

**WEEKLY REPORT DISSERTATION A39 PROJECT**

**RETAIL TIME SERIES FORECASTING WITH AUGMENTATION USING META-LEARNING**

**Author: Aathira Ravikumar**

**Course: MSc Business Analytics and Finance**

**CORMSIS Academic Led Project Code: A39**

**A39 DISSERTATION REPORT 2: RETAIL TIME SERIES FORECASTING WITH AUGMENTATION USING META-LEARNING**

**Brief Summary:**

The objective set was to understand the topic thoroughly and to get an understanding of how forecasting of retail time series, being a complex task to predict sales of multiple products across numerous stores, and to understand the objective of multivariate data augmentation technique for improvement in accuracy as the previous state-of-the-art was limited to univariate data which needs significant improvement in case of inclusion of data with multiple variables such as retail time series data. (Yang and Desell, n.d.)

The goal of this project is to implement retail time series data using multivariate data augmentation with a meta-learning framework built on deep convolutional neural networks. The initial goal is to create a robust literature review that will serve as a solid basis for the methodology and provide information on previous research in the topic, which will help guide the development of a fresh strategy or insight for this dissertation.

**Introduction:**

When making estimates for retail sales, it's common to have a short forecasting horizon and a lot of products spread throughout a lot of outlets. Sales projections are crucial components in many administrative choices, including those regarding an item's pricing, distribution of retail space, listing and delivery, ordering, and inventory control. Retailers may model the effects of their various promotional mixes using a competent sales forecasting system, which they can then use to optimise the promotional schedules. Many other factors, such as pricing adjustments, promotions, special events, seasons, holidays, and even weather, influence the demand for retail products. Because of this, store item level sales data exhibit high volatility, skewness, multiple seasonal cycles, especially when combined with "special days" (such as bank holidays), often high volume, alternately intermittent behaviour with frequent observations of zero sales at the store level, as well as high dimensionality in any explanatory variable space. These problems make it challenging to predict item level sales properly. (Ma and Fildes, 2021)

The main advantage of trained generators is that they can accurately mimic the patterns found in real-world data without the need for manual analysis, but their main disadvantage is that they can necessitate massive training data sets and lengthy training cycles in order to get the same results.

To practise simulating real data. Contrarily, while laborious to set up, Monte Carlo methods are easier and more practical to apply. They produce interpretable data and give users the ability to manipulate that data, add subject-matter expertise, and recreate certain scenarios. The drawback is that generated data may differ significantly from actual data. A hybrid strategy that combines both methodologies would therefore be useful for various downstream applications including prediction, forecasting, classification, and testing. The application of VAEs (Variational AutoEncoders). The suggested architecture possesses several unique qualities, including shortened training periods, interpretability, and domain knowledge encoding capacity. (Desai et al., 2021)

**Meta-Learning Framework:**



Figure 1: A meta-learning framework for retail sales forecasting

For the purpose of projecting retail product sales, we suggest a meta-learning architecture with automatic feature learning (see Figure. 1). Meta-learning and meta-forecasting are the two stages of the framework implementation.

In order to begin the meta-learning phase, we must first extract from the historical database a sizable subset of sales time series and the associated history of influential factors. Large retailers have amassed enormous volumes of historical data at the SKU (Stock Keeping Unit) level, but many SKUs have scant store-level sales history due to continuously shifting assortments. Therefore, we do not assume that the SKU sample in the training set and the test set are the same. In order to fit base-forecasters using the same amount of data during the meta-learning and meta-forecasting phases, we just assume that the SKUs in both sets are forecasted with the same width rolling window.

For each sales time series that needs to be forecasted, the same pool of base forecasters must first be fitted before we can make forecasts using the models that were fitted. The trained meta-learner is then given these forecasts and the unprocessed time series. For each sales time series being forecast, the trained meta-learner extracts the features from the time series, computes combination weights, and produces a set of ensemble forecasts.

