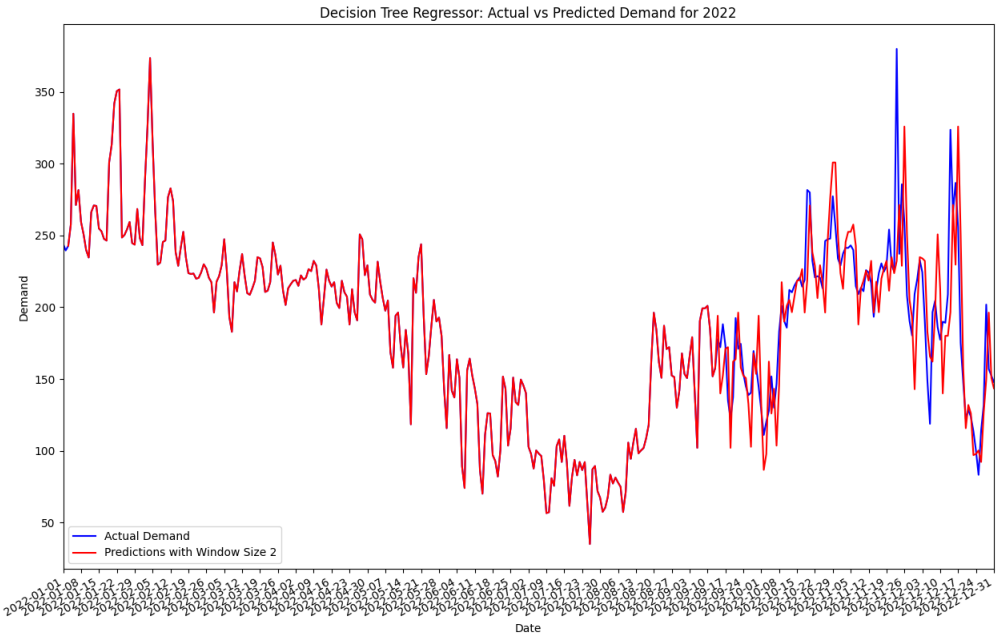
1)

I would analyse the data and get more features than I have with just the demand, like day of week, and if there were more years available, we could also have time of year, holidays and such as well, which would impact the sales. The sales vary wildly per day, especially in 2023 with the big peak late January.

2)

I tried to make a function that created the best possible sliding window size, but I believe it to be overfitting at the small window size of 2, even though it is predicting pretty good with a single tree as long as you have the test data. If I am predicting without any actual data for a extended period, it just goes flat. Not much to say here, I made a tree, trained it, and checked the error. I would say the tree seems to do pretty well, even on the test set (which is unseen data), but it didn't to particularly well when trying to predict far ahead in the future. I don't think any of these models are good choices for doing this, there are so many parameters to adjust, and they are so finicky to get correct. I could have used GridSearchCV to find the best hyperparameters, but I chose to not do this, as I would have to do that for every single task, and I don't have time or a computer to do that right now. It is simply to demanding (I tried for MLP, it ran for 1 hour, then I had to stop it)

