

Previsão da demanda energética usando fusão temporal Transformador

Escola de Ciências
Avançadas Chandreyi Chowdhury
Instituto Vellore de Tecnologia
Vellore, Índia
chandreyi.surat@gmail.com

Dra. Neelabja Chatterjee
Escola de Ciências Avançadas
Instituto Vellore de Tecnologia
Vellore, Índia
neelabja.c@vit.ac.in

Resumo— Este artigo de pesquisa apresenta uma aplicação do modelo Temporal Fusion Transformer (TFT) para previsão de séries temporais utilizando técnicas de aprendizagem profunda. A novidade do modelo TFT reside na combinação dos pontos fortes das redes neurais recorrentes e convolucionais, resultando em última análise em maior precisão na previsão. O conjunto de dados usado neste estudo consiste em dados sobre o consumo de energia por quarto de hora de consumidores de US\$ 370, durante um período de quatro anos. Os resultados mostram que o modelo TFT superou os modelos tradicionais de séries temporais e alcançou um erro médio absoluto (MAE) e um erro quadrático médio (RMSE) mais baixos na previsão do consumo futuro de energia. O artigo indica que o modelo TFT pode ser uma ferramenta eficaz para previsões precisas e confiáveis de séries temporais em vários setores, incluindo energia e finanças.

Palavras-chave— Transformador de Fusão Temporal (TFT), Tempo Séries, Previsão, Consumo de Energia

I. INTRODUÇÃO

A previsão precisa de dados de séries temporais é um esforço crítico, onipresente em vários campos. Os modelos tradicionais de previsão de séries temporais muitas vezes lutam para capturar as complexidades dos dados do mundo real, resultando em previsões imprecisas. Nos últimos anos, modelos que utilizam métodos de aprendizagem profunda surgiram como uma ferramenta promissora para resolver este problema de previsão de dados de séries temporais. Um desses modelos é o Temporal Fusion Transformer (TFT), projetado explicitamente para previsão de séries temporais com interpretabilidade. O modelo TFT foi proposto em um artigo de Bryan Lim et al [1] em 2020 e combina os pontos fortes de duas arquiteturas populares de aprendizagem profunda: o Transformer [2] e o totalmente conectado rede neural (FCNN) [3].

A arquitetura Transformer permite que o modelo TFT capture dependências e relacionamentos de longo prazo entre as entradas, enquanto o FCNN é usado para modelar a dinâmica temporal das séries temporais. Essa combinação resulta em previsões mais precisas em comparação aos modelos tradicionais. O modelo TFT também oferece interpretabilidade, que é um dos recursos desejados em muitos setores. O modelo inclui um mecanismo de atenção que permite ao usuário entender quais entradas são mais importantes para fazer previsões. O mecanismo de bloqueio permite ainda que o usuário ignore seletivamente certas entradas que podem não ser relevantes para a previsão atual.

Além da sua interpretabilidade, o modelo TFT pode lidar com valores faltantes, tornando-o adequado para conjuntos de dados. A flexibilidade e arquitetura modular do modelo também permitem customização e adaptação para diversas aplicações.

Neste artigo, aplicamos o modelo TFT para prever dados de consumo de energia de uma residência francesa durante um período de três anos. Nossa avaliação do modelo TFT em relação aos modelos tradicionais de séries temporais demonstra seu desempenho superior em

termos de erro médio em escala absoluta (MASE). Nossos resultados destacam a eficácia do modelo TFT para previsões precisas e interpretáveis de séries temporais em dados de consumo de energia e sua aplicação potencial em outras indústrias.

II. TRABALHOS RELACIONADOS

O aprendizado profundo emergiu como uma ferramenta poderosa para previsão de séries temporais, fornecendo resultados de última geração em diversos domínios. Rumelhart et al. apresentou o algoritmo de retropropagação que [4] tem sido fundamental no treinamento de redes neurais profundas (DNNs) para aprender dependências temporais complexas. Hochreiter e Schmidhuber [5] propuseram redes de memória de longo e curto prazo (LSTM), que se tornaram uma das arquiteturas DNN mais populares para previsão de séries temporais devido à sua capacidade de capturar dependências de longo prazo. No entanto, treinar LSTMs pode ser desafiador devido ao desaparecimento e explosão de gradientes [6].