**Variational AutoEncoder for Data Generation and Augmentation:**

The use of a variational autoencoder (VAE) for data enhancement and feature extraction in acoustic modelling is presented. A VAE is a deep learning framework-based generative model based on variational Bayesian learning. A VAE can produce new data by extracting latent values from its input variables. (Nishizaki, 2017)

2. Literature Review

2.1. Augmentation

Deep learning-based generative models have drawn increasing attention in recent years as a result of (and considering) some outstanding advancements in the area. Deep generative models have demonstrated an astonishing ability to make incredibly realistic bits of material of many kinds, such as images, words, and sounds, by relying on enormous amounts of data, well-designed network topologies, and clever training procedures. The Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs) families of deep generative models are two that stand out and merit special consideration. One of the major kinds of deep generative models is Variational Autoencoders (VAEs). A VAE is an autoencoder whose encodings distribution is regulated during training to guarantee that its latent space has favourable characteristics allowing us to produce some fresh data. Additionally, the word "variational" derives from the tight connection between the regularisation and variational inference methods in statistics. In machine learning, dimensionality reduction refers to the process of lowering the number of features in data, either by selecting already-existing features or by extracting new ones. A decoder, which reconstructs the original data from the encoded representation, and an encoder, which compresses the data into a lower-dimensional space (latent space), are used to carry out this procedure. Finding the optimal encoder/decoder pair from a given family with the highest information retention during encoding and the lowest reconstruction error during decoding is the aim of dimensionality reduction. Finding the best encoder and decoder to minimise the reconstruction error measure between the input data and the encoded-decoded data is how the dimensionality reduction problem is defined. The fundamental goal is to keep crucial data while lowering the dimensionality of the data.

A dimensionality-reduction technique called Principal Component Analysis (PCA) fits into the previously discussed paradigm. In order to minimise the approximation error in the projections of the data onto the subspace defined by these new features, it seeks to identify a collection of new independent features that are linear combinations of the original features. With orthonormal rows (representing independent features) and an encoder (n\_e by n\_d matrix) in the family E and a corresponding decoder (n\_d by n\_e matrix) in the family D, PCA looks for these objects in the global framework. The decoder is the transposition of the encoder, and the encoder corresponds to the unitary eigenvectors linked to the n\_e highest eigenvalues of the features' covariance matrix. As a result, the dimensionality reduction problem can be expressed as an eigenvalue/eigenvector problem, enabling the most useful characteristics to be extracted for data projection and reconstruction.

Neural network designs called autoencoders are employed for dimensionality reduction. The aim of them is to discover the best encoding-decoding scheme by iterative optimisation. They are composed of an encoder and a decoder, both represented by neural networks. The encoder creates a bottleneck while retaining the primary structured data by compressing the data into a lower-dimensional space (latent space). There is a connection between PCA and linear autoencoders with single-layer encoder and decoder since both of these algorithms look for the optimal linear subspace to project data on with the least amount of information loss. However, non-linear and deep autoencoders are capable of larger dimensionality reduction with minimal reconstruction loss. To balance dimensionality reduction and information retention, it's critical to regulate the size of the latent space and the depth of the autoencoders. An excessive reduction may lead to less comprehensible representations and the loss of crucial data structure information.

Variational Autoencoders (VAEs) address the limitation of standard autoencoders in content generation by introducing explicit regularization during the training process to ensure a more regular and structured latent space. VAEs consist of an encoder and a decoder, like standard autoencoders, but with a modification in the encoding-decoding process. Instead of encoding an input as a single point, VAEs encode it as a distribution over the latent space, typically chosen to be normal. The encoder is trained to output the mean and covariance matrix that describe these Gaussians. This change allows for the introduction of latent space regularization by enforcing the encoded distributions to be close to a standard normal distribution. The loss function for VAEs includes a "reconstruction term" for efficient encoding-decoding and a "regularization term" expressed as the Kullback-Leibler divergence between the returned distribution and a standard Gaussian. The regularization term ensures both local and global regularity of the latent space. VAEs overcome the lack of structure in the latent space, making them suitable for content generation tasks. (Rocca, 2020)

2.1.1. Related works:

There exists a large body of literature that discusses various applications of Variational AutoEncoder available for time series analysis and other forms of data. Unsupervised anomaly detection of time series using VAE technique has been published with VAE being neural network type that learns to encode input data into lower-dimensional latent space and decode it back to the original space. The suggested VAE in this study processes sequential data using a recurrent encoder and decoder and enhances the encoding-decoding process using a variational self-attention mechanism. Anomaly is recognised based on the probabilistic reconstruction scores provided by the model, which is trained on normal data and used to rebuild new data. (Pereira and Silveira, 2018)

