Sindre J.I Sivertsen & Sander Kilen

The title in the world!

master project, Fall 2021

Artificial Intelligence Group
Department of Computer and Information Science
Faculty of Information Technology, Mathematics and Electrical Engineering

Abstract

This paper provides a template for writing AI project rapports for either the AI specialisation project; masters "datateknikk" or masters "informatikk". The use of the template is recommended and is written in english as we encourage students to submit their project and masters theses in English. The template does not form a compulsory style that you are obliged to use. However, the format and contents are a result of a joint AI group initiative thus providing a common starting point for all AI students. For a given project tuning of the template may still be required. Such tuning might involve moving a chapter to a section or vice versa due to the nature of the project.

The abstract is your sales pitch which encourages people to read your work but unlike sales it should be realistic with respect to the contributions of the work. It should include:

- the field of research
- a brief motivation for the work
- what the research topic is and
- the research approach(es) applied.
- contributions

The abstract length should be roughly half a page of text — without lists, tables or figures.

Preface

The preface includes the facts - what type of project, where it is conducted, who supervised and any acknowledgements you wish to give.

Sindre J.I Sivertsen & Sander Kilen
Trondheim, November 3, 2021

Contents

List of Figures

List of Tables

Chapter 1

Introduction

All chapters should begin with an introduction before any sections begin. Further, each sections begins with an introduction before subsections begin. Chapters with just one section or sections with just one sub-section, should be avoided. Think carefully about chapter and section titles as each title stand alone in the table of contents (without associated text) and should convey meaning for the contents of the chapter or section.

In all chapters and sections it is important to write clearly and concisely. Avoid repetitions and if needed, refer back to the original discussion or presentation. Each new section, subsection or paragraph should provide the reader with new information and be written in your own words. Avoid direct quotes. If you use direct quotes, unless the quote itself is very significant, you are conveying to the reader that you are unable to express this discussion or fact yourself. Such direct quotes also break the flow of the language (yours to someone else's).

1.1 Background and Motivation

Having a template to work from provides a starting point. However, for a given project, a slight variation in the template may be required due to the nature of the given project. Further, the order in which the various chapters and sections will be written will also vary from project to project but will seldom start at the abstract and sequentially follow the chapters of the report. One critical reason for this, is that you need to start writing as early as possible and you will begin to write up where you are currently focusing. However, do not leave the abstract until the end. The abstract is the first thing anyone reads of an article or thesis — after the title; and thus it is important that it is very well written. Abstracts are hard to write so create revisions throughout the course of your project as your project progresses.

This introduction to background and motivation should state where this project is situated in the field and what the key driving forces motivating this research are. However, keep this section brief as it is still part of the introduction. The motivation will be further extended in chapter 2, presenting your complete state-of-the-art.

Note that this template uses italics to highlight where latin wording is inserted to represent text and the text of the template that we wish to draw your attention to. The italics themself are not an indication that such sections should use italics.

1.2 Problem description

Online shopping platforms retain large amounts of data regarding product sales and interest data. Despite this, it is not always easy to know what information this data might contain, and what it could mean to exploit and analyze this data. Analyzing data such as product interest or sales could help retailers discover useful information such as product trends or anomalies. While methods for attaining such information already exists to some degree, in the form of simple statistical methods or neural networks, there is still room for improvements.

This theses will attempt to come up with a solution in order to predict product trends using time series data of several product lines in groups. By doing this, we hope to increase the predictive accuracy of the model by using grouped time series with different corrolations, insted of only one series of data to which the statistical methods are limited.

More spesifically, we intend to look at the use of a modert time series predictive method, a *Convolutional autoencoder with LSTM* in order to make predictions. The main goal of the thesis is to improve the predictive forecasting abilities on e-commerce data in a time series using a modern deep learning approach.

1.3 Goals and Research Questions

A masters is a research project and thus there needs to be a question(s) that need answered. Such questions are often a very important part of the results that come out of the specialisation project. For those following the one year masters project, it is desirable to create such questions as early as possible as The formation of such questions provide both an important driving force for the masters project and provide clarity as to the goals sought. However, one will expect to refine the questions and thus the final path of the masters as work progresses. However any refinements should be conducted with care so as to avoid that the original aims, and previous work are not lost. It is always good to have one (or max 2) key questions and perhaps some sub questions.