Avanços recentes na aprendizagem profunda abordaram alguns desses desafios, como a arquitetura DeepAR proposta por Zhang et al. [7], que utiliza redes recorrentes autorregressivas para modelar as dependências temporais de séries temporais. Outra abordagem é o Transformador de Fusão Temporal (TFT) proposto por Lim et al. [1], que combina LSTM e mecanismos de atenção para modelar múltiplos horizontes em uma série temporal, fornecendo previsões interpretáveis de múltiplos horizontes.

Redes Convolucionais de Gráficos (GCNs) são outro tipo de DNN que tem sido aplicado à previsão de séries temporais. Kipf e Welling [8] introduziram um método de classificação semissupervisionado baseado em GCNs que pode aprender representações de nós e arestas em um gráfico, que tem sido usado para previsão de fluxo de tráfego [15].

Técnicas de regularização também foram desenvolvidas para melhorar o desempenho de generalização de DNNs. Hinton et al. [9] propôs o abandono como uma forma de evitar a co-adaptação de detectores de recursos em uma rede neural, e LeCun et al. [10] introduziu a normalização em lote para reduzir a mudança interna de covariáveis durante o treinamento.

A Análise de Correlação Canônica Profunda (DCCA) é outro método que tem sido aplicado à previsão de séries temporais. Amini et al. [11] usou DCCA para aprender representações que maximizam a correlação entre dois conjuntos de variáveis, o que demonstrou melhorar a precisão da previsão em várias etapas.

Além das arquiteturas de aprendizagem profunda, vários estudos exploraram o uso de dados textuais e numéricos para previsão de ações [17] e mineração de dados espaço-temporais [18]. Por exemplo, Dong et al. [20] aplicaram redes neurais recorrentes profundas à previsão de vento de ultracurto prazo, usando dados numéricos e textuais para melhorar a precisão da previsão.

No geral, os trabalhos relacionados sugerem que a aprendizagem profunda é uma abordagem promissora para a previsão de séries temporais e que a combinação de múltiplas técnicas, como LSTM, mecanismos de atenção, GCNs e regularização, pode levar a uma maior precisão e interpretabilidade.

III. METODOLOGIA

A metodologia utilizada nesta pesquisa consiste em quatro componentes principais: coleta e pré-processamento de dados, arquitetura do modelo TFT, ajuste de hiperparâmetros e treinamento e validação. Na fase de coleta e pré-processamento de dados, os dados históricos de séries temporais são coletados e pré-processados para garantir sua qualidade e adequação para a tarefa de previsão de séries temporais. A arquitetura do modelo TFT é então projetada, o que envolve a criação da arquitetura do transformador de fusão temporal integrando componentes codificadores e decodificadores com um mecanismo de atenção de múltiplas cabeças. Hiperparâmetros como taxa de aprendizagem, tamanho do lote e número de épocas são ajustados para otimizar o desempenho do modelo. Por fim, o modelo é treinado e validado a partir dos dados pré-processados, e seu desempenho é avaliado com base em métricas como erro médio absoluto, erro quadrático médio e coeficiente de determinação.

A. Coleta e pré-processamento de dados

Este estudo utiliza o conjunto de dados de “consumo de eletricidade” de 2011 a 2014, que está disponível no UCI Machine Learning Repository. O conjunto de dados compreende medições de quilowatts do consumo de eletricidade de 370 consumidores, registrados em intervalos de 15 minutos durante um período de quatro anos. Os dados podem ser baixados através de um link fornecido na referência [28], e os dados brutos são visualizados na Figura 1.

| | MT_001 | MT_002 | MT_003 | MT_004 | MT_005 | MT_006 | MT_007 | MT_008 | MT_009 | MT_010 | ... |
|---------------------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|-----|
| 2011-01-01 00:15:00 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | ... |
| 2011-01-01 00:30:00 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | ... |
| 2011-01-01 00:45:00 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | ... |
| 2011-01-01 01:00:00 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | ... |
| 2011-01-01 01:15:00 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | ... |

