solener.2019.02.044.

[5] Y. Wang, J. Wang, and X. Wei, "A hybrid wind speed forecasting model based on phase space reconstruction theory and Markov model: a case study of wind farms in northwest China," *Energy*, vol. 91, pp. 556–572, Nov. 2015, doi: 10.1016/j.energy.2015.08.039.

[6] L. Wu, X. Gao, Y. Xiao, Y. Yang, and X. Chen, "Using a novel multi-variable grey model to forecast the electricity consumption of Shandong Province in China," *Energy*, vol. 157, pp. 327–335, Aug. 2018, doi: 10.1016/j.energy.2018.05.147.

[7] T. Samardzioska, V. Zileska Pancovska, S. Petrusheva, and B. Sekovska, "Prediction of energy consumption in buildings using support vector machine," *Tehnicki Vjesnik*, vol. 28, no. 2, pp. 649–656, 2021, doi: 10.17559/TV-20190822153751.

[8] J. Ruiz-Rangel, C. J. Ardila Hernandez, L. M. Gonzalez, and D. Jabba Molinares, "ERNEAD: Training of artificial neural networks based on a genetic algorithm and finite automata theory," *International Journal of Artificial Intelligence (IJ-AI)*, vol. 16, no. 1, pp. 214–253, 2018.

[9] C. Lu, S. Li, and Z. Lu, "Building energy prediction using artificial neural networks: A literature survey," *Energy and Buildings*, vol. 262, May 2022, doi: 10.1016/j.enbuild.2021.111718.

[10] F. Mahia, A. R. Dey, M. A. Masud, and M. S. Mahmud, "Forecasting electricity consumption using ARIMA model," in *2019 International Conference on Sustainable Technologies for Industry 4.0 (STI)*, Dec. 2019, pp. 1–6, doi: 10.1109/STI47673.2019.9068076.

[11] H. Cai, S. Shen, Q. Lin, X. Li, and H. Xiao, "Predicting the energy consumption of residential buildings for regional electricity supply-side and demand-side management," *IEEE Access*, vol. 7, pp. 30386–30397, 2019, doi: 10.1109/access.2019.2901257.

[12] R. Wang, X. Xia, Y. Li, and W. Cao, "Clifford fuzzy support vector machine for regression and its application in electric load forecasting of energy system," *Frontiers in Energy Research*, vol. 9, Dec. 2021, doi: 10.3389/fenrg.2021.793078.

[13] R. F. K. Basha, M. L. Bharathi, and K. Venusamy, "Dynamic prediction of energy and power usage cost using linear regression-machine learning analysis," *Journal of Physics: Conference Series*, vol. 1921, no. 1, May 2021, doi: 10.1088/1742-6596/1921/1/012067.

[14] D. Lee, S. Kang, and J. Shin, "Using deep learning techniques to forecast environmental consumption level," *Sustainability*, vol. 9, no. 10, Oct. 2017, doi: 10.3390/su9101894.

[15] C. Li, Z. Ding, D. Zhao, J. Yi, and G. Zhang, "Building energy consumption prediction: an extreme deep learning approach," *Energies*, vol. 10, no. 10, Oct. 2017, doi: 10.3390/en10101525.

[16] J. Palet, V. Manquinho, and R. Henriques, "Multiple-input neural networks for time series forecasting incorporating historical and prospective context," *Data Mining and Knowledge Discovery*, vol. 38, no. 1, pp. 315–341, Oct. 2023, doi: 10.1007/s10618-023-00984-y.

[17] J.-Y. Franceschi, A. Dieuleveut, and M. Jaggi, "Unsupervised scalable representation learning for multivariate time series," *arXiv:1901.10738*, Jan. 2019.

[18] "Pecan Street Dataport," Pecan Street, https://www.pecanstreet.org/dataport/ (accessed Feb. 05, 2024).

[19] A. Desiani, S. Yahdin, A. Kartikasari, and I. Irmeilyana, "Handling the imbalanced data with missing value elimination SMOTE in the classification of the relevance education background with graduates employment," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 10, no. 2, pp. 346-354, Jun. 2021, doi: 10.11591/ijai.v10.i2.pp346-354.

[20] H. Omara, M. Lazaar, and Y. Tabii, "Effect of feature selection on gene expression datasets classification accuracy," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 8, no. 5, pp. 3194–3203, 2018, doi: 10.11591/ijece.v8i5.pp.3194-3203.

[21] Z. Niu, J. Wu, X. Liu, L. Huang, and P. S. Nielsen, "Understanding energy demand behaviors through spatio-temporal smart meter data analysis," *Energy*, vol. 226, Jul. 2021, doi: 10.1016/j.energy.2021.120493.

