

EP2420 Project 2 - Forecasting Service Metrics - Peer Review

André Silva

January 5, 2021

1 Project Overview

This section provides a good, concise overview of the project, although it could be improved by providing a clearer mention to the time series analysis models used. It manages to touch the relevant goals of the project, but misses to mention the final achieved results. Another aspect that would improve it would be to include a reference to the paper where the dataset was produced.

2 Background

Again, this section also provides a good and clear introduction to the models and essential concepts, with a good in-depth explanation of the ARIMA model. This being said, the readability would be improved if better formatting would be applied, namely using \LaTeX objects such as equations and references to both equations and figures.

Including more references, particularly on the ARIMA sub-section, would improve the section.

3 Data Sets and Data Pre-Processing

This section comprehends everything that is expected, although it would be nice to clarify what the threshold used for outlier removal actually means (i.e. standard deviation σ).

4 Tasks

In Task II, you could explain the hyper-parameter search you did. It is also not necessary to include the pre-processing of the dataset again, as this was already mentioned in its own section. When you compare LSTM and Linear Regression you mention that LSTM is better, "except for several points". It would be nice to provide an explanation as to why this happens, as well as where.

In Task III, it would be nice to provide a general formula for the methods utilized.

In general, introducing \LaTeX objects such as references to figures, tables and equations would facilitate an easier read, by creating hyperlinks, as well as automatically updating the text when new structures are added. This would prevent some referencing errors that occur throughout the report.

Result wise, you mention that you expected to achieve a lower accuracy with a higher l . My intuition tells me otherwise, as more information is available. The results for the KV periodic dataset also contradict this, so it could be something related with the dataset and/or the step size used.

Overall, I found these sections clarifying, and was able to extract the information I was looking for without much hassle.

5 Discussion

The text is a bit dense and confusing at some times. It would be nice to format it in a different way, for example by separating into several paragraphs.

It would also be valuable to mention future work on the several topics studied.

This being said, all the information I was looking for was available and not difficult to find.