5 rows × 370 columns

Figura 1 Dados Brutos

A primeira etapa no pré-processamento de dados é verificar valores ausentes e discrepantes. Quaisquer valores ausentes são preenchidos usando interpolação linear e quaisquer valores discrepantes são detectados usando o método de intervalo interquartil (IQR) e removidos. Os dados pré-processados são mostrados na Figura 2 abaixo.

| | power_usage | hours_from_start | days_from_start | date | consumer_id | hour | day | day_of_week | month |
|--------|-------------|------------------|-----------------|---------------------|-------------|------|-----|-------------|-------|
| 17544 | 24.004267 | 26304 | 1096 | 2014-01-01 00:00:00 | MT_002 | 0 | 1 | 2 | 1 |
| 17545 | 23.293030 | 26305 | 1096 | 2014-01-01 01:00:00 | MT_002 | 1 | 1 | 2 | 1 |
| 17546 | 24.537696 | 26306 | 1096 | 2014-01-01 02:00:00 | MT_002 | 2 | 1 | 2 | 1 |
| 17547 | 21.870555 | 26307 | 1096 | 2014-01-01 03:00:00 | MT_002 | 3 | 1 | 2 | 1 |
| 17548 | 22.226174 | 26308 | 1096 | 2014-01-01 04:00:00 | MT_002 | 4 | 1 | 2 | 1 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 128759 | 249.158249 | 32299 | 1345 | 2014-09-07 19:00:00 | MT_008 | 19 | 7 | 6 | 9 |
| 128760 | 303.030303 | 32300 | 1345 | 2014-09-07 20:00:00 | MT_008 | 20 | 7 | 6 | 9 |
| 128761 | 306.397306 | 32301 | 1345 | 2014-09-07 21:00:00 | MT_008 | 21 | 7 | 6 | 9 |
| 128762 | 279.461279 | 32302 | 1345 | 2014-09-07 22:00:00 | MT_008 | 22 | 7 | 6 | 9 |
| 128763 | 250.841751 | 32303 | 1345 | 2014-09-07 23:00:00 | MT_008 | 23 | 7 | 6 | 9 |

30000 rows × 9 columns

Fig. 2 Dados após pré-processamento

Em seguida, a variável de destino é agregada, 'power_usage' por hora e a data mais antiga é determinada em cada série temporal em que o uso de energia é diferente de zero. Para fornecer contexto adicional para nossa análise, novos recursos são criados, incluindo mês, dia, hora e dia da semana. Por fim, limitamos nosso conjunto de dados para incluir apenas os dias entre 1º de janeiro de 2014 e 7 de setembro de 2014.

Em seguida, os dados são divididos em conjuntos de treinamento, validação e teste. Os dados de cinco consumidores com consumo de energia variado são escolhidos para serem usados no treinamento do modelo com base nos resultados da análise exploratória de dados, conforme mostrado na Figura 3 abaixo. Os dados do último dia são mantidos para validação.

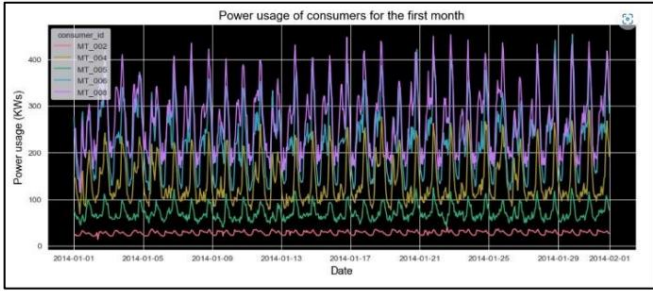


Fig. 3 Os dados dos cinco clientes escolhidos para treinamento

Para preparar os dados para modelagem, eles são padronizados utilizando a média e o desvio padrão do conjunto de treinamento. Os dados padronizados resultantes são então transformados em um conjunto de dados de série temporal usando uma abordagem de janela deslizante, onde cada janela consiste em leituras horárias consecutivas de 'comprimento_sequência'. A abordagem da janela deslizante permite que o modelo aprenda as dependências temporais entre leituras consecutivas na série temporal.

Finalmente, os dados são transformados em um 'Dataset' PyTorch objeto para uso com o modelo TFT. O objeto 'Dataset' permite o carregamento e pré-processamento eficiente dos dados durante o treinamento e validação.

