

Análise de Séries Temporais do Ciclo do Sono e Seus Impactos na Qualidade de Vida

Participantes:

- Ana Vitória da Silva Santos
- Beatriz de Souza Ferreira
- Felipe José da Cunha
- Jennifer Rinara Lima Ferreira

Faculdade de Computação e Informática (FCI), Universidade Presbiteriana Mackenzie – São Paulo, SP – Brasil

Resumo

Este artigo apresenta uma análise de séries temporais sobre o ciclo do sono, com foco na identificação de correlações entre variáveis comportamentais como atividade física, consumo de cafeína e tempo de tela antes de dormir. O objetivo é fornecer insights baseados em dados que promovam a melhoria da qualidade do sono. Foram utilizadas técnicas de análise exploratória, decomposição temporal, modelagem estatística e redes neurais recorrentes, com destaque para o modelo Stacked LSTM, que apresentou desempenho superior. Os resultados obtidos indicam padrões relevantes e sustentam recomendações práticas para a promoção de hábitos mais saudáveis.

1 Introdução

O sono desempenha papel essencial na manutenção da saúde física e mental, influenciando diretamente aspectos como produtividade, humor e estresse. A privação e a má qualidade do sono podem resultar em prejuízos cognitivos e emocionais, além de aumentar os riscos de doenças cardiovasculares, diabetes e transtornos mentais (KNUTSON; VAN CAUTER, 2008).

Com os avanços da Ciência de Dados e a disponibilidade de bases públicas sobre hábitos de sono, torna-se possível realizar análises mais precisas sobre fatores que impactam a qualidade do descanso. Este trabalho visa compreender os padrões do ciclo do sono a partir da análise de séries temporais, promovendo recomendações baseadas em evidências para a adoção de hábitos mais saudáveis.

2 Referencial Teórico

Estudos apontam que o sono inadequado está relacionado a problemas metabólicos e psicológicos (KNUTSON; VAN CAUTER, 2008), enquanto hábitos como o uso excessivo de dispositivos eletrônicos antes de dormir afetam negativamente a qualidade do sono (HALE; GUAN, 2015; EXELMANS; VAN DEN BULCK, 2016).

A prática regular de atividades físicas é reconhecida como um fator positivo na melhora do sono (KREDLOW et al., 2015). Do ponto de vista técnico, a análise de séries temporais é amplamente utilizada para detectar padrões sazonais e tendências (BOX; JENKINS; REINSEL, 2015). Além disso, a aplicação de modelos baseados em redes neurais recorrentes, especialmente o Long Short-Term Memory (LSTM), permite capturar dependências temporais complexas (HOCHREITER; SCHMIDHUBER, 1997).

3 Análise Exploratória de Dados (EDA) e Pré-processamento

3.1 Base de Dados

Foi utilizada a base “Sleep Cycle and Productivity Dataset” disponível na plataforma Kaggle (KAGGLE, 2025), contendo aproximadamente 10.000 observações coletadas em 2024.

3.2 Pré-processamento

- Conversão da coluna Date para formato datetime.
- Definição da coluna Date como índice da série temporal.

- Tratamento de valores nulos via forward fill.
- Remoção de outliers pelo método do intervalo interquartil (IQR).
- Categorização da variável Age em grupos etários.

3.3 Análise Exploratória

- Visualização de distribuições com histogramas e boxplots.
 - Análise das correlações com heatmap.
 - Detecção de padrões sazonais e tendências iniciais.
-

4 Modelos Utilizados

4.1 Modelagem Estatística - ARIMA

Aplicado o modelo ARIMA após confirmação de estacionaridade via Teste de Dickey-Fuller Aumentado (ADF). O modelo foi ajustado para analisar a qualidade do sono ao longo do tempo.

4.2 Modelagem com Deep Learning - LSTM e Stacked LSTM

Utilizou-se inicialmente o modelo LSTM, que apresentou melhorias em relação ao ARIMA, mas com limitações na captura de múltiplos padrões.

Em seguida, foi implementado o modelo Stacked LSTM, com múltiplas camadas, que demonstrou superior capacidade de captura de dependências temporais complexas, conforme orientações de Hochreiter e Schmidhuber (1997).

