POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

Executive Summary of the Thesis

Business Time Series Forecasting through Data Science

Laurea Magistrale in Management Engineering - Ingegneria Gestionale

Author: Valeria Maranesi

Advisor: Prof. Carlo Vercellis

Academic year: 2021-2022

Abstract

The advent of Data Age has come. The amount of data produced and managed every day is continuously, exponentially growing in almost every sector. Each organisation or company must deal with more and more data. Data is currently one of the most valuable resources that each company could own, since it can potentially generate many precious business insights.
Technologies are continuously evolving and, with them, the usage of data is constantly improving. Nowadays, we don't simply analyse data but we try to predict them, making machines learn from the past data. We try, in some ways, to "forecast the future", investigating over the possible connections between past and future data.
Here the Data Science finds place: it is able to manipulate large amounts of historical data to obtain these insights. The Forecasting Science is a clear example of technology at the service of the business: it supports companies of different markets in executing several core activities, such as organising the processes, managing the flows of both materials and information and monitoring the business KPIs. All these aspects of the business give to the decision makers precious information about the development of the business; thus, the Forecasting Science represents a way to make conscious, science-based and bias-free decisions.
The objectives of this thesis are to investigate over the most largely used mathematical methods that enable the time series forecasting and to identify the most appropriate ones in terms of predictions' accuracy. The aim is to evaluate them not only in absolute terms, but also in relation to the hyperparameters they assume. Moreover, the aim is to spot any eventual connection between the performances of each model and the main characteristics of the time series such as granularity, seasonality, trend, noise and autocorrelation.
The results clearly show that the choice of the model significantly impacts the accuracy of the forecasting: a good choice of the model is able to generate quite affordable predictions. The main evidences suggest the existence of a model that usually outperforms all the other algorithms. Moreover, the choice of the hyperparameters that fit each model strongly affects the performances: great attention should be put on the tuning process. At the same time, it is not possible to find any high correlation between the characteristics of the time series and the optimal model identified. Thus, a big effort should be put on the automatization of the whole process of model testing to try each (time series - model - hyperparmeters) combination and identify the optimal model tuned for each dataset. Finally, results

clearly evidence the possibility to obtain different conclusions - and therefore to make different decisions - depending on the choice of the accuracy metric.

1. Background and Business Objectives

Theoretical Background

In mathematics, a **time series** is a sequence $\{y_t\}$ of values assumed by a quantity under interest indexed in time order t.
The time series are usually characterised by three main properties, called **components**: trend, seasonality and random noise. We report the main ones, as explained by literature in [9].

- **Trend.** The trend component M_t is responsible for the long-time behaviour of the time series. The mathematical object which is responsible for the modeling of the trend component is called moving average $m_t(h)$, defined as the average of h successive values that the time series assumes.
- **Seasonality.** The seasonality component Q_t is responsible for the short-term fluctuations, which often present a regular frequency during the time span considered. It is mathematically represented by a periodic function.
- **Random noise.** The random noise component is able to model the irregular and unexpected fluctuations that a random variable usually has. In mathematical terms, the random noise is the time series $\{\varepsilon_t\}$, which represents the white noise, equivalent to a sequence of independent random variables which are normally distributed with mean equal to 0 and constant variance.

Regarding the **models** for time series forecasting, the most largely used are reported below.

- **Exponential smoothing models** class includes models that predict the future as a linear combination of a previous value and a shock. Exponential smoothing assumes that a series extends infinitely into the past, and that this influence of past on future decays smoothly and exponentially. This model, in its extended shape, can capture the main components of the time series [3].
- **Autoregressive (AR)** models are based on the idea of identifying possible relationships between the observations of a time series analysing the autocorrelation between observations taken at different time moments [2].
- **Stepwise regression** models use previous time steps as input variables and the next time step as the output variable. This technique is called **sliding window** method [5]. This procedure is iterated many times in order to obtain all the predictions. In particular, the following regressors proved to be among the most efficient ones for time series analysis [7]:
    - Linear Regressor,
    - Ridge Regressor,
    - Lasso Regressor,
    - Support Vector Machine (SVMs),
    - K-nearest neighbors method (k-NN),
    - Gradient Boosting Regressor.
- **Prophet** model is a procedure for forecasting time series data developed by the core data science team of Facebook [8]. It is an additive model that is able to handle non-linear trends, yearly, weekly and daily seasonality and holiday effects. In general, the standard mathematical shape of a **Prophet model** is similar to a Generalised Additive Model (GAM), a class of regression models with potentially non-linear smoothers applied to the regressors.