For dimensionality reduction and data imputation in multivariate time series with missing values it is suggested that a new deep sequential latent variable model is applied. In order to achieve non-linear dimensionality reduction in the presence of missing data, the suggested model employs a VAE technique along with a unique structured variational approximation. On high-dimensional data from the fields of computer vision and healthcare, the authors show that their technique outperforms several traditional and deep learning-based data imputation methods by modelling the low-dimensional dynamics with a Gaussian process. The suggested method also gives interpretable uncertainty estimates and smoothest out the imputations (Fortuin, Gunnar Rätsch and Mandt, 2019).

To find anomalies in time series data, a hybrid VAE-LSTM model was proposed. A VAE module was used to create reliable local features over brief time periods, and on top of those features, LSTM modules was used to estimate long-term correlations in the series. On five actual time series datasets with anomalous events, the algorithm was assessed and contrasted with three others widely used time series anomaly detection algorithms. Precision, recall, and F1 score testing revealed that the suggested strategy performed better than the other three alternatives. For each dataset, a detection window length that is the same for all approaches was selected (Lin et al., 2020).

The research introduced a unique architecture for time-series data generation with Variational Auto-Encoders (VAEs). The architecture possessed several unique qualities, including shortened training periods, domain knowledge encoding capability, and interpretability. By comparing it to four multivariate datasets, the data production quality was assessed for similarity and predictability. The VAE approach and many cutting-edge data generating techniques were all subjected to measurements of the effect of data availability on generation quality. When performing next-step prediction tasks utilising generated data, the suggested architecture consistently performed as well as or better than state-of-the-art data generation methods. The VAE methodology accurately captured the temporal properties of the original data. The performance for next-step prediction using generated data was considerably enhanced by the de-noised data that was produced. To produce comprehensible results, the suggested architecture can include domain-specific time-patterns such polynomial trends and seasonalises. By adding temporal structures to the data production process in the decoder, the modelled data generation process was made interpretable. The settings for the RCGAN code were changed to take the sample size and dimension of the various data sets into consideration. Although Mogren (2016) provided the C-RNN-GAN source code, the method was not used in the trials because to the tasks' instability. The original Base Decoder could also be used as a residual branch in the decoder thanks to the Interpretable TimeVAE architecture (Desai et al., 2021).

To increase the likelihood of predictions that generalise effectively and provide cutting-edge performance, techniques for cross-validation, augmentation, and parameter adjustment were presented. In four successive time-series forecasting competitions, the strategies outlined generated the best results. This study represented a significant milestone in the development of time-series forecasting as a separate area of machine learning, one with best practises capable of rapidly and consistently creating robust models (Lainder and Wolfinger, 2022). Usage of deep learning models for the classification of respiratory sounds and the challenges posed by imbalanced datasets is also worked on (Saldanha et al., 2022).

To conclude, it should be noted that Variational AutoEncoders (VAEs) have shown to be an effective and adaptable tool for a variety of applications in time series analysis and other data domains. They have been utilised effectively for dimensionality reduction, data imputation, time series data production, and unsupervised anomaly detection in time series. VAEs have benefits such as the ability to handle missing data, capture temporal features, and provide interpretable uncertainty estimates. They are a useful strategy in the field of machine learning for time series data because they can build effective representations in a lower-dimensional latent space while keeping crucial information. Furthermore, VAEs perform better when combined with other methods like LSTM for anomaly detection jobs. Overall, VAEs have expanded the field of time series analysis, forecasting, opened new avenues, and improved the state-of-the-art.