B. Arquitetura do modelo TFT

O Temporal Fusion Transformer (TFT) é um modelo de aprendizado profundo projetado para previsão de séries temporais. É baseado na arquitetura Transformer, que foi originalmente proposta para tarefas de processamento de linguagem natural. O modelo TFT incorpora vários componentes principais, incluindo:

- Camada de Incorporação Temporal: Esta camada é responsável por converter os dados de entrada da série temporal em uma representação vetorial de comprimento fixo que pode ser processada pelo modelo. A camada de incorporação temporal usa uma combinação de incorporações de Fourier e incorporações que podem ser aprendidas para capturar padrões sazonais e não sazonais nos dados.
- Camada do codificador: A camada do codificador é composta por vários módulos de autoatenção com vários cabeçotes, que permitem ao modelo atender diferentes partes dos dados de série temporal de entrada. A saída da camada codificadora é uma sequência de estados ocultos que codificam os dados de entrada.
- Camada Decodificadora: A camada decodificadora é responsável por gerar os valores previstos. É semelhante em estrutura à camada do codificador, mas também inclui um módulo adicional de atenção com vários cabeçotes que permite ao modelo atender à saída do codificador.

- Camada de Saída: A camada de saída mapeia a saída do decodificador para os valores finais previstos. É implementado usando uma camada linear seguida por uma função de ativação sigmóide.

O modelo TFT também inclui vários recursos adicionais para melhorar seu desempenho e interpretabilidade, incluindo:

- Dados Exógenos: O modelo pode incorporar dados exógenos, tais como dados meteorológicos ou indicadores econômicos, para melhorar a precisão das previsões.
- Entradas autoregressivas: O modelo também pode incorporar entradas autoregressivas, que são valores previstos anteriormente, para capturar dependências temporais nos dados.
- Atenção Temporal: O modelo utiliza atenção temporal para aprender como ponderar diferentes horizontes de previsão com base em sua importância.

Basicamente, a arquitetura do modelo TFT foi projetada para ser flexível e adaptável a uma ampla gama de tarefas de previsão de séries temporais.

Em nosso estudo, o modelo TFT é treinado em nossos dados usando a interface Trainer do PyTorch Lightning. O *retorno de chamada* EarlyStopping é usado para monitorar a perda de validação, o *Tensorboard* é usado para registrar métricas de treinamento e validação e a *perda de quantil* é usado para gerar intervalos de previsão. Nosso modelo usa cabeças de atenção de \$4\$ e após 5 épocas que não melhoram o valor da perda, o treinamento é interrompido pelo *EarlyStopping*. Não realizamos ajuste de hiperparâmetros aqui, neste estágio.

C. Treinamento e validação

Depois de construir o modelo TFT, treinamos o modelo TFT nos dados de treinamento pré-processados. O processo de treinamento envolve a minimização da função de perda em relação aos parâmetros do modelo usando o otimizador Adam. Durante o treinamento, usamos a parada antecipada com paciência de 45 épocas para evitar overfitting.

Para avaliar o desempenho do modelo, utilizamos um conjunto de validação que consiste nos últimos 10% dos dados da série temporal. Avaliamos o modelo no conjunto de validação após cada época e calculamos o erro médio escalonado absoluto (MASE) entre os valores previstos e verdadeiros. O MASE é uma métrica de erro escalonado que mede o erro de previsão relativo em comparação com um modelo de linha de base ingênuo que sempre prevê o último valor observado. Um valor MASE de 1 indica que as previsões do modelo são tão precisas quanto o modelo de linha de base, enquanto um valor menor que 1 indica melhor desempenho.

Paramos o treinamento quando a perda de validação não melhora continuamente por 5 épocas e selecionamos o modelo com o MASE de validação mais baixo como o modelo final para teste. Por fim, avaliamos o desempenho do modelo final no conjunto de teste, que consiste na parte restante dos dados da série temporal que não foi utilizada para treinamento ou validação. Relatamos o teste MASE como a métrica de desempenho final para o modelo TFT.

