Review 1

it is not clear whether the authors have investigated all the variations of Transformer.

V - Ci è o ci fa? Ci fa. Cambierei il titolo, mettiamo minimalistic così non pensa che parliamo di tutti i transformer

It is not clear what is the fundamental novelty of the proposal.

V - Ridiciamo che il contributo è pedagogico

It is not clear whether the authors have made sufficient comparisons with all the variations of Transformer.

V - Come il primo

Review 2

The manuscript is primarily pedagogical, aiming at interpretability and transparency. While this is valuable, it does not propose fundamentally new architectures or theoretical insights.

V - Da aims della rivista: “The aim of Algorithms is to encourage scientists to publish their experimental and theoretical results”, il nostro è expermental

Compared with recent transformer adaptations (Informer, Autoformer, FEDformer, PatchTST), the novelty is limited.

V - v. sopra

Mathematics in the article is largely a restatement of known transformer operations.

V - v. sopra

The “minimalist” approach does not introduce new theoretical insights.

V - v. sopra

1. The title is too generic; the authors must have to modify the title to be specifically linked with the proposed contribution.

V - ok

1. The abstract repeatedly uses the pronoun “we” (i.e., “we describe,” “we implement,” “we validate”). I recommend rephrasing these sentences in the passive voice or by referring directly to the work. The same goes for the conclusion, too.

V - Tocca farlo, ma vedi Transformers is all you need e altri 1000, è meglio usare we

1. Clearly articulate the study’s contributions and novelty, supported by key findings/quantitative results for abstract enhancement.

V - Diciamo che l’abbiamo fatto

1. Evaluation is limited to univariate series; multivariate testing would provide stronger validation. Could the minimalist approach be extended to multivariate data, which is more realistic in forecasting applications?

V - Aggiungo una frase

1. A block diagram of the overall architecture should be included. Adding a visual representation of the transformer-based model will improve clarity and help readers better understand the workflow and component interactions.

F - ok

1. Hyperparameter tuning procedures are not sufficiently detailed. Results may depend strongly on initialization, learning rate, and sequence length choices. How robust is the model to missing data, noise, or seasonality beyond simple normalization?

V - ok

1. Statistical validation (Mann-Whitney U-test) is appropriate, but the depth of interpretation is minimal.

V - Aggiungo una frase

1. Performance differences across domains are reported but not deeply analyzed (e.g., why finance performed better than demographics).

VF - Aggiungo una frase

1. Scalability test is not included for the proposed scheme (e.g., long-horizon forecasts on very large datasets).

F - Facciamo due o tre prove

1. Discuss the trade-offs between interpretability and accuracy more explicitly.

F - Non sa di cosa parla

1. Include an ablation study showing how much accuracy is lost at each simplification step compared to a standard transformer.

F - Tocca farlo, credo sia semplice

1. It is recommended to enhance Figures 2, 3, 4 & 5 and their font sizes, as these are not properly readable.

F - ok

1. Computational efficiency (training time, memory footprint, scalability) is not analyzed. Since the model is advertised as “minimalist,” these aspects should be emphasized.

F - ok

1. Strengthen the discussion on computational efficiency (training time, parameter count vs. accuracy).

VF - Aggiungo una frase

1. Benchmarking is limited to Random Forest, which is a relatively weak baseline. Comparisons with ARIMA/ETS and recent transformer variants would provide a more meaningful context.

V - ok

1. A few areas for improvement are needed for English. Some sentences are overly long and could be made more concise (e.g., in the Introduction and Methodology). Some redundancy in stating that the architecture is “minimalist” and “transparent” multiple times.

V - ok

Review 3

1) The parameters adopted in the different experiments are introduced at page 4 “…in the case of the running example we used the following argument values: n = 7, m = 4, k = 2, dk = dv = 2, p = 16” but they are not described. Perhaps a table explaining them and later on a short motivation for their selection in the different cases could help the readers understand how they should select them.  
V - ok

2) Perhaps adding convolutional neural networks with attention could be interesting to the study. [1]

[1] Temporal Convolutional Attention Neural Networks for Time Series Forecasting Yang Lin, Irena Koprinska, Mashud Rana Code: https://github.com/YangLIN1997/TCAN-IJCNN2021

F - Aggiungiamo anche questa citazione