[22] M. T. Sarker, M. Jaber Alam, G. Ramasamy, and M. Nasir Uddin, "Energy demand forecasting of remote areas using linear regression and inverse matrix analysis," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 14, no. 1, pp. 129–139, Feb. 2024, doi: 10.11591/ijece.v14i1.pp129-139.
[23] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, "Convolutional neural networks: an overview and application in radiology," *Insights into Imaging*, vol. 9, no. 4, pp. 611–629, Jun. 2018, doi: 10.1007/s13244-018-0639-9.
[24] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.
[25] N. Mounir, H. Ouadi, and I. Jrhilifa, "Short-term electric load forecasting using an EMD-BI-LSTM approach for smart grid energy management system," *Energy and Buildings*, vol. 288, Jun. 2023, doi: 10.1016/j.enbuild.2023.113022.
[26] U. Naseem, I. Razzak, S. K. Khan, and M. Prasad, "A comprehensive survey on word representation models: from classical to state-of-the-art word representation language models," *ACM Transactions on Asian and Low-Resource Language Information Processing*, vol. 20, no. 5, pp. 1–35, Jun. 2021, doi: 10.1145/3434237.
[27] D. Janka, F. Lenders, S. Wang, A. Cohen, and N. Li, "Detecting and locating patterns in time series using machine learning," *Control Engineering Practice*, vol. 93, Dec. 2019, doi: 10.1016/j.conengprac.2019.104169.
[28] A. Alexandrov *et al.*, "GluonTS: probabilistic time series models in Python," *arxiv.org/abs/1906.05264*, Jun. 2019.
[29] D. Salinas, V. Flunkert, J. Gasthaus, and T. Januschowski, "DeepAR: probabilistic forecasting with autoregressive recurrent networks," *International Journal of Forecasting*, vol. 36, no. 3, pp. 1181–1191, Jul. 2020, doi: 10.1016/j.ijforecast.2019.07.001.
[30] Y. Yuan, W. Chen, Y. Yang, and Z. Wang, "In defense of the triplet loss again: learning robust person re-identification with fast approximated triplet loss and label distillation," Jun. 2020, doi: 10.1109/cvprw50498.2020.00185.

BIOGRAPHIES OF AUTHORS

Fariza Nussipova is a developer with a rich background in programming. She holds a bachelor's degree in information systems from the International Information Technologies University and is currently pursuing a master's in software engineering at Kazakh-British Technical University. Fariza has a strong track record in developing engaging apps, leveraging her skills in HTML, CSS, JavaScript, and Python. Her professional journey includes significant contributions at Jusan Bank, where she played a key role in developing their marketplace and other critical projects. Prior to that, she enhanced operational efficiency at DevArt and honed her mentoring skills at Morrison code, guiding young developers. With a passion for coding and a commitment to advancing web development, Fariza is an asset to any project aiming to create innovative and efficient solutions. She can be contacted at email: f_nussipova@kbtu.kz.

Shynggys Rysbekov is a talented video editor with extensive experience in both pre and post commercial video production. He earned his bachelor's degree in information technologies from the International University of Information Technologies and is currently pursuing a master's degree in the same field at KBTU Business School, with an expected completion in 2023. Shynggys has demonstrated his expertise during his tenure at Oberfilm in Almaty, Kazakhstan, from January 2018 to October 2020, where he was instrumental in producing commercial videos. His educational background and professional experience equip him with a deep understanding of the technical and creative aspects of video editing, making him a valuable asset to projects aiming to captivate and engage audiences through high-quality media content. He can be contacted at email: s_rysbekov@kbtu.kz.

Zukhra Abdiakhmetova is an associate professor and Vice Dean at the Faculty of Information Technology, Al-Farabi Kazakh National University, specializing in computer science. She has a remarkable academic background, having earned all her degrees in computer science from Al-Farabi Kazakh National University. Her research spans various aspects of machine learning and wavelet transform, notably in medical applications like ECG analysis for heart fibrillation prediction. She has also worked on projects such as homomorphic encryption on microcontrollers and machine learning for heart disease study. Dr. Abdiakhmetova's academic and research pursuits demonstrate a deep engagement with applying complex computational techniques to practical and significant real-world problems. She can be contacted at zuhra.abdyahmetova@kaznu.edu.kz.

Amandyk Kartbayev is an associate professor at Kazakh-British Technical University, Almaty, Kazakhstan. He obtained his bachelor's degree in Math from Satbayev University, Master of Science, and Ph.D. in information systems from Al-Farabi Kazakh National University. Kartbayev has also worked as a Senior IT Engineer at KMG Engineering and had a research internship at Fondazione Bruno Kessler in Italy. He holds several certifications in areas like mathematics, machine learning, and Microsoft technologies. Kartbayev's research contributions include papers on machine translation and data models, reflecting his expertise in computational linguistics and data engineering. He can be contacted at a.kartbayev@gmail.com.