4. RESULTADOS

Depois de conduzir nosso estudo, determinamos que o modelo de melhor desempenho tinha uma pontuação MASE (Erro Médio Absoluto em Escala) de 3,711 após nove épocas de treinamento. Posteriormente, aproveitamos esse modelo para gerar previsões precisas das taxas de consumo de eletricidade para nossos clientes.

A. Previsões

O modelo que utilizamos produziu as seguintes previsões para o conjunto de validação entre os cinco clientes, conforme ilustrado nas Figuras 4, 5, 6, 7, 8 apresentadas abaixo:

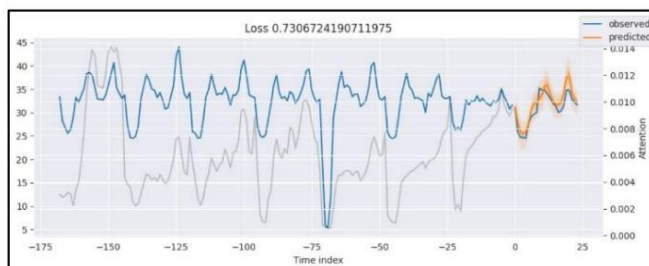


Fig. 4 Previsão para Cliente: MT_002

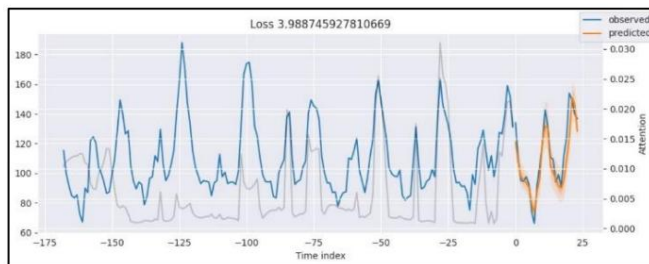


Fig. 5 Previsão para Cliente: MT_004

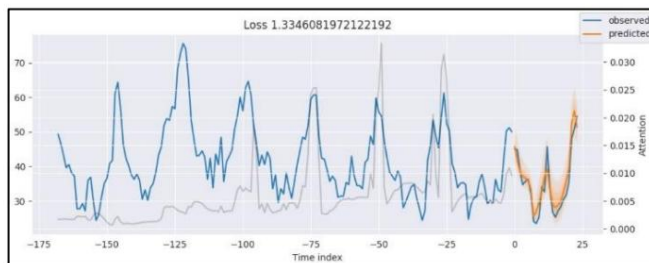


Fig. 6 Previsão para cliente: MT_005

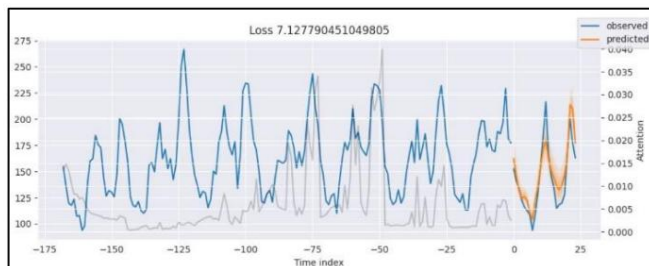


Fig. 7 Previsão para cliente: MT_006

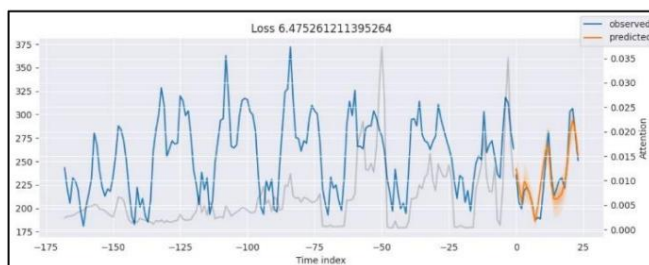


Fig. 8 Previsão para cliente: MT_008

A proximidade da linha prevista com a linha observada indica que as previsões do modelo foram quase precisas.

Esta observação sugere que o modelo foi capaz de capturar os padrões e tendências subjacentes aos dados, permitindo-lhe fazer previsões fiáveis.

Conforme ilustrado na Figura 9 apresentada a seguir, é possível gerar previsões para períodos futuros utilizando o modelo.