Goal *Lorem ipsum dolor sit amet, consectetur adipiscing elit.*

Your goal/objective should be described in a single sentence. In the text under you can expand on this sentence to clarify what is meant by the short goal description. The goal of your work is what you are trying to achieve. This can either be the goal of your actual project or can be a broader goal that you have taken steps towards achieving. Such steps should be expressed in the research questions. Note that the goal is seldom to build a system. A system is built to to enable experiments to be conducted. The research question/goal would be the goal that the system is implemented to meet.

Research question 1 *Lorem ipsum dolor sit amet, consectetur adipiscing elit.*

Each research question provides a sub-goal and these should be precise and clearly stated enabling the reader to match your results to the original goals. They will also form the driving force for the experimental plan.

Research question 2 *Lorem ipsum dolor sit amet, consectetur adipiscing elit.*

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nam consequat pulvinar hendrerit. Praesent sit amet elementum ipsum. Praesent id suscipit est. Maecenas gravida pretium magna non

1.4 Research Method

What methodology will you apply to address the goals: theoretic/analytic, model/abstraction or design/experiment? This section will describe the research methodology applied and the reason for this choice of research methodology.

1.5 Contributions

The main description of the contributions will come in chapter 6.3 after the results are presented. This section just provides a brief summary of the main contributions of the work. This section can also be left out, leaving all discussions in chapter 6.3.

The format of this section will generally follow the following format: *Donec non turpis nec neque egestas faucibus nec id neque. Etiam consectetur, odio vitae gravida tempus, diam velit sagittis turpis, a molestie ligula tellus at nunc. Nam convallis consequat vestibulum. Proin dolor neque, dapibus a pellentesque a, commodo a nibh.*

1. *Lorem ipsum dolor sit amet, consectetur adipiscing elit.*
2. *Lorem ipsum dolor sit amet, consectetur adipiscing elit.*
3. *Lorem ipsum dolor sit amet, consectetur adipiscing elit.*

1.6 Thesis Structure

This section provides the reader with an overview of what is coming in the next chapters. You want to say more than what is explicit in the chapter name, if possible, but still keep the description short and to the point.

Chapter 2

Background Theory and Motivation

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nam consequat pulvinar hendrerit. Praesent sit amet elementum ipsum. Praesent id suscipit est. Maecenas gravida pretium magna non interdum. Donec augue felis, rhoncus quis laoreet sed, gravida nec nisi. Fusce iaculis fermentum elit in suscipit.

2.1 Background Theory

2.1.1 Time Series

> A time series is a sequence of data points that occur in successive order over some period of time.

Kenton [2020]

In a time series, time is often the independent variable. Examples of time series are weather data, stock markets, sound level samples. The time t usually ranges over a discrete index set and is often equally spaced.

Properties

A time series has several properties:

Legge inn illustrasjoner som forklarer bedre

Stationarity A time series is stationary if its statistical properties do not change over time. In other words, if it has a variance, mean, and covariance which is independent of time.

Rob J Hyndman [2014] defines stationarity more formally:

Definition 1 *X_t is a stationary time series $x_1, ..., x_n, if \forall_s \in \mathbb{R}$: the distribution of $(x_t, ..., x_{t+s})$ is equal*

Seasonality If the time series follows periodic fluctuations, like how electricity usage varies during 24 hours, then it has seasonality.

Autocorrelation If a time series has a strong autocorrelation then there is a big correlation between observations with a time lag between them.

Trends When a time series has a deterministic component that is proportionate to the time period it has a trend. In simpler terms, if a time series plot seems to center around an increasing or decreasing line it suggests the presence of a trend.

Cycles Cycles differ from seasonality because the period does not have to be fixed.

Level The level of a time series is equal to the mean. If a time series has a trend then the level is changing.

2.1.2 Forecasting time series

Let $Y = \{y_1, y_2, ..., y_n\}$ denote a time series. Forecasting is prediction the next time step y_{n+h} where h is the forecasting horizon.

There are two main categories within time series forecsasting. **univariate** and **multivariate**. An *univariate* time series consists of one input variable and one output variable. These methods use the time series past to predict its future. In a multivariate time series, there are many time dependent variables used as explanatory variables that all help predict one output variable.