Finally, regarding the **metrics** that are used to **evaluate the accuracy** of the models, the most common ones are reported in the Table 1 with their mathematical formulations [4].

Metric	Formula
MAE	$\frac{1}{n}\sum_{t=1}^{n} \lvert y_t - f_t \rvert$
MAPE	$\frac{100}{n}\sum_{t=1}^{n} \frac{\lvert y_t - f_t \rvert}{y_t}$
MSD %	$\frac{1}{n}\sum_{t=1}^{n} (y_t - f_t)$
MSE	$\frac{\sum_{t=1}^{n}(y_t - f_t)^2}{n}$
RMSE	$\sqrt{\frac{\sum_{t=1}^{n}(y_t - f_t)^2}{n}}$

Table 1: Main accuracy metrics with their related mathematical formulations

Business Objectives

The objective of this thesis is to **analyse the most largely used forecasting models** to

identify the most appropriate ones for time series forecasting. Their performances in terms of accuracy are assessed not only in absolute terms, but also in relation to the hyperparameters they assume and in connection with the characteristics of the time series.
From a business perspective, predicting customer demand is a core objective for every company. It allows the enterprise to offer a better service and therefore to gain competitive advantage over the competitors: it then relies on the effectiveness layer. Moreover, all the internal activities can be better managed, correctly allocating the scarce resources such as materials, time and information. In this way, the waste is reduced and a higher level of efficiency is reached. Finally, **business insights** coming from data constitute a solid base that relies on the scientific method and that supports managers in any **decision-making process** [6].

2. Experiments settings

Regarding the methodologies that we follow in the experimental phase, in this section we report the main steps.
First of all, the analysis has been done on **two different datasets** containing data related to sales: the first one denoted as Pharma dataset is composed by **9 different sub series** while the second one, the Food Demand dataset, is made by **11 sub series**. This choice is due to the objective of analysing the models on different datasets.
As a first step, a deep **exploratory analysis** is performed in order to highlight the main characteristics of each time series in statistical terms.
Then, the **models** that are tested in the experiments are those mentioned in Section 1. Each of these models has its own **hyperparameters** that are tuned through the **Grid search optimizer**. Each couple (time series - model), with its optimal hyperparameters, is evaluated through an **accuracy metric**.
For what regards the metrics, some of them are evaluated at each step: the main ones are reported in Section 1. Moreover, the Adjusted MAPE is added to this list. It is defined exactly as the MAPE, since it enables the comparison of time series with different orders of magnitude, but it is computed only on those observations that are non null. This allows to avoid the problem of obtaining an infinite MAPE value in case of null time series observations.
Anyway, the **MAE** is chosen as **principal metric** to assess the goodness of each experiment (dataset - model). This choice is due to the very common use of this metric in literature and, moreover, in assessing the accuracy of different models evaluated on the same time series any magnitude problem arises. The other metrics are mainly used to make some benchmarks among the various series and models assessed.
Made these premises, we want to perform a set of experiments, where an experiment is defined as the combination of a dataset and an ad-hoc tuned model. The output of each experiment is the set of metrics mentioned above. Due to their very high number, we **automate** the whole experimental process.
We implement in Python an algorithm which performs the steps of tuning, forecasting and evaluating each model on each dataset, and whose main outline is reported below.