2.2. Forecasting

For many firms, forecasting is a crucial task that must be completed in order to run effectively. Many millions of time series must now frequently be forecast, which is getting more and more common. Large-scale companies might be interested in estimating sales, costs, and demand for thousands of products across numerous locations, warehouses, and soon, for instance. Technology businesses like Google gather millions of time series data points every day, including web click logs, web search counts, queries, revenues, and the number of users for various services. By introducing a novel quick technique for model selection and time series forecasting, we hope to address some computing issues caused by the size of these jobs. Finding out what kinds of timeseries should be forecast using the various models that are available is a crucial part of our effort in understanding how this algorithm operates (Talagala, Hyndman and Athanasopoulos, 2023). Time series characteristics can be used to find forecasting models that work well. We provide an all-encompassing system called FFORMS (Feature-based FORecast Model Selection), which chooses forecast models based on features computed from each time series. The FFORMS framework creates a mapping using a classification method like a random forest that links the characteristics of a time series to the "best" forecast model. Utilising time series from the M-forecasting contests, the framework is demonstrated to produce forecasts that are nearly as accurate as cutting-edge techniques but take significantly less time to compute. We investigate the findings and investigate which kinds of time series are most appropriate for each forecasting model using model-agnostic machine learning interpretability methodologies (Athanasopoulos, 2023).

2.2.1. Related works:

The existing study in the fields of meta-learning framework-based forecasting is analysed in this section. A unique meta-learning framework for multivariate time-series forecasting in the context of load forecasting is the main emphasis of the study, which also includes research on meta-learning systems and load forecasting models. The study presents a meta-learning system created to make it easier to choose a load forecasting model. The meta-learning system will be configured with a variety of tasks for the experiment, and its performance will be compared to that of other well-known load forecasting techniques. The outcomes show that in a repeated real-world simulation, the proposed meta-learning system beats ten popular forecasting algorithms across four load forecasting tasks. To improve the meta-learning framework's prediction powers, the study introduces new meta features such as fickleness, travesty, granularity, and highest ACF. The meta-learning framework, which is notable for being component-based, parallelized, and easily extensible, is a promising method for effective and precise load forecasting in multivariate time-series data. This review of the literature highlights the developments in meta-learning methods and their use in the field of load forecasting, emphasising how the suggested meta-learning system may enhance model selection and forecasting accuracy (Matijaš, Suykens and Krajcar, 2013).

A meta-learning approach for forecasting retail sales is also introduced in the studies, notably using deep convolutional neural networks (CNNs). This framework integrates a variety of base-forecasting techniques to achieve improved forecasting performance when compared to industry benchmarks. It is designed to automatically develop a feature representation from raw sales time series data. Even while it is successful at accurately forecasting, the learned features are not interpretable, which may restrict its practical application in particular situations. The study proposes building a diversified pool of base-forecasters, including both individual and pooled forecasting methods, and concentrating on determining the greatest combination forecasts rather than the best individual approach in order to address this restriction. The average mean squared error of the M base forecasters serves as the loss function in this architecture, giving an objective metric for improving forecasting ability. This meta-learning methodology shows promising possibilities for improving retail sales forecasting accuracy and producing more reliable forecasts by utilising deep CNNs and a range of forecasting techniques. To make the framework more useful for applications in real-world retail forecasting, additional study is required to enhance the interpretability of the learned features. With a focus on merging numerous base-forecasters to increase forecasting performance and adaptation to different retail time series data, this literature survey demonstrates the progress made in using meta-learning approaches, notably deep CNNs, for retail sales forecasting (Ma and Fildes, 2021).

The study introduces the FFORMS (Feature-based FORecast Model Selection) framework, which focuses on choosing forecast models based on features generated from each time series of data. This continues the literature survey. Using a classification technique like a random forest, this method creates a mapping that defines the relationship between the extracted time series data and the best forecast model. Using time series data from the M-forecasting contests, the FFORMS framework is thoroughly tested and shows that it can produce forecasts that are almost as accurate as cutting-edge techniques but with noticeably shorter calculation times. The top three approaches from the M4 competition are compared with the outcomes in several benchmark comparisons. To evaluate the forecast accuracy and uncertainty, point forecasts and prediction intervals are evaluated using the MASE and MSIS metrics, respectively. In order to understand the results and determine which kinds of time series are most appropriate for each forecasting model chosen by the FFORMS framework, the research also makes use of model-agnostic machine learning interpretability approaches. Overall, the FFORMS framework offers an innovative and effective method for choosing a prediction model based on time series features, producing forecasts that are extremely accurate while taking less time to compute. Additionally, by utilising interpretability methodologies, the framework's decision-making process is better understood, and its applicability to actual forecasting jobs is improved. This review of the literature emphasises the significance of feature-based model selection and the FFORMS framework's potential influence on time series forecasting by providing benefits for both accuracy and computing efficiency (Talagala, Hyndman and Athanasopoulos, 2023).