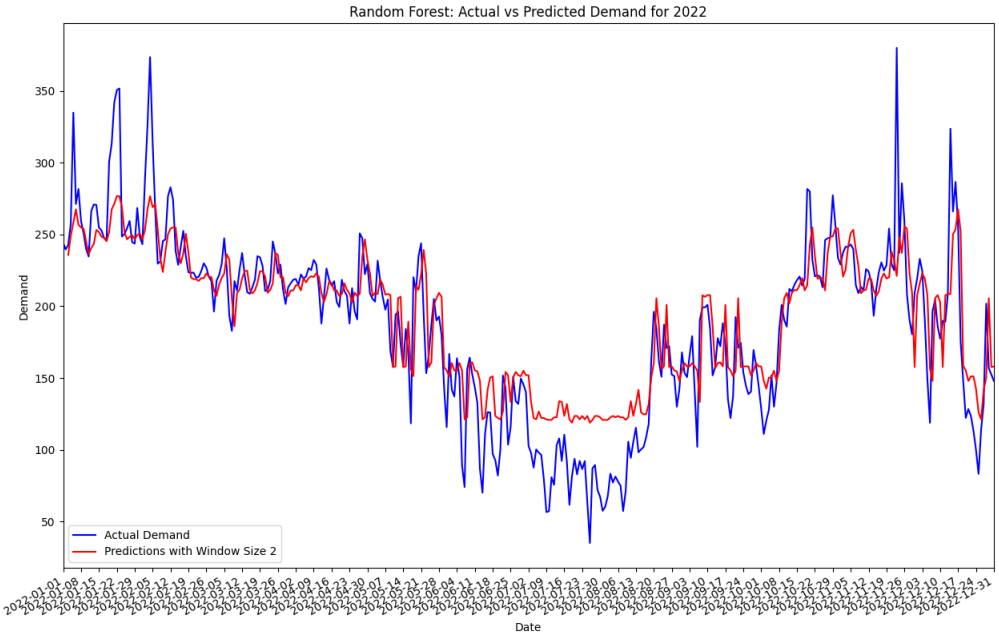


It seems to be doing pretty well on the unseen data here, of course, it is fed with each day as it goes along in time, so it will always know the actual data of the previous day when predicting the next, which the last task does not provide.

3)

To implement a random forest, I simply created a custom class that created n DecisionTreeRegressors, and I used bootstrapping to give each tree different samples. Since this data is time series, it has to be sequential data, it can not be random. I set the block size to 100, so each bootstrap sample is 100 days long, this should give enough overlap between all the trees to give reasonable results. I have randomised the parameters as requested, and I tried different values but I did not see any big changes except if I locked it to set values.

For the sliding window, I kept it at the window size the first single tree got, which is too small, and probably leads to overfitting, but it seemed to work fine on a single tree, and I am reasonably happy with the random forest output. It has to be said, the MAE for a single tree was pretty good, while it was not as good for the random forest, so in this case, I think a single tree is a better solution than a random forest. For aggregation I used mean, as that is what makes most sense in regression in my eyes.