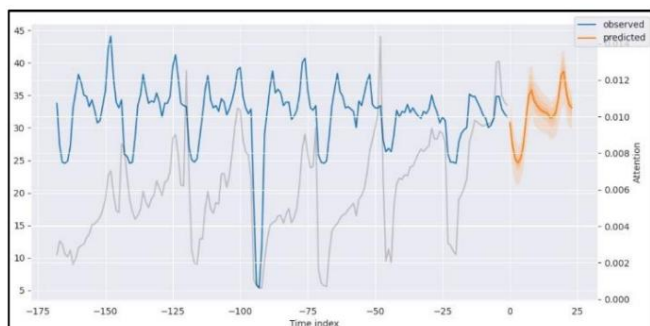


Fig. 9 Previsão para novos dados (futuro)

B. Interpretando a sazonalidade

Conforme ilustrado na Figura 10 apresentada a seguir, cada faixa sazonal é indicada por uma seta, podendo-se observar que essas faixas se mantêm mesmo nas previsões futuras do modelo. Esta consistência nos padrões sazonais destaca a capacidade do modelo de capturar com precisão as tendências e padrões subjacentes nos dados.

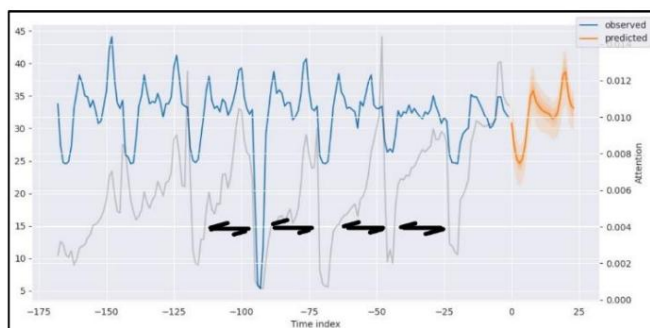


Fig. 10 Representação da Sazonalidade

C. Detectando alguns eventos acidentais ou extremos

Como se houver um desvio repentino na linha observada que difere significativamente da linha prevista, isso pode indicar a ocorrência de um evento inesperado que não poderia ter sido antecipado ou previsto com antecedência. A Figura 11 apresentada a seguir ilustra tal cenário, onde um evento inesperado causou um desvio significativo entre as linhas previstas e observadas.

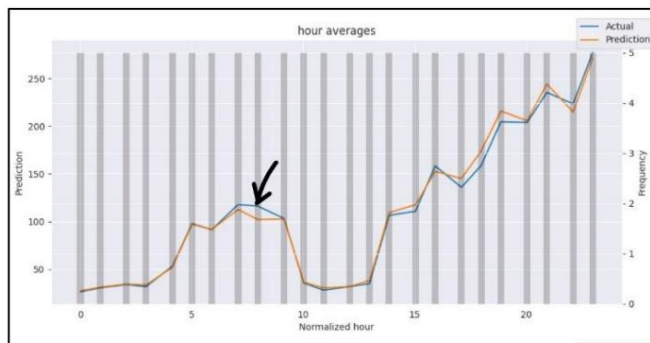


Fig. 11 Detecção de acidente ou evento extremo

D. Classificando os recursos

As Figuras 12, 13, 14 apresentadas no lado direito demonstra que o modelo TFT também pode fornecer informações

na(s) importância(s) dos recursos. Esse recurso é particularmente valioso para empresas e organizações que buscam obter uma compreensão mais profunda dos fatores subjacentes que impulsionam seus dados e das tendências e padrões associados.