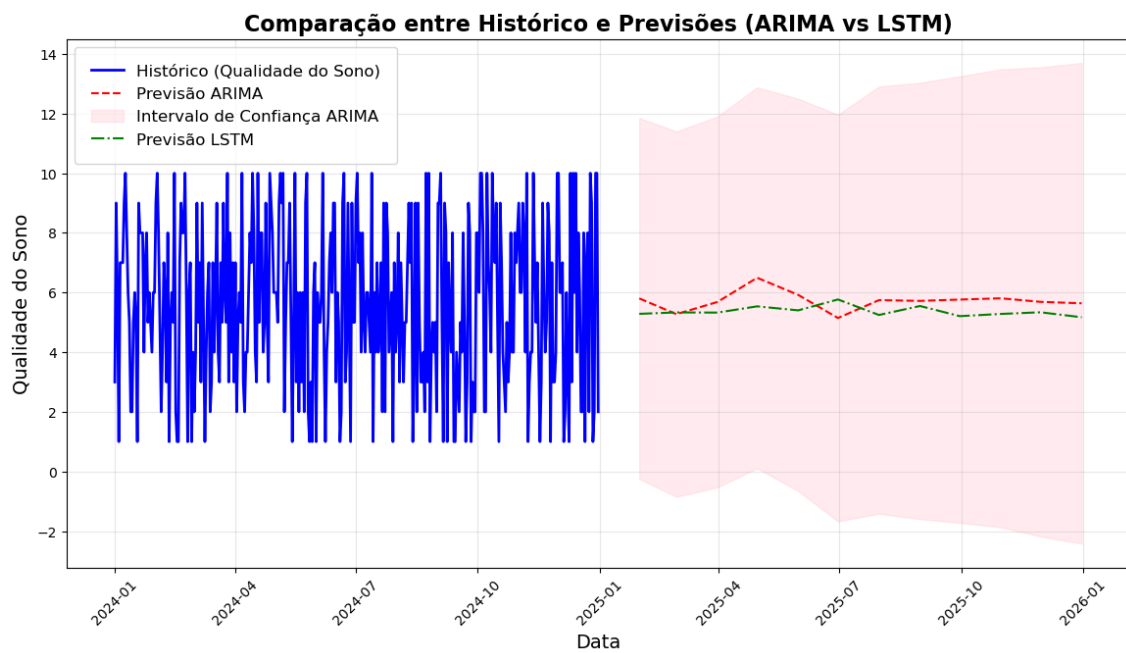
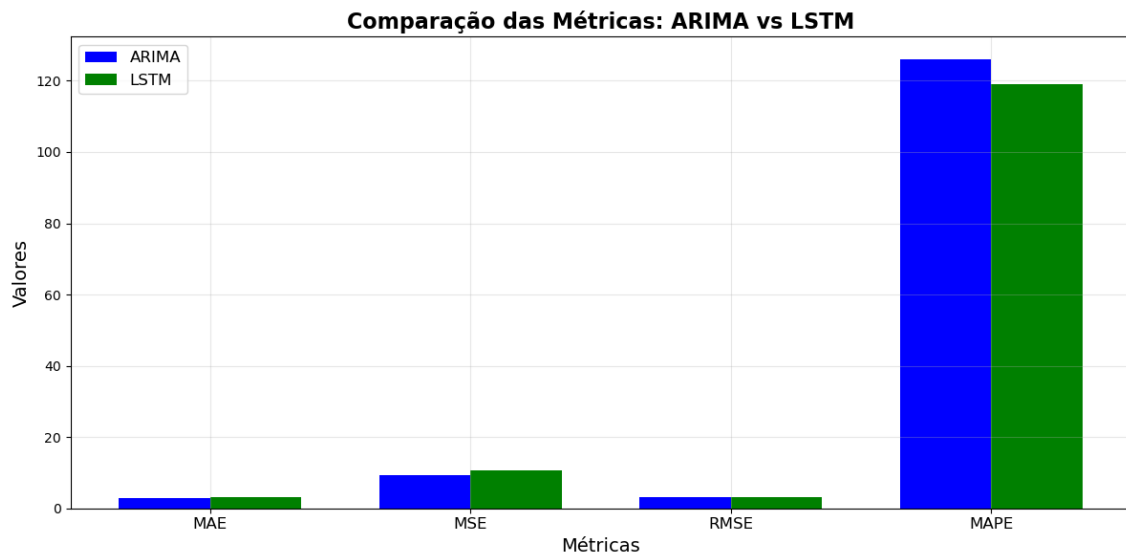
5 Resultados