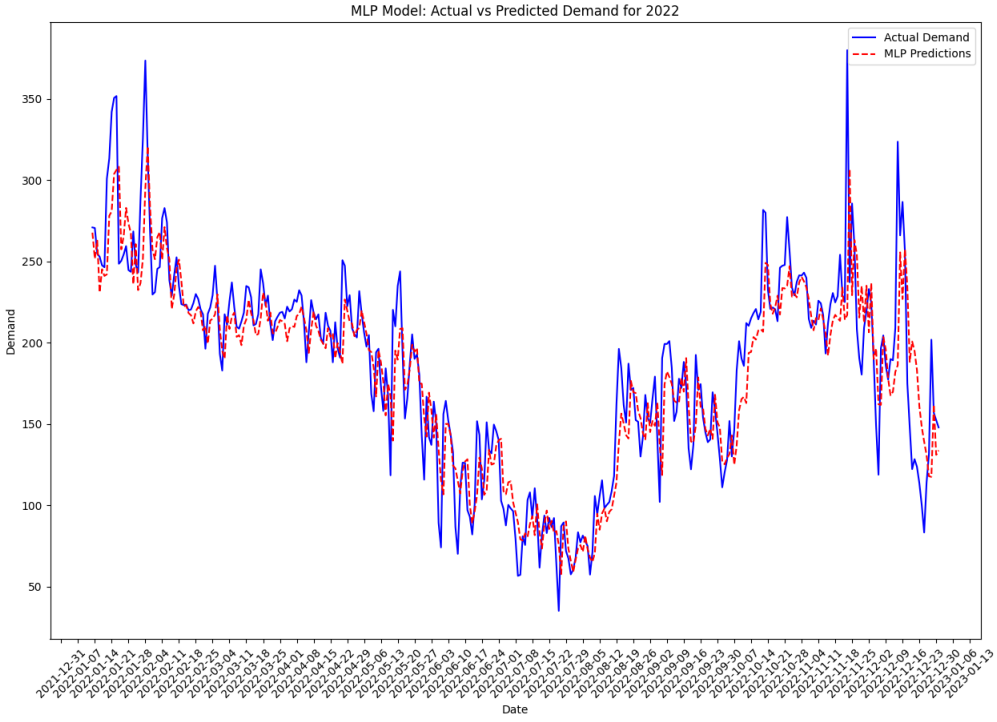


You can see that it is smoothing out the data since it is using mean here, and that is probably why the MAE is higher than on the single tree. The cut for training/test is 70/30, so 256/109 days. I have been consistent with this throughout.

4)

For the MLP, which we also used in Neural Networks earlier, I again feel like the single perceptron did better than the ensemble. I used the same tactic as for the random forest, I made a custom class, and used bootstrapping with sequences (not sure if this is actually called bootstrapping at this point? It is changed from the Random Forest, as I am using resample, which states that it is the first step of bootstrapping) as data to feed to each MLP. I experimented with different layer sized, layer counts, learning rates, optimisers, loss functions and iterations (most of all because I thought we were to randomise them like for the random forest), but I did not find anything useful. As I mentioned before, running GridSearchCV to find the best hyperparameters is the way to go, but it took too long, and maxed my CPU so I could not use the computer for anything else (6 cores/12 threads were maxed).

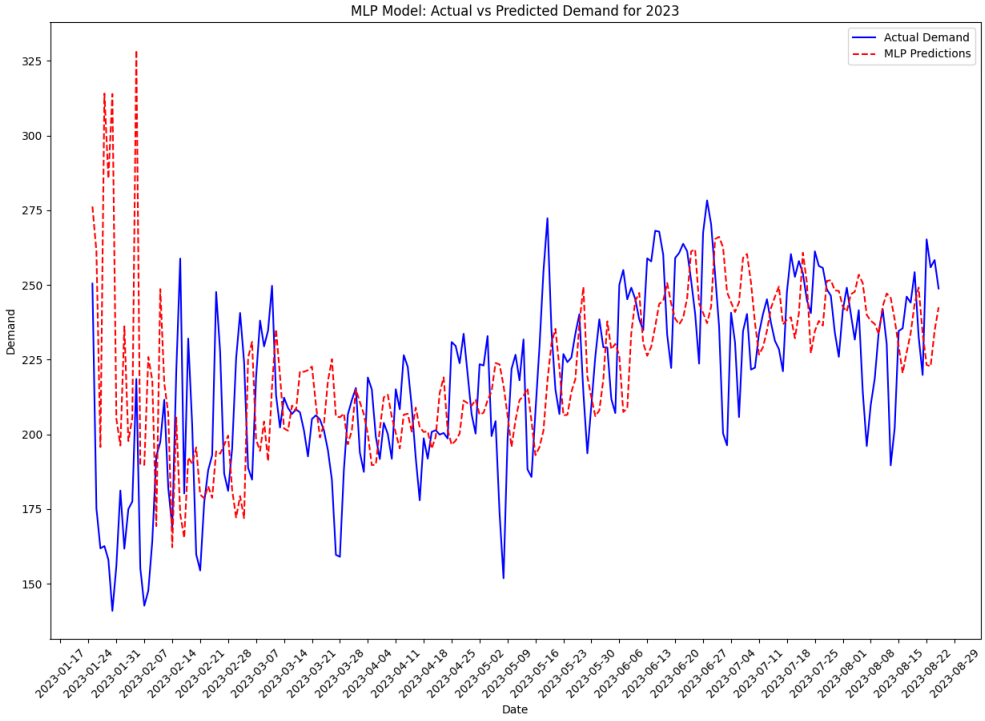
I tried changing it up, and doubled the calculated window size so I could get better results, but it did not help.