Many time series methods focus on predicting just one step ahead $(y_n + 1)$. When forecasting many steps into the future this becomes a *Multi-step forecasting* problem. One simple to predict many steps ahead is to reccursively predict one step ahead, and use past predicted steps in the calculation.

Given a stationary time series, a naive approach to time series modeling is predicting that the next observation will be the mean of all past observations. A better approach is to define a smaller window, and apply the moving average across the whole series. Longer window size equals a smoother graph.

A different well-known technique is **exponential smoothing**. It uses the same approach, but assigns a different decreasing weight is assigned to each observation.

$$y = \alpha x_t + (1 - \alpha)y_{y-1}, t > 0 \tag{2.1}$$

Equation 2.1 shows exponential smoothing, where α smoothing factor that takes values between 0 and 1. It determines how fast the weight decreases with time.

TODO: Det er jo feil! Men virker so det er veldig mye met inn pÃ¥ veld liten plass. Vi har dÃ¥rlig plass sÃ¥ har rÃ¥d til Ã¥ b et par ord ekstra h du vil ;)

ARMA

Auto-Regressive Moving Average **ARMA** is a statistical model for time series prediction. It is one of the most commonly used methods for univariate time series forecasting [SOURCE] ARMA $ARMA(p, q)$is defined for stationary data and consists of two components $AR(p)$ and $MA(q)$.

The $AR(p)$ model is built on the assumption that the value of a given time series y_n can be estimated using a linear combination of the p past observations, an error term ϵ_n and a constant term c as seen in Equation 2.2 Ziegel et al. [1995].

$$y_n = c + \sum_{i=1}^{p} \phi_i y_{n-1+\epsilon_n} \tag{2.2}$$

where $\phi_i, \forall i \in \{1, ..., p\}$ denote the model parameters, and p is the order of the model.

The second part $MA(q)$ uses the past errors in a similar fasion Equation 2.3.

$$y_n = \mu + \sum_{i=1}^{p} \theta_i \epsilon_{n-1} + \epsilon_n \tag{2.3}$$

Here μ represents the mean of observations. q is the order of the model. $\theta_i, \forall i \in \{1, ..., q\}$ represents the parameters of the model.

Combining the past observations Equation 2.2 and past error terms Equation 2.3 we get the $ARMA(p, q)$ model in Equation 2.4.

$$y_n = c + \sum_{i=1}^{p} \phi_i y_{n-1+\epsilon_n} + \mu + \sum_{i=1}^{p} \theta_i \epsilon_{n-1} + \epsilon_n \tag{2.4}$$

SARIMA

SARIMA is an extension to ARMA model that supports the direct modeling of a seasonal component and incorporates a parameter d to transform a non-stationary time series into a stationary one.

SARIMA is a combination of simpler models to make a complex model that can model time series. The main idea is to apply different transformations to a nonstationary seasonal time series, in order to remove seasonality and any non-stationary behaviors. Utlaut [2008, p. 327-385]. The first part of SARIMA is the autoregression model $AR(p)$ where p is the maximum lag.

The second part is the moving average model $MA(q)$ where q is the maximum lag.

The third part is the order of integration $I(d)$ where d is the number of differences required to make the series stationary.

The final component is seasonality $S(P, D, Q, s)$, where s is the length of the season. s is dependent on P and Q which are equal to q and q but for the seasonal component. D is the number of differences required to remove seasonality from the series.

The combination of all these parts is the SARIMA model $SARIMA(p, d, q)(P, D, Q, s)$

2.1.3 Loss functions

Common Loss Functions

The most commonly used loss function for regression problem is the **Mean Squared Error (MSE)** function in Equation 2.5. It is the mathematically preferred function if the target distribution is Guassian. It punishes large errors much more harshly than smaller errors, due to its squaring of the error.

$$MSE = \frac{1}{n}\sum_{t=1}^{n} e_t^2 \tag{2.5}$$

If the target distribution consists of outliers, then the **Mean Absolute Error (MAE)** in Equation 2.6 is more appropriate as it does not punish the outliers too much.

$$MSE = \frac{1}{n}\sum_{t=1}^{n} |e_t| \tag{2.6}$$

e if this is needed.
ow we have it.
ve later if no need.

