

# Review of: "Machine Learning Methods in Algorithmic Trading: An Experimental Evaluation of Supervised Learning Techniques for Stock Price"

S. E. Schaeffer<sup>1</sup>

<sup>1</sup> McGill University

Potential competing interests: No potential competing interests to declare.

Small edits:

1) decisionmaking → decision-making / decision making

2) extra newline at "employed \n for high" (there are a few other places, too, with extra newlines such as "full \n training set")

At the start of Section 3, an overview paragraph of the steps would be helpful. Same goes for Section 4. An opening paragraph sets the stage for what you are about to describe and helps your reader share your goal.

In Section 3.1, links to potential reputable databases would be useful. Particularly the yfinance API mentioned in Section 3.2 would be best accompanied by a reference to identify the source (and the precise version used). Same goes for the tf reference in Section 3.4.

For metrics (Section 3.5), it is understandable that you won't include the definitions, but please cite something that does so that your readers have an easy way to freshen up on which is which. The metrics mentioned in Section 3.5 differ from the ones listed in Section 4; consistency would be good.

It is a bit hard to read when you switch between future, present, and past tenses. Anything that refers to this present work could just be written in the present tense, prior work in the past tense, and upcoming work in the future tense. Makes it easier to understand the scope of the manuscript.

In Table 1 (or any other table really), numerical columns should be right-aligned for ease of comparison and the number of decimal places should be kept consistent. Mixing scientific notation (e-05) with decimal representation makes it hard to compare the information. For such a long table, a heat map with colours instead of numbers might be easier to digest. Another option would be putting the whole table in a linked CSV file hosted on GitHub and just using the top 10 in some sense for the table that is shown in the manuscript. The present form of the table is not very helpful for the reader.

As you well state in your discussion, model performance is largely a question of tuning the parameters, which is not done as far as I understand. If defaults are used, it is a game of luck which ones do well on a particular data set. It would be good to discuss this in more detail in Section 7 as part of the limitations.

If you were to show a heat map over multiple datasets and using more than one set of parameter values per method, it would be easier to argue whether one outperforms another.

Section 8 confuses me. It sort of sounds like future work since no results are reported, but it is written as if the bot was already made. It might be better to discuss the bot in a separate contribution and cross-reference the two manuscripts.

I would cut out the first sentence of Section 9 “Our research endeavors culminate in a succinct yet impactful conclusion.” — it is vague, contributes little, and may come off as arrogant.

The beginning of the conclusions should be written in past tense (you are using future tense, which again makes it hard to understand what is happening).

I highly advise against not disclosing the code. Reproducible research is only possible if we all share our code.

Figure 2 is not very legible since the font size of the labels is too small. The caption text should also indicate what is the data that is being forecast. The closing price of something in particular?

All in all, this is a good start but not quite yet a solid contribution to the state of the art on financial forecasting.