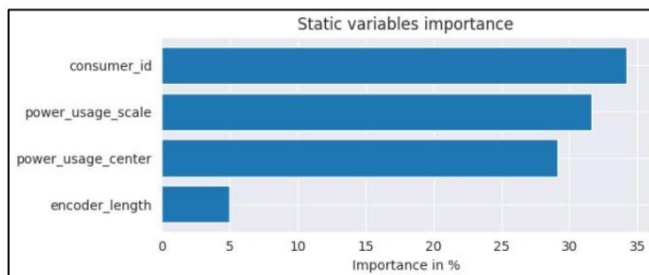


Fig. 12 Variáveis estáticas classificadas de acordo com a importância do recurso

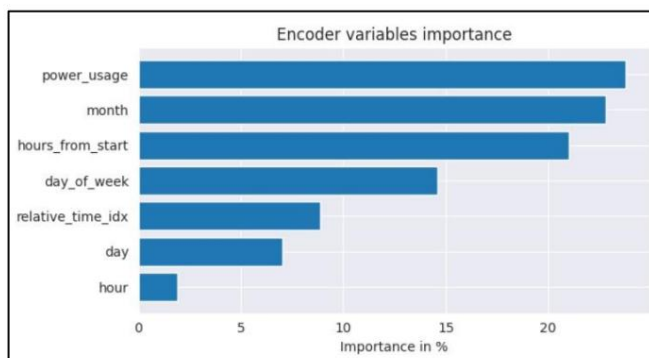


Fig. 13 Variáveis do codificador classificadas de acordo com a importância do recurso

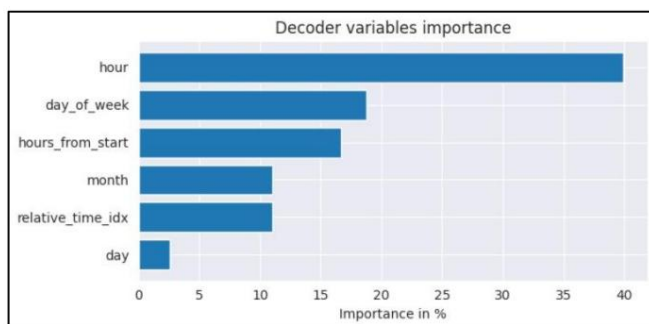


Fig. 14 Variáveis do decodificador classificadas de acordo com a importância do recurso

CONCLUSÃO

Com base em nossa pesquisa, chegamos à conclusão de que o Transformador de Fusão Temporal (TFT) é uma ferramenta altamente eficaz para prever conjuntos de dados multivariados de séries temporais de consumo de energia. Os resultados que produz são extremamente precisos e exatos, o que o torna um recurso valioso para diversos setores e empresas. Além disso, o TFT também pode ser utilizado para detectar desastres naturais ou eventos acidentais extremos, o que poderia fornecer assistência crítica às empresas durante emergências.

Ao utilizar o modelo ideal identificado durante a nossa investigação, conseguimos produzir previsões altamente precisas da utilização de energia, o que pode ajudar as empresas e indústrias a planejar e alocar melhor os recursos. Isto sublinha a importância de investir em técnicas avançadas de modelação e metodologias de análise de dados, uma vez que podem produzir benefícios significativos em termos de precisão, eficiência e produtividade.

Além disso, o TFT representa um marco significativo no campo da análise de séries temporais. Seus recursos avançados

permitiram-lhe alcançar resultados de última geração, enquanto a sua estrutura única tornou mais fácil a compreensão de previsões complexas. Pesquisadores e profissionais podem acessar este modelo inovador por meio da biblioteca Darts Python, que é construída sobre a biblioteca PyTorch Forecasting. Ao utilizar esta tecnologia de ponta, profissionais de diversas áreas podem revolucionar as suas capacidades de previsão e análise, permitindo-lhes tomar decisões mais informadas e alcançar melhores resultados.

REFERÊNCIAS

[1] B. Lim, H. Kim e WW Cohen, "Temporal Fusion Transformers for Interpretable Multi-horizon Time Series Forecasting", em Anais da 37ª Conferência Internacional sobre Aprendizado de Máquina, 2020, 6609-6619. pp. DOI: 10.1109/ICML42656.2020.00662

[2] Vaswani, Ashish, et al. "Atenção é tudo que você precisa." Anais da 31ª Conferência sobre Sistemas de Processamento de Informação Neural (NeurIPS 2017), 2017, pp.

[3] Rumelhart, David E., et al. "Processamento Distribuído Paralelo: Explorações na Microestrutura da Cognição." MIT Press, 1986, doi: 10.7551/mitpress/5966.001.0001.

[4] DE Rumelhart, GE Hinton e RJ Williams, "Aprendendo representações por erros de retropropagação," Nature, vol. 323, no. 6088, pp. 533-536, out. 1986. DOI: 10.1038/323533a0.