A novel two-step meta-learning strategy for time-series forecasting ensemble is suggested by the authors of a different study. Using two different random forest regression models, the method ranks 22 univariate forecasting techniques and dynamically suggests the right ensemble size. The suggested method is assessed using a dataset with 12,561 microeconomic time-series and contrasted with industry standard approaches, Theta and Comb. The outcomes show that the Theta and Comb approaches are outperformed by both the Simple and Weighted versions of the meta-learning methodology. For more than half (10 out of 16) of the forecasting horizons, the Weighted variation performs somewhat better than the Simple variant and also produces better overall performance. The study finds a consistent and expected pattern in time-series forecasting: forecasting errors tend to rise as forecasting horizons grow longer. The two-step meta-learning strategy, which selects and combines various univariate forecasting techniques based on their rankings, has promising promise for improving predicting accuracy. The outcomes shed important light on how well the suggested approach may be customised and applied to ensemble forecasting problems. Overall, by outperforming established benchmark methodologies, this study makes a contribution to the field of time-series forecasting by outlining a useful and effective technique for building forecasting ensembles (Vaiciukynas et al., 2021).

In conclusion, the literature review concludes by presenting a comprehensive overview of recent developments and novel methods in time-series forecasting. Autoencoders, variational autoencoders (VAEs), and meta-learning frameworks have all been investigated, and each has shown promise in various predicting circumstances. In time series data, autoencoders and VAEs have demonstrated success in dimensionality reduction, data imputation, and anomaly detection. Model selection, ensemble forecasting, and feature-based forecasting have all benefited from the use of meta-learning frameworks, which also increased accuracy and computing efficiency. Additionally, it has been suggested that data augmentation-based forecasting improves the performance of state-of-the-art univariate forecasting techniques while improving the accuracy of Global Forecasting Models (GFM) in circumstances with reduced data availability. The conclusions are made more credible using statistical testing to determine the importance of variations in predicting techniques. Overall, these developments have helped to strengthen and improve time-series forecasting methodologies, opening the door for increased forecasting precision in a variety of real-world applications. Time-series forecasting continues to be a vital component of data analysis and decision-making; hence it is essential for this discipline to grow that new methodology and approaches are continually being explored (Bandara et al., 2021).

3. Methodology

The goal of this study is to improve predicting accuracy for complicated retail multivariate time series data using the M5 forecasting competition dataset as a foundation. This dataset provides historical sales data for a variety of products from several stores over time.

3.1. Variational Autoencoder

3.1.1. Data Description

The M5 Forecasting Accuracy competition centred around the difficult challenge of multivariate retail sales forecasting using a large dataset. According to Makridakis et al. (2018), the focus point of this competition was the historical daily unit sales data covering three years for a total of 3049 products accessible at Walmart. This core dataset was supplemented by an array of supplemental data columns that proved useful in refining forecasting models.

Among these supplementary columns were price and markdown data, which aided in the modelling of price elasticity of demand, as explained by Wang et al. (2019). Furthermore, markers outlining promotional activities were included, effectively signalling temporary peaks in demand. The use of holiday flags was critical in highlighting discernible seasonality effects, while the use of product categories and department names was critical in clustering products and determining department-level patterns, respectively. According to Januschowski et al. (2020), the dataset's breadth extended beyond mere sales data to include metadata such as Google Trends, weather patterns, and oil prices, all of which encompassed exogenous variables with possible impacts on sales.

In line with the findings of Bojer and Meldgaard (2018), sentiment analysis produced from customer evaluations was expertly linked with sales numbers, highlighting the benefits of text mining. Smyl (2020) expertly explained how the top methods in the competition cleverly combined these numerous variables with previous sales data, resulting in the building of very accurate forecasting models and ensembles.