Algorithm 1 Experiments outline

```
define sub series, models and hyperparameter spaces
for each time series ts do
    import time series ts
    for each model m do
        preprocess data for m
        tune hyperparameters for m on ts
        fit model m on d
        predict future data
        evaluate predictions
        save metrics
    end for
end for
```

In particular, each step of this algorithm is in charge of a specific function which makes the whole process modular. For example, the `preprocessing` function executes the specific data preparation for each model. Instead, the `fit_predict_evaluate` tunes the hyperparameters, fits the model and evaluates the predictions.

3. Data Exploration

This section is dedicated to the execution of the preliminary analyses on both datasets to evaluate which are the most important features of

the time series and the technical characteristics they have.
First of all, each sub time series has been **decomposed into its main components** described in Section 1, and an analysis of the components is reported.
For what regards the Pharma dataset, in terms of trend every sub series seems to be stationary since any particular trend effect is visible and the Adfuller test [1] confirms the **stationarity** of all the sub series. Regarding the seasonality, a **weekly periodicity** is observable in all the time series, even if it is different in every specific case. The **yearly seasonality**, instead, presents **very different** behaviours depending on the sub series. For example, a time series presents an evident yearly seasonality, while for another one the seasonal behaviour along the year is almost absent.
Regarding the Food Demand dataset, every sub series presents a **yearly seasonal behaviour**, but it assumes a different shape in every case. The trend component of the time series is also very different in the cases under analysis. Indeed, we see a **monotonically increasing trend** in a couple of time series, while a much more variable trend in the other cases. In terms of stationarity, the behaviour of these time series is much variable: some of them are almost stationary, others aren't.
Then, a statistical analysis is performed in order to assess the main statistical indicators of each time series. In particular, the **coefficient of variation** is the most interesting one since it enables the comparison of the variability rate of the various time series. In general, the coefficient of variation is very different in the different series. In general, it assumes lower values in the Food Demand dataset than in the Pharma dataset. The values **range from 15% to 189%**.
Therefore, **many time series with different characteristics** are analysed so that it is possible to spot any eventual correlation between the results of this preliminary analysis and the accuracy of the forecasting.

4. Results and Evidences

In this section we report the main points that emerge as results of this research work and the main **takeaways** that should be considered in making time series forecasting.

Regarding the models, Table 2 and Table 3 report the results obtained with the optimal model and its best hyperparameters' configuration in terms of MAE and adjusted MAPE values.

Series	Opt model	MAE	MAPE
Aggregate	Prophet	13.71	28.83
M01AB	ExpSmooth	2.19	68.18
M01AE	Prophet	1.68	114.57
N02BA	Prophet	1.46	86.67
N02BE	GradBoost	8.85	33.87
N05B	ExpSmooth	3.22	63.09
N05C	ExpSmooth	0.70	49.06
R03	KNN	5.42	110.17
R06	Prophet	1.85	76.64

Table 2: The best model reported with its respective MAE and adjusted MAPE values for each sub series of the Pharma dataset

Series	Opt model	MAE	MAPE
Aggregate	Prophet	68816	9.35
TYPE_A	Prophet	41978	9.62
TYPE_C	GradBoost	14620	10.95
region_56	GradBoost	35357	8.58
region_93	Prophet	1373	27.42
meal_2290	SVR	11741	14.36
meal_2956	ExpSmooth	1486	25.11
city_473	SVR	742	8.32
city_713	Prophet	1447	9.75
email	Prophet	77449	187.34
homepage	Prophet	70128	48.45

Table 3: The best model reported with its respective MAE and adjusted MAPE values for each sub series of the Food Demand dataset

In general, what is observable is the dominance of the **Prophet model** and the Exponential Smoothing one in the time series related to the Pharma dataset. They perform optimally on those series which present a marked seasonality.