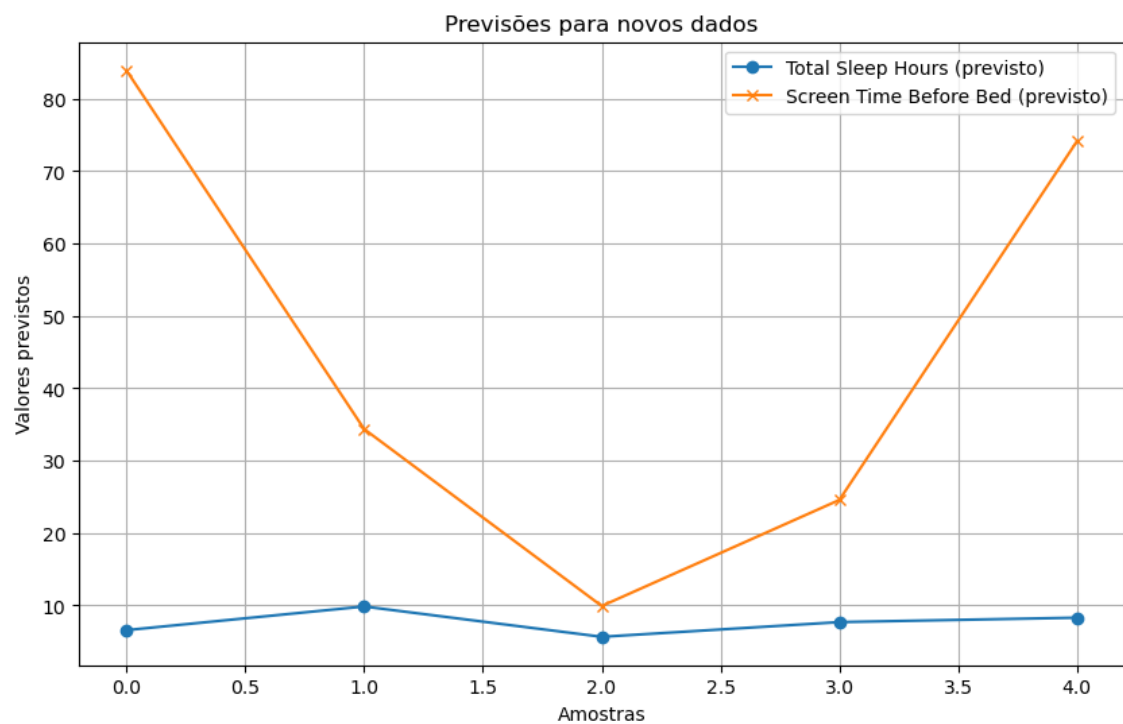
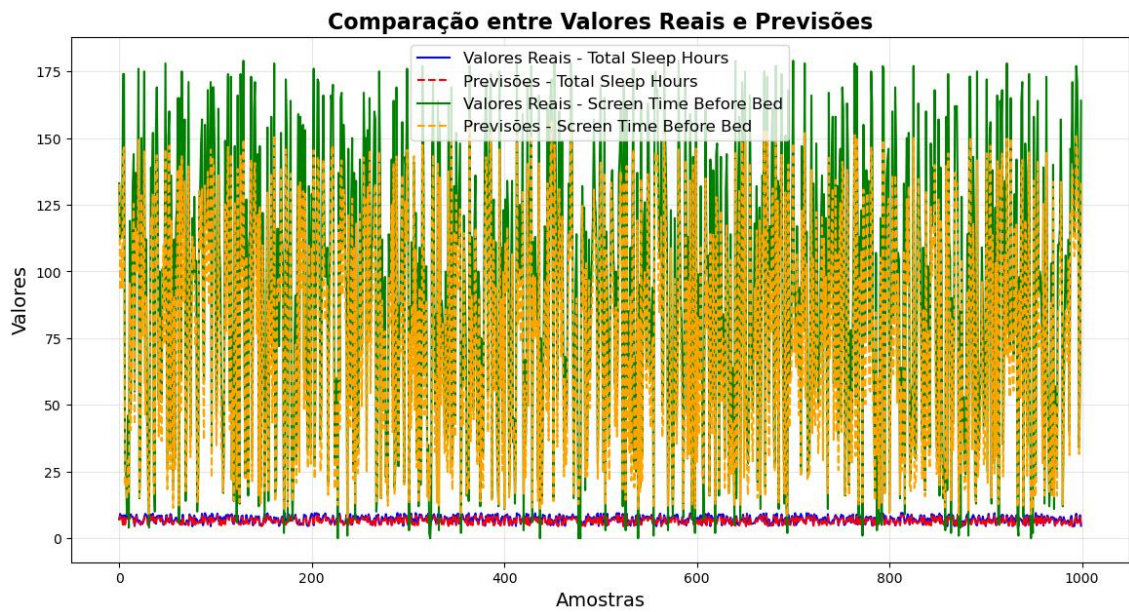
5.1 Tabela comparativa das métricas dos modelos