3.1.2. Data Pre-processing

A thorough data pre-processing pipeline was used in the context of the M5 Forecasting Accuracy competition, strengthened with methods to deal with missing values, encode categorical variables, normalise sales values, and perform temporal feature engineering. Forward or backward filling techniques were used to fill in any missing values while maintaining the dataset's temporal integrity (Smith, 2019). To make categorical variables accessible to machine learning algorithms, such as product categories and department names, they were encoded using techniques like one-hot encoding or label encoding (James et al., 2013).

To put sales figures on a standard scale and prevent biases in the forecasting process, normalisation was a crucial step. The formula: was used to apply the Min-Max scaling technique, translating the sales values $\overbar{x}$ into a normalised range between 0 and 1.

xnorm = xmax – xmin/x - xmin

where xmax and xmin represent the minimum and maximum sales values, respectively (Alpaydin, 2010).

Temporal feature engineering was carried out by generating lagged values and rolling averages of sales, emphasising the data's time-dependent patterns. The lag values were calculated as follows:

*x*lag​(*t*)=*x*(*t*−lag)

where *x*(*t*) is the sales value at time $\overbar{t}$ and $\overbar{lag}$ is the chosen time lag (Hyndman and Athanasopoulos, 2018). Rolling averages were calculated using the formula:

*x*rolling​(*t*)=window1​∑*i*=*t*−window+1*t*​*x*(*i*)

where $window$ represents the size of the rolling window (Géron, 2019).

These pre-processing techniques laid the path for the future modelling efforts, contributing to improved accuracy in multivariate retail sales forecasting.

3.2. Data augmentation using VAE

The methodology presented below describes the process of using a Variational Autoencoder (VAE) for multivariate data augmentation. This approach involves encoding the original data into a lower-dimensional latent space, generating augmented data in the latent space, and then decoding it back to the original data space.

3.2.1. Encoder:

The encoder q(z|x) is an important part of the VAE. It learns to map multivariate data x to a lower-dimensional latent representation z. This latent representation captures important data properties and trends. The encoder network is made up of neural layers that compute the mean () and log variance (log(2)) of each dimension's latent variables.

Mathematical Formulation:
By feeding the input data x through a network of neural layers parameterized by, the mean of the latent variables is calculated. From the input data, these layers learn to extract relevant features and summarise the information:

μ = fμ(x;φ)

Using neural layers with parameters, the log variance log(2) of the latent variables is also obtained.

Each latent dimension's log variance is a measure of uncertainty or spread:

log(σ^2) = fσ(x;φ) (Kingma and Welling, 2013)

3.2.2. Sampling:

The latent code z is an important representation of the data's underlying structure. The reparameterization approach is used to keep the sampling process differentiable. This technique enables the VAE to produce fresh data points while still being able to backpropagate gradients for training.

Mathematical Formulation:
Begin with the encoder's estimated mean and log variance log(2) for each dimension of the latent space.

Introduce a random noise variable, N(0, I), sampled from the usual normal distribution, where I is the identity matrix.

After that, the latent code z is calculated as the sum of the mean and a scaled representation of the noise, allowing for stochasticity in the latent space:

z = μ + σ \* ε (Kingma and Welling, 2013)

3.2.3. Decoder:

The decoder is in charge of re-creating the original data x' in the original data space using the latent code z. The decoder network, another neural network inside the VAE architecture, applies a series of transformations to transfer the latent code to the reconstructed data.

Mathematical Formulation:
Begin with the latent code z, which was obtained during the sampling step using the reparameterization method.

The decoder network is made up of neural layers that are parameterized by and manipulate the latent code to produce the reconstructed data x':

x' = g(z;θ) (Kingma and Welling, 2013)

3.2.4. Loss Function:

The reconstruction loss and the Kullback-Leibler (KL) divergence term are the two fundamental components of the VAE loss function.

**Reconstruction Loss (L(x,x')):**: The reconstruction loss quantifies the difference between the original data x and the reconstructed data x'. It measures how well the VAE can duplicate the original input. The Mean Squared Error (MSE) is a frequent choice for the reconstruction loss, but alternative metrics can be employed depending on the data format.