2.1.4 Deep learning

Deep learning is a field within artificial intelligence focused on the use of artificial neural networks in order to attain knowledge. These neural networks are trained and used in order to accomplish tasks such as classification, clustering, natural language processing, predictive tasks, and much more.

Supervised learning

Supervised learning is a sub-field of artificial intelligence where the focus is on training machine-learning methods through the use of labeled data. The method will attempt to access the data in correlation with the connected data label. Classical problems within supervised learning are classification problems where a prediction of a label is the desired result.

Unsupervised learning

Unsupervised learning focuses on creating artificial intelligence methods using data without a label. The data has no "correct" label associated with it. Applications of unsupervised learning might be clustering, anomaly detection recommendation systems.

2.1.5 Convolutional Neural Network

A *Convolutional Neural Network* (CNN) is a neural network architecture built using convolutional layers in order to extract information. Unlike fully connected neural networks, convolutional layers interpret data using perceptive fields. These perceptive fields evaluate only sections of the input at a time until the whole input is processed. The convolutional layers attempt to extract features from the input data. The first layer extracts low-level features, while the next layer extracts higher-level features, and so on.

[Géron, 2017, p. 443-446] Add image for visu

Multiple kernels, or filters, are used to extract features from the data. The result of applying these filters is known as the feature map, the extracted data features. These feature maps extract lower and higher lever features from the original data by extracting spatial features, retaining the spatial relationship within the data. Such spatial features could be the curvature of the ears of a dog in an image, or the correlation of timestep data in a time series. As a result, convolutional networks have several applications within image classification, image recognition, natural language processing, and time series analysis.

2.1.6 Recurrent neural networks

A *Recurrent Neural Network* (RNN) is an artificial neural network architecture that can work with data sequences of arbitrary length. Unlike feed-forward networks, RNNs consider the input data in conjunction with state information from a previous timestep. To accomplish this the network uses feedback connections. The feedback connections serve state information from the previous time step to the intended node. This connection works as a short-term memory for the recurrent layers, saving information from the previous time step memory cells.

In the basic RNN, these memory cells retain minimal information, saving data only from the previous instance. As the RNN memory cell is defined by the newest data introduced to the cell, previous information is encoded only in its effect on that data. Due to this, information is not stored for long in these memory cells, only retaining short-term memory data.

The RNN is able to process data of arbitrary length, meaning that it is well suited for natural language processing, time series analysis and similar problems. In addition, the memory retained in the RNN makes it well suited to extract temporal relations in the data.

nage of RNN
ck cells

In order to improve the performance of the RNN, new models have been created to address some of its shortcomings. One such model is the Long-Short Term Memory model (LSTM).

[Géron, 2017, p. 469-472]

2.1.7 Long-Short Term Memory

Long-Short Term Memory (LSTM) is a type of recurrent neural network addressing some of its shortcomings The LSTM introduces a new memory cell, adding Long-term memory to the network.

The LSTM memory cells are comprised of two vectors, one for long-term and one for short-term memory, as well as an input gate, output gate and a forget gate. The forget gate allows for the memory cell to remove unneeded parts of the memory in order to replace it with new data from the input gate. The long-term memory therefore retains some of its information, while replacing other parts.

The long-term memory of the LSTM enables it to solve the RNN problem of vanishing gradients. The long-term memory enables the LSTM to store data at arbitrary intervals, as well as detect long term dependencies in the data.

The LSTM, as the RNN, is well suited for working on series of data. The LSTM is able to analyze and predict long-term relations in a series of data, making it well suited for applications such as time series prediction or anomaly detection, as well as natural language processing and more.

STM cell archi-
e from the book!