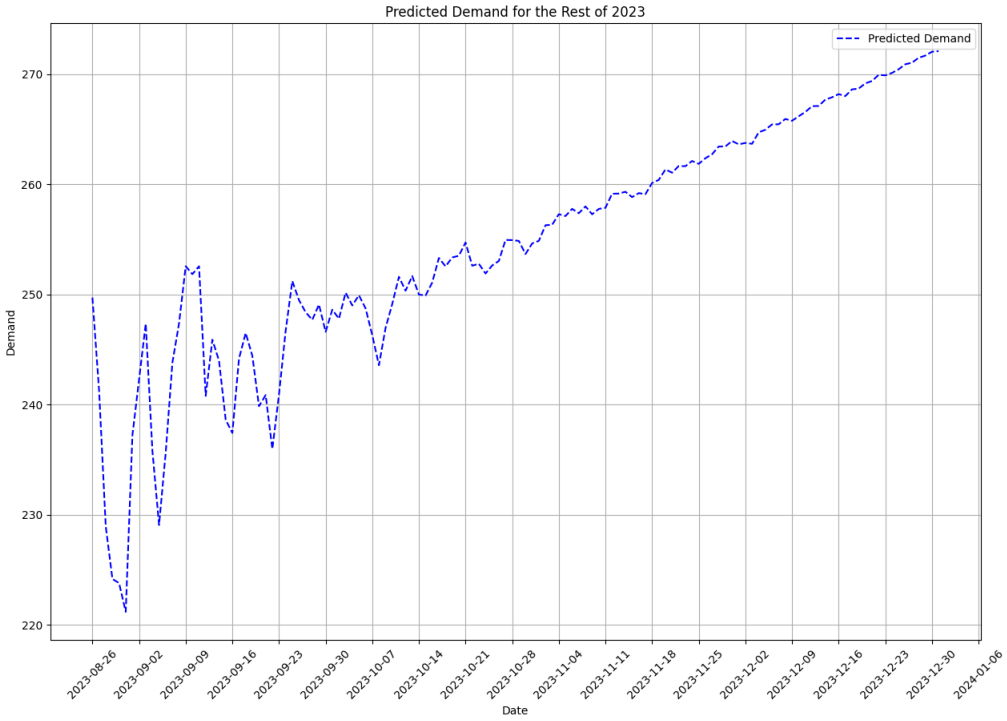
This is using the default 12 window size. It seems to be each and every one of these methods react, and do not act, so they are always too slow to get anything that would be considered really usable.

5)

I chose to go with the MLP Ensemble, and I am not happy with the results. They seem to start reasonably well, but then they start going upwards indefinitely. I am going to be honest and say I am not sure why, they seem to be reasonable in the start, so maybe the values are just increasing for so long that the window size is only seeing increase, and then it will only go in one direction, and that is up. I could try to increase the window size, but that gave me issues with feature counts, for some reason.



This is the training of 2023 data, I have a higher window size compared to 2022, and you can see it as it is acting up in the start due to the big spike in January that is not seen here as the window size is too large. For the rest of the period you can see the smoothing effect using mean on multiple MLPs or trees has on the graph, compared to the blue actual data.



Here is the prediction graph, as I said, it is diverging and going crazy, and I think it is due to window size, perhaps overfitting as well, but it could just as well be my hyperparameters, which I haven't adjusted properly to my liking.

As for some last words, I have actual use for this! I am going to use my Christmas to try doing something similar for my old workplace, a bowling centre in Svolvær. I will not use these simple models, but rather deep learning with a Recurrent Neural Network, and I will add data for weather like clouds, precipitation and wind, as well as seasons, holidays and similar. I think this will be an interesting task, and with more features and more data (I have 8 years of income available), I can probably get something more useful than what I did with the data I had in this assignment.

All in all it was an interesting assignment, but maybe a bit much to do on top of everything else with computer graphics project, exam, e-test in AI and neural network graded assignment, especially since the data has to be read in in different ways for tree/random forest and MLP/MLP Ensemble, same goes for the graphing, lacking information online, and not too much information about this in this subject either.