In these high frequency cases, also some regressors such as the SVR and the Gradient Boosting seem to have a good performance. For less granular time series, like those belonging to the Food Demand dataset, the optimal models are the Prophet, the SVR and the Gradient Boosting.
Regarding the regressors in general, in addition to the ones mentioned above, the K-NN is very good in terms of predictions' accuracy. Therefore, the **dominance of complex algorithms** seems clear in each case of analysis.
Regarding the models, in Figure 1 and Figure 2 we report the performances (MAE) of the various models in the analysed time series. In both cases, the chart is divided into sub charts in order to make them more readable, since the time series present very different orders of magnitude.

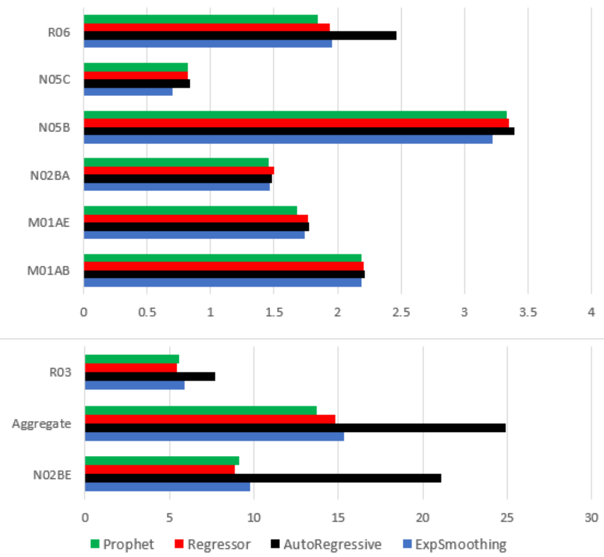

Figure 1: Plot of the performances of the models in terms of MAE - Pharma dataset

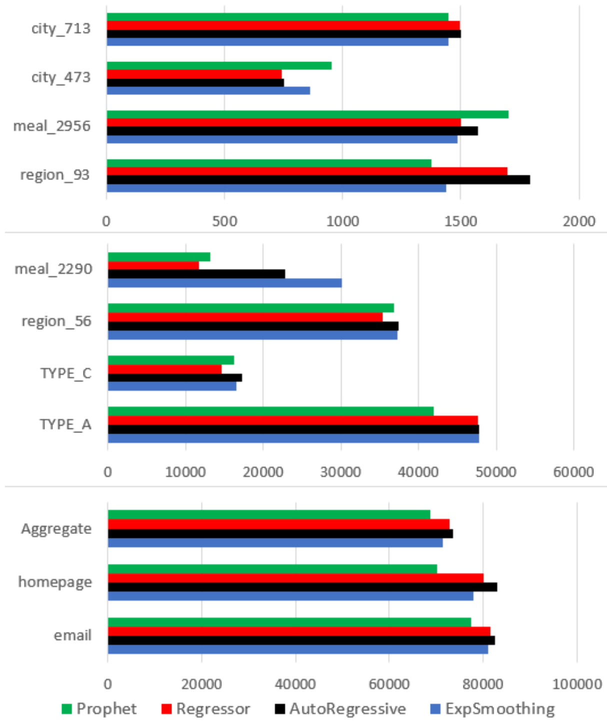

Figure 2: Plot of the performances of the models in terms of MAE - Food Demand dataset

What emerges from the analysis is the dominance of the **Prophet models**: even if the optimal model is not the Prophet, it usually obtains performances that are not so far from the ones of the optimal configuration.
Regarding the forecasts, we report below the plots of a good prediction and a bad one in terms of adjusted MAPE.