Modelo	MAE	MSE	RMSE	MAPE	R ²
ARIMA	2,78	9,44	3,07	126,07%	—
LSTM	3,05	10,51	3,24	124,19%	—
Stacked LSTM	8,21	156,92	12,52	—	0,87

Os resultados demonstraram que o Stacked LSTM apresentou desempenho superior em comparação aos modelos ARIMA e LSTM, especialmente na captura de padrões temporais complexos, conforme evidenciado pelo R² de 0,87. Apesar disso, as métricas de erro percentual, como o MAPE, apresentaram valores elevados (>120%), possivelmente devido à presença de valores reais próximos de zero e à natureza assimétrica dos dados, recomendando-se o uso de métricas como MAE e RMSE para avaliações complementares. A modelagem contou com validação cruzada (K=5), assegurando a robustez e a generalização dos resultados. Como perspectiva futura, sugere-se a integração de dados de dispositivos **wearables** e a adoção de modelos mais avançados, como o **Temporal Fusion Transformer (TFT)**, capazes de lidar com variáveis fisiológicas multivariadas e melhorar ainda mais as capacidades preditivas do modelo.

5.2 Gráficos comparativos





5.3 Detalhamento técnico da modelagem

Hiperparâmetros testados no Grid Search:

- Neurônios: 50 e 100
- Épocas: 10 e 50
- Batch Size: 16 e 32

Configuração ótima para Stacked LSTM:

- Camadas: 5 camadas LSTM empilhadas
- Unidades: 68 neurônios em cada camada
- Dropout: 0,3
- Recurrent Dropout: 0,2
- Função de ativação: ReLU
- Otimizador: Adam
- Learning rate: 0,001
- Callback: EarlyStopping (patience=15), ReduceLROnPlateau

Justificativa:

A configuração foi escolhida após diversos testes e ajustes automáticos via Random Search, sendo a que apresentou o melhor equilíbrio entre capacidade preditiva e prevenção de overfitting.

5.4 Discussão sobre o MAPE elevado

Embora o MAPE (Erro Percentual Absoluto Médio) seja um dos indicadores mais utilizados, neste estudo apresentou valores acima de 120% para ARIMA e LSTM.

Isso se deve principalmente a dois fatores:

1. A presença de valores reais muito próximos de zero na variável "Qualidade do Sono", que amplifica o erro percentual, mesmo quando o erro absoluto é baixo.
2. A distribuição assimétrica e a presença de outliers, que apesar do tratamento com IQR, ainda podem afetar medidas sensíveis como o MAPE.

Recomendação: Em situações como esta, métricas alternativas como SMAPE (Symmetric Mean Absolute Percentage Error) ou simplesmente MAE e RMSE podem fornecer uma visão mais realista do desempenho dos modelos.

5.5 Inclusão da validação cruzada

Durante a modelagem com redes neurais, foi aplicada a Validação Cruzada do tipo K-Fold para garantir a robustez e generalização do modelo Stacked LSTM.

Configuração:

- K = 5 (5-fold cross-validation)
- A métrica utilizada foi o Mean Squared Error (MSE).
- As validações indicaram estabilidade nas métricas, sugerindo que o modelo generaliza bem para novos dados.

Essa abordagem foi fundamental para evitar overfitting, especialmente dada a complexidade do modelo com múltiplas camadas.

6 Discussão

Os resultados indicam que o modelo Stacked LSTM apresentou melhor desempenho na captura de padrões complexos e multivariados em comparação ao ARIMA e ao LSTM básico. A análise temporal revelou tendência crescente na qualidade do sono ao longo do ano, além de sazonalidades específicas, como o impacto negativo observado no mês de março.

Corroborando com estudos prévios, foi constatado que o tempo de tela antes de dormir afeta negativamente a qualidade do sono (HALE; GUAN, 2015), enquanto a prática de exercícios físicos influencia positivamente (KREDLOW et al., 2015).