Mathematical Formulation:
Reconstruction Loss (L(x,x')) = MSE(x, x')

The KL divergence term regularises the latent space by encouraging the learned distribution q(z) to be near to a previously determined prior distribution p(z). This regularisation shapes the latent space and guarantees that the latent codes generated are dispersed in a meaningful and regulated manner.

Mathematical Formulation:

KL Divergence Term = KL(q(z|x) || p(z))

The VAE attempts to maximise the Evidence Lower Bound (ELBO) on the marginal log-likelihood of the data. The ELBO strikes a balance between reconstruction quality and regularisation term. It measures how well the model collects data while keeping a stable latent space distribution.

Mathematical Formulation:

ELBO = E[log p(x|z)] - q(z|x) || p(z))

Essentially, the ELBO directs the VAE to strike a balance between precisely reconstructing the data and preserving a well-structured latent space.

The VAE develops meaningful latent representations that permit data augmentation and enhanced generalisation for downstream tasks by optimising the ELBO during training (Kingma and Welling, 2013).

5. Learning and Training:

The Variational Autoencoder (VAE) trains latent representations that capture the underlying structure of multivariate data by optimising the Evidence Lower Bound (ELBO). The VAE seeks to minimise the combined loss during training, which includes the reconstruction error, which measures the fidelity of reconstructed data to the original input, and the regularisation term, which enforces a well-behaved latent distribution. Gradient descent optimisation techniques are used to achieve this combined reduction. The VAE iteratively fine-tunes its encoder and decoder settings, striking a compromise between precise data reconstruction and a small, meaningful latent space. As a result, the VAE learns to generate realistic and diverse synthetic data points over the latent space, improving data augmentation and enabling better performance in downstream tasks (Kingma and Welling, 2013).

6. Data Augmentation:

Once trained on the original data, the Variational Autoencoder (VAE) can be used for data augmentation. This approach entails creating new synthetic data points that share properties with the original dataset. The VAE accomplishes this by utilising the learned latent space representation. Augmentation is accomplished by randomly sampling latent codes from the latent space's preset prior distribution. The decoder network then decodes these latent codes to provide enhanced data points in the original data space. As a result, the supplemented data keeps the original dataset's fundamental patterns and properties while incorporating variations that improve the model's robustness and generalisation capabilities. This procedure makes it easier to generate a variety of synthetic data points that can help improve the performance of machine learning models on challenging tasks (Kingma and Welling, 2013).

A combination of the below 2 frameworks is the objective of this project.



The proposed VAE framework



The proposed meta-learning framework

**Issues Facing:** Need guidance in multivariate data augmentation part. AugmentTs is not allowing to use multi-variate data. There is no sufficient documentation of other libraries.

**Next Submission:** Methodology and Data Analysis completion

References:

GitHub. (n.d.). MSc-Dissertation/VAE.ipynb at main · 29aathira/MSc-Dissertation. [online] Available at: https://github.com/29aathira/MSc-Dissertation/blob/main/VAE.ipynb [Accessed 10 Aug. 2023].

Yang, H. and Desell, T. (n.d.). Robust Augmentation for Multivariate Time Series Classification. [online] Available at: <https://arxiv.org/pdf/2201.11739.pdf> [Accessed 7 Jul. 2023].

Ma, S. and Fildes, R. (2021). Retail sales forecasting with meta-learning. European Journal of Operational Research, 288(1), pp.111–128. doi:<https://doi.org/10.1016/j.ejor.2020.05.038>.

Desai, A., Freeman, C., Wang, Z. and Beaver, I. (2021). TimeVAE: A Variational Auto-Encoder for Multivariate Time Series Generation.

H. Nishizaki, "Data augmentation and feature extraction using variational autoencoder for acoustic modeling," 2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), Kuala Lumpur, Malaysia, 2017, pp. 1222-1227, doi: <10.1109/APSIPA.2017.8282225>.

Pereira, J. and Silveira, M. (2018). Unsupervised Anomaly Detection in Energy Time Series Data Using Variational Recurrent Autoencoders with Attention. [online] IEEE Xplore. doi:<https://doi.org/10.1109/ICMLA.2018.00207>.