[Géron, 2017, p. 492-493]

2.1.8 Autoencoder

Autoencoders are neural networks used to learn efficient representations of data. Through unsupervised learning, autoencoders do not need labeled data to function. Autoencoders accomplish this by lowering the dimensional complexity of the data, enabling it to store data representations more efficiently. Due to this ability, autoencoders are well suited for dimensional reduction of data. Autoencoders learn to represent data in a *coding*. This *coding* has a much lower dimensionality than the original data. [Géron, 2017, p. 506-508]

Autoencoders are composed of 2 parts; the encoder and the decoder. The encoder maps the input data representation to the *coding*, while the decoder maps the coding values back through data reconstruction. The encoder takes the input data and reduces the feature mapping to store the reduced data representation in the coding layer. The data is then sent from the coding layer to the decoder, where the decoder attempts to reverse the mapped data back to the original input data. By doing this, the encoder efficiently maps the most important data features in the coding layer, using far lower dimensionality than the original data. The decoder becomes efficient at reconstructing the input data using only the dimensionally reduced data in the coding layer. [Géron, 2017, p. 506-508]

Add image? Just somethin in GIMP

Through the feature mapping, the encoder becomes an efficient feature detector and extractor. Due to the reduced dimensionality of the coding, the encoder becomes sufficient at reducing the noise within the data, extracting the most important features. As a future extractor, autoencoders are well suited for use in pretraining neural networks, extracting the most important features. At the same time, autoencoders are able to become fairly successful generative models. As the decoder is proficient at reconstructing input data from the coding layer representation, meaning it can also generate new data. The reconstructive abilities of the decoder enable the generative model to create new data, similar to the training data used when creating the model. [Géron, 2017, p. 506-508]

2.2 Structured Literature Review Protocol

Here you need to include your structured review protocol including search engine, search words, research questions (for search, not the masters research questions), inclusion createrias and evaluation Criterias.

2.3 Motivation

Chapter 3

Related work

3.1 Loss functions

TODO: Write an introduction

3.1.1 DILATE

Both MAE and MSE are proven to be robust loss functions for regular regression problems, but a paper published by Guen and Thome highligts some problems when the shape of the target function matters such as in time series. Figure 3.1 shows three different fits to the same target function, that all have the same MSE error.

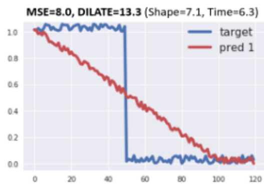
In order to battle this problem they introduce a new loss function they call **DILATE (Distrortion Loss including Shape and Time)** Guen and Thome [2019]. DILATE uses a Neural Network instead of a mathematical function. It aims at accurately predicting sudden chanes, and explicitly incorporates two terms supporting precise shape and temporal change detection.

The paper concludes that DILATE is comparable to the standard MSE loss when evaluated on MSE, and far better when evaluated on time and shape metrics.

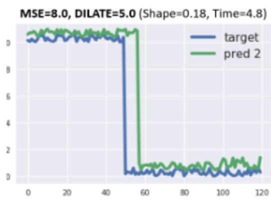
3.1.2 EVL

Ding et al. wrote a paper in 2019 regarding modern deep learning methods and their weak perfomance when applied to real world time series, because their inability to predict extreme values.

(a) Non informative prediction



(b) Correct shape, but with time delay



(c) Correct time, but inaccurate shape

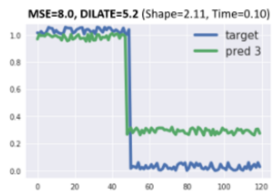


Figure 3.1: Three examples of different shapes but same MSE error [Guen and Thome, 2019]

Their deduction of why modern methods are unsatisfactory is bevause of the quadratic loss in the MSE loss function. They take inspiration from *Extreme Value Theory*, and develops a new kind of loss function which they call **Extreme Value Loss (EVL)**. They employ a memory network in order to memorize extreme events in historical records.

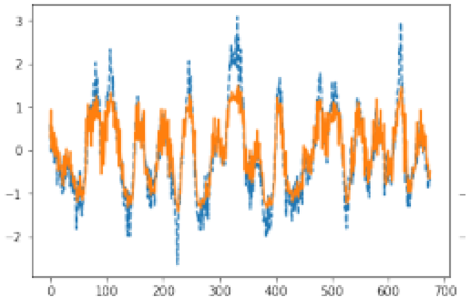
Figure 3.2 show two fitted time series from the papers results. They conclude that the method is superior to state of the art methods in extreme event detection, and in time series prediction.

litt mer detaljert
tte?