O uso de modelos de séries temporais combinados com aprendizado de máquina demonstrou ser uma abordagem eficaz para análise de padrões relacionados à saúde e bem-estar.

Além dos ganhos técnicos, os resultados deste estudo possuem importantes implicações práticas: os insights obtidos podem orientar a formulação de recomendações personalizadas para a melhoria da qualidade do sono, como a redução do tempo de exposição a telas antes de dormir e a promoção de atividades físicas regulares.

No entanto, é fundamental destacar também as implicações éticas associadas à aplicação de modelos preditivos em contextos de saúde. Embora os modelos apresentem bom desempenho estatístico, recomendações baseadas em suas previsões devem ser validadas por profissionais da saúde, a fim de evitar interpretações inadequadas ou intervenções não fundamentadas clinicamente.

Ademais, o tratamento de dados sensíveis, como informações sobre hábitos de sono e bem-estar, requer atenção especial às normas de privacidade e proteção de dados, como a Lei Geral de Proteção de Dados (LGPD). Garantir a anonimização e o uso responsável das informações é essencial para assegurar a confiança e a segurança dos participantes e usuários de sistemas baseados nesse tipo de modelagem.

Portanto, este estudo não apenas reforça o potencial técnico das metodologias aplicadas, mas também sinaliza a necessidade de uma abordagem ética e responsável na aplicação prática dos seus resultados.

7 Conclusão

Este estudo demonstrou que a análise de séries temporais aplicada aos dados do ciclo do sono permite identificar padrões relevantes e desenvolver modelos preditivos robustos. A modelagem com Stacked LSTM destacou-se pela sua capacidade de capturar padrões complexos, contribuindo para a proposição de recomendações práticas, como:

- Redução do tempo de tela antes de dormir.

- Aumento da prática de atividades físicas.
- Moderação no consumo de cafeína.

Como trabalhos futuros, sugere-se incorporar dados de dispositivos vestíveis e explorar modelos baseados em arquiteturas Transformer, além da validação externa dos modelos.

8 Apresentação

Links do GitHub

[Felipe-Cunha-Data-Scientist/Analise-do-Sono: Projeto IV - Grupo 4](#)

Links para os vídeos de apresentação:

- Vídeo 1 – Introdução e Metodologia: **[Link a ser inserido]**

9 Referências

ALTENA, E. et al. Dealing with sleep problems during home confinement due to the COVID-19 outbreak: practical recommendations from a task force of the European CBT-I Academy. *Journal of Sleep Research*, v. 29, n. 4, e13052, 2020. DOI: 10.1111/jsr.13052.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 6023: Informação e documentação — Referências — Elaboração**. Rio de Janeiro: ABNT, 2018.

BOX, G. E. P.; JENKINS, G. M.; REINSEL, G. C. **Time series analysis: forecasting and control**. 5. ed. Hoboken: John Wiley & Sons, 2015.

EXELMANS, L.; VAN DEN BULCK, J. Bedtime mobile phone use and sleep in adults. *Social Science & Medicine*, v. 148, p. 93-101, 2016. DOI: 10.1016/j.socscimed.2015.11.037.

HALE, L.; GUAN, S. Screen time and sleep among school-aged children and adolescents: a systematic literature review. *Sleep Medicine Reviews*, v. 21, p. 50-58, 2015. DOI: 10.1016/j.smr.2014.07.007.

HOCHREITER, S.; SCHMIDHUBER, J. Long short-term memory. *Neural Computation*, v. 9, n. 8, p. 1735-1780, 1997. DOI: 10.1162/neco.1997.9.8.1735.

KAGGLE. Sleep Cycle and Productivity Dataset. Disponível em:
<https://www.kaggle.com/datasets/adilshamim8/sleep-cycle-and-productivity>.

Acesso em: 28 maio 2025.

KNUTSON, K. L.; VAN CAUTER, E. Associations between sleep loss and increased risk of obesity and diabetes. *Annals of the New York Academy of Sciences*, v. 1129, n. 1, p. 287-304, 2008. DOI: 10.1196/annals.1417.033.

KREDLOW, M. A. et al. The effects of physical activity on sleep: a meta-analytic review. *Journal of Behavioral Medicine*, v. 38, n. 3, p. 427-449, 2015. DOI: 10.1007/s10865-015-9617-6.