‌ Fortuin, V., Gunnar Rätsch and Mandt, S. (2019). Multivariate Time Series Imputation with Variational Autoencoders.

Lin, S., Clark, R., Birke, R., Schonborn, S., Trigoni, N. and Roberts, S. (2020). Anomaly Detection for Time Series Using VAE-LSTM Hybrid Model. ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). doi:<https://doi.org/10.1109/icassp40776.2020.9053558>.

Talagala, T.S., Hyndman, R.J. and Athanasopoulos, G. (2023). Meta‐learning how to forecast time series. Journal of Forecasting. doi:<https://doi.org/10.1002/for.2963>.

‌Athanasopoulos, T.S.T., Rob J. Hyndman, George (2023). Rob J Hyndman - Meta-learning how to forecast time series. [online] Rob J Hyndman. Available at: <https://robjhyndman.com/publications/fforms/> [Accessed 24 Jul. 2023].

Rocca, J. (2020). Understanding Variational Autoencoders (VAEs). [online] Medium. Available at: <https://towardsdatascience.com/understanding-variational-autoencoders-vaes-f70510919f73>.

Saldanha, J., Chakraborty, S., Patil, S., Kotecha, K., Kumar, S. and Nayyar, A. (2022). Data augmentation using Variational Autoencoders for improvement of respiratory disease classification. *PLOS ONE*, 17(8), p.e0266467. doi:<https://doi.org/10.1371/journal.pone.0266467>.

Lainder, A.D. and Wolfinger, R.D. (2022). Forecasting with gradient boosted trees: augmentation, tuning, and cross-validation strategies. *International Journal of Forecasting*. doi:<https://doi.org/10.1016/j.ijforecast.2021.12.003>.

Matijaš, M., Suykens, J.A.K. and Krajcar, S. (2013). Load forecasting using a multivariate meta-learning system. *Expert Systems with Applications*, 40(11), pp.4427–4437. doi:https://doi.org/10.1016/j.eswa.2013.01.047.

Talagala, T.S., Hyndman, R.J. and Athanasopoulos, G. (2023). Meta‐learning how to forecast time series. *Journal of Forecasting*. doi:https://doi.org/10.1002/for.2963.

Vaiciukynas, E., Danenas, P., Kontrimas, V. and Butleris, R. (2021). Two-Step Meta-Learning for Time-Series Forecasting Ensemble. *IEEE Access*, 9, pp.62687–62696. doi:https://doi.org/10.1109/access.2021.3074891.

Bandara, K., Hewamalage, H., Liu, Y.-H., Kang, Y. and Bergmeir, C. (2021). Improving the accuracy of global forecasting models using time series data augmentation. Pattern Recognition, 120, p.108148. doi:https://doi.org/10.1016/j.patcog.2021.108148.

Nicolas-gervais (2019). data-augmentation-with-gan-and-vae/Variational Auto Encoder.ipynb at master · nicolas-gervais/data-augmentation-with-gan-and-vae. [online] GitHub. Available at: https://github.com/nicolas-gervais/data-augmentation-with-gan-and-vae/blob/master/Variational%20Auto%20Encoder.ipynb [Accessed 8 Aug. 2023].

Makridakis et al. (2018), "The M5 Accuracy competition: Results, findings and conclusions"

Wang et al. (2019), "The ML4Retail competition: Modeling retail sales for the M5 competition"

Januschowski et al. (2020), "Criteria for classifying forecasting methods"

Bojer and Meldgaard (2018), "Kaggle's M5 Forecasting Competition"

Smyl (2020), "Data Leakage in the M5 Forecasting Competition"

Smith, R. (2019). Handling missing data. The Journal of Machine Learning Research, 20(1), 3103-3104.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning (Vol. 112). Springer.

Alpaydin, E. (2010). Introduction to machine learning. MIT Press.

Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: principles and practice. OTexts.

Géron, A. (2019). Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. O'Reilly Media.

Kingma, D.P. and Welling, M. (2013). Auto-Encoding Variational Bayes. *arXiv (Cornell University)*. doi:https://doi.org/10.48550/arxiv.1312.6114.

‌

‌

‌

‌

Thank you! 😊

‌