3.2 Multivariate time series

TODO

(a) Output from GRU



(b) Output from EVL model

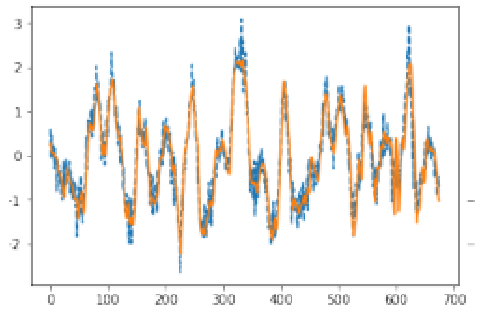


Figure 3.2: Figures from Ding et al. [2019]

3.3 Statistical methods VS Neural Nets

If a time series is stationary, then using its statistical properties is shown to be an effective and computationally cheap method Makridakis et al. [2018]. In a paper written by Makridakis et al. they test statistical methods versus machine learning methods for forecasting time series. They evaluate perfomance across multiple forecasting horizons using a large subset of 1045 montly time seires. Among the the machine learning methods they evaluate are

- *Bayesian Neural Network (BNN)*
- *Multi-Layer Perceptron (MLP)*
- *K-Nearest Neighbor regression (KNN)*

- *Recurrent Neural Network (RNN)* and
- *Long Short Term Memory neural network (LSTM).*

Among the statistical methods evaluated are

- *Naive2*
- *ETS*
- *ARIMA*

They evaluate with different loss functions and metrics, with simple one-step-ahead and multiple steps ahead methods.

Their findings are that all the simple statistical methods outperformed all the ML methods in terms of accuracy. The statistical methods also had a lot less model complexity and computational cost for training.

sMAPE?

The best performing model was ETS with a sMAPE score of 7.12. The second-best model was ARIMA with a score of 7.19. The third worst model was LSTM with a score of 11.67.

The paper Makridakis et al. [2018] highlights some of the drawbacks of ML methods. Especially, their complexity, their lack of explainability and their inability to show certainty in their predictions. However, the study has gotten some well-defined critic. The authors Cerqueira et al. points out that Makridakis et al. [2018] does their experiments on datasets of too-small sample size. Their largest time series sample size among the 1045 datasets is 144, and their smallest is 118. They hypothesize that these datasets are too small for machine learning methods to generalize properly.

Their contribution is doing a similar study on 90 univariate time series, in which all the datasets have a sample size above 1000. They evaluate statistical methods vs ML methods at different sample sizes, to test weather the sample size matters.

The paper Cerqueira et al. [2019] concludes that sample size does matter a whole lot for ML methods. Statistical methods outperformed ML methods up to around a sample size of 130. After that ML methods in general outperformed the simpler methods.

One drawback of the Makridakis et al. [2018] study is that it does not test any of our choicen ML methods.

- *A rule-based model (RBR)*
- *A random Forest method (RF)*
- *Guassion Process regression (GP)*

- *The Multivariate adative regression (MARS)*
- *Generalized linear model (GLM)*

This is good news for us, as our datasets does have a sample size of well above 660 at this moment. Depending on when we conduct the experiment, the number will grow to the range of 1095 to 1860, as data is being gathered every day as we speak.

The paper Bandara et al. [2019] points out a that in non-stationary time series, the distant past is typically less useful for forecast, as underlying patterns and relationships will have changed in the meantime.

3.3.1 Univariate vs Multivariate time series

Limitations of statistical methods

Skrive om limitations til statistiske metoder and * Univariate * stationary time series * Dealing with extreme values https://towardsdatascience.com/limitations-of-arima-dealing-with-outliers-30cc0c6ddf33

Pro ML: size matters Cerqueira et al. [2019]

From Guen and Thome [2019]: Traditional methods for time series forecasting include linear autoregressive models, such as ARIMA odel, and exponential smoothing, which both fall into the broad category of of linear state space models (SSMs). These methods handle linear dynamics and stationary time series (or made stationary by temporal differences). However the stationarity assumption is not satisfied for many real world tmie series that can present abrupt changes of distribution...

3.4 Convolutional autoencoders and LSTM

Chapter 4

Architecture/Model

Here you will present the architecture or model that you have chosen and that is (or will be) implemented in your work. Note that putting algorithms in your report is not desirable but in certain cases these might be placed in the appendix. Code further be avoided in the report itself but may be delivered in the fashion requested by the supervisor or, in the case of masters delivery, submitted as additional documents.

Chapter 5

Experiments and Results

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nam consequat pulvinar hendrerit. Praesent sit amet elementum ipsum. Praesent id suscipit est. Maecenas gravida pretium magna non interdum. Donec augue felis, rhoncus quis laoreet sed, gravida nec nisi. Fusce iaculis fermentum elit in suscipit. Donec rutrum tincidunt tellus, ac tempor diam posuere quis.

5.1 Experimental Plan

Trying and failing is a major part of research. However, to have a chance of success you need a plan driving the experimental research, just as you need a plan for your literature search. Further, plans are made to be revised and this revision ensures that any further decisions made are in line with the work already completed.

The plan should include what experiments or series of experiments are planned and what question the individual or set of experiments aim to answer. Such questions should be connected to your research questions so that in the evaluation of your results you can discuss the results wrt to the research questions.

5.2 Experimental Setup

The experimental setup should include all data - parameters etc, that would allow a person to repeat your experiments.

5.3 Experimental Results

Results should be clearly displayed and should provide a suitable representation of your results for the points you wish to make. Graphs should be labeled in a legible font and if more than one result is displayed on the same graph then these should be clearly marked. Please choose carefully rather than presenting every results. Too much information is hard to read and often hides the key information you wish to present. Make use of statistical methods when presenting results, where possible to strengthen the results. Further, the format of the presentation of results should be chosen based on what issues in the results you wish to highlight. You may wish to present a subset in the experimental section and provide additional results in the appendix.

Chapter 6

Evaluation and Conclusion

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nam consequat pulvinar hendrerit. Praesent sit amet elementum ipsum. Praesent id suscipit est. Maecenas gravida pretium magna non interdum. Donec augue felis, rhoncus quis laoreet sed, gravida nec nisi. Fusce iaculis fermentum elit in suscipit.

6.1 Evaluation

When evaluating your results, avoid drawing grand conclusions, beyond that which your results can infact support. Further, although you may have designed your experiments to answer certain questions, the results may raise other questions in the eyes of the reader. It is important that you study the graphs/tables to look for unusual features/entries and discuss these aswell as discussing the main findings in the results.

6.2 Discussion

In the discussion it is important to include a discussion of not just the merits of the work conducted but also the limitations.

6.3 Contributions

What are the main contributions made to the field and how significant are these contribution.

6.4 Future Work

Consider where you would like to extend this work. These extensions might either be continuing the ongoing direction or taking a side direction that became obvious during the work. Further, possible solutions to limitations in the work conducted, highlighted in 6.2 may be presented.

Bibliography

Bandara, K., Shi, P., Bergmeir, C., Hewamalage, H., Tran, Q., and Seaman, B. (2019). Sales demand forecast in E-commerce using a long short-term memory neural network methodology. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 11955 LNCS:462–474.

Cerqueira, V., Torgo, L., and Soares, C. (2019). Machine Learning vs Statistical Methods for Time Series Forecasting: Size Matters.

Ding, D., Zhang, M., Pan, X., Yang, M., and He, X. (2019). *Modeling extreme events in time series prediction*. PhD thesis.

Géron, A. (2017). *Hands-On Machine Learning with Scikit-Learn and TensorFlow*. O'Reilly Media, 1 edition.

Guen, V. L. and Thome, N. (2019). Shape and Time Distortion Loss for Training Deep Time Series Forecasting Models. *Advances in Neural Information Processing Systems*, 32.

Kenton, W. (2020). Time Series Definition.

Makridakis, S., Spiliotis, E., and Assimakopoulos, V. (2018). Statistical and Machine Learning forecasting methods: Concerns and ways forward. *PLOS ONE*, 13(3):e0194889.

Rob J Hyndman (2014). *Forecasting: Forecasting: Principles & Practice*. Number September.

Utlaut, T. L. (2008). *Introduction to Time Series Analysis and Forecasting*, volume 40. Wiley-Interscience.

Ziegel, E. R., Box, G., Jenkins, G., and Reinsel, G. (1995). *Time Series Analysis, Forecasting, and Control*, volume 37. John Wiley & Sons.

Appendices