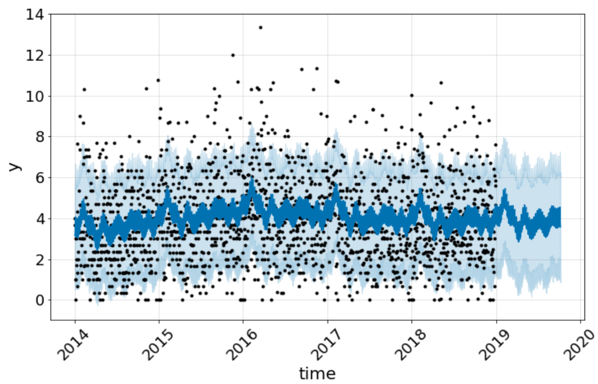

Figure 3: Plot of the *M01AE* time series and its future predictions

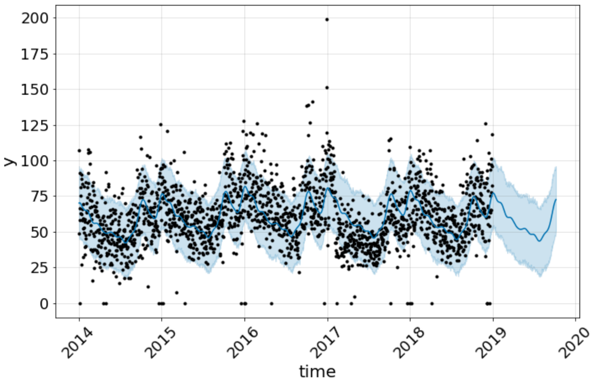

Figure 4: Plot of the *Aggregate* time series and its future predictions - Pharma dataset

The light blue band in these graphs represents the range between a lower and an upper bound that constitute an uncertainty interval and are responsible for the residual component of the time series (white noise). Then, we notice that the series which has the worst performance presents the thickest dark blue line and largest band.

Moreover, the **impact of hyperparameters' tuning** is analysed. Due to the considerations concerning the Prophet method, the impact of the hyperparameters' tuning is assessed on this model only. The improvement and degradation rates reported below are calculated referring to a baseline value, which is the MAE calculated fitting the model without any parameter.
What emerges from this analysis is that the choice of the best model improves the performance evaluated on the Pharma dataset by 5.3% and a bad choice of the hyperparameters degradates the performance by the 17.8% on average. Regarding the Food Demand dataset, we obtain an average improvement around the 19.7% and an average degradation rate equal to 42.4% with respect to the baseline.
The improvements and degradation rates registered for the two datasets are represented in Figure 5 and Figure 6.

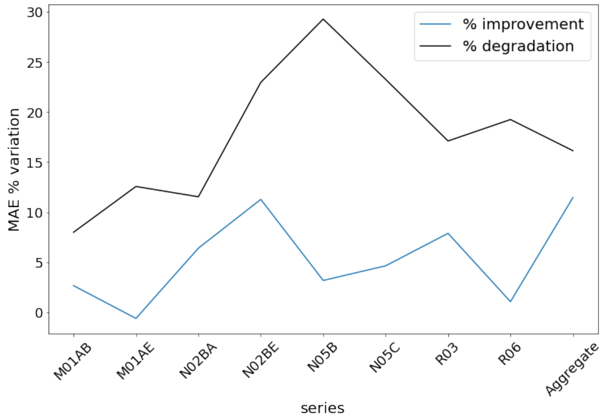

Figure 5: Plot of improvement and degradation rates - Pharma dataset

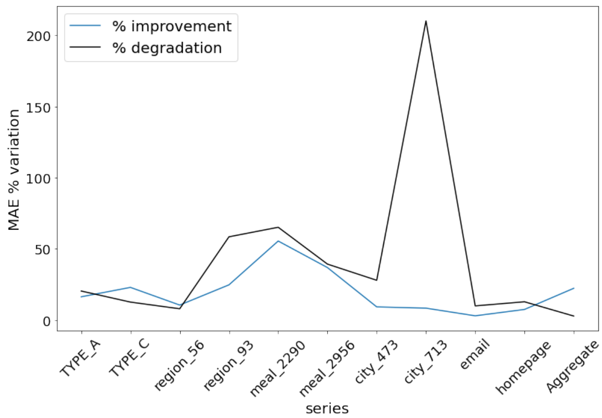

Figure 6: Plot of improvement and degradation rates - Food Demand dataset

As it is visible in the graphs, a big improvement does not always correspond to a big degradation and vice versa: this phenomenon cannot be spotted in advance. Moreover, in general, the