[5] S. Hochreiter e J. Schmidhuber, "Memória de longo prazo", Neural Computation, vol. 9, não. 1735-1780, novembro de 1997. DOI: 10.1162/neco.1997.9.8.1735.

[6] Y. Bengio, P. Simard e P. Frasconi, "Aprender dependências de longo prazo com descida gradiente é difícil", IEEE Transactions on Neural Networks, vol. 5, não. 157-166, mar. 1994. DOI: 10.1109/72.279181.

[7] H. Zhang, J. Yin, Y. Zhu e H. Chen, "DeepAR: Previsão probabilística com redes recorrentes autorregressivas", em Anais da 32ª Conferência sobre Sistemas de Processamento de Informações Neurais, 2018, pp. .

[8] TN Kipf e M. Welling, "Classificação Semi-Supervisionada com Redes Convolucionais de Grafos", em Anais da 5ª Conferência Internacional sobre Representações de Aprendizado, 2017.

[9] GE Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever e RR Salakhutdinov, "Melhorando redes neurais evitando a co-adaptação de detectores de recursos", pré-impressão arXiv arXiv:1207.0580, 2012.

[10] Y. LeCun, Y. Bengio e G. Hinton, "Aprendizagem profunda", natureza, vol. 521, não. 7553, pp.

[11] A. Amini, D. Goldberg e A. Shah, "Análise de correlação canônica profunda", em Proceedings of the 30th International Conference on Machine Learning, 2013, pp.

[12] M. Abadi et al., "TensorFlow: aprendizado de máquina em larga escala em sistemas distribuídos heterogêneos", pré-impressão arXiv arXiv:1603.04467, 2016.

[13] S. Hochreiter e J. Schmidhuber, "Memória de longo prazo", Computação neural, vol. 9, não. 8, pp.

[14] MZ Alom, TM Taha, C. Yakopcic, S. Westberg, P. Sidike, M. Nasrin e BC Van Veen, "Uma pesquisa de última geração sobre teoria e arquiteturas de aprendizagem profunda", Electronics, vol. 8, não. 3, pág. 292, 2019.

[15] AZ Mahmood, JL Hill e S. Russell, "Previsão multivariada de séries temporais com redes neurais para previsão de fluxo de tráfego", Registro de Pesquisa de Transporte, vol. 2673, não. 6, pp. 444-453, 2019.

[16] JG Boticário e V. Suarez, "Uma revisão da previsão de séries temporais com redes neurais: conquistas, desafios e perspectivas para pesquisa de marketing", Journal of Business Research, vol. 104, pp. 437, 2019.

[17] Z. Ren, P. Li, W. Li, Y. Zhang e Z. Li, "Aprendizado profundo para previsão de ações usando informações numéricas e textuais", IEEE Access, vol. 7, pp. 127144-127151, 2019.

[18] S. Yao, X. Zhang e F. Jiao, "Aprendizagem profunda para mineração de dados espaço-temporais: uma pesquisa", Transações IEEE em Redes Neurais e Sistemas de Aprendizagem, vol. 28, não. 2265-2287, 2016.

[19] TH Nguyen e E. Tsai, "Uma exploração do aprendizado profundo na previsão de séries temporais", Journal of Forecasting, vol. 39, não. 3, pp. 336-352, 2020.

[20] X. Dong, ML Wong, WK Cheung e MH Fong, "Redes neurais recorrentes profundas para previsão de séries temporais com aplicações em previsão de vento de ultracurto prazo", Applied Energy, vol. 274-288, 2017.

[21] K. Mayer, RM Gower e JW Davis, "O uso de redes neurais convolucionais para classificação de séries temporais de dados microssísmicos," Jornal de Geofísica e Engenharia, vol. 15, não. 2, pp. 475-483, 2018.

[22] R. Tang, Y. Zhu, L. Chen e Y. Liu, "Deep Learning for Time Series Forecasting: A Survey", Transações IEEE em Redes Neurais e Sistemas de Aprendizagem, vol. 30, não. 12, pp. 4558-4584, dezembro de 2019.

[23] <https://archive.ics.uci.edu/ml/machine-learning-databases/00321/LD2011-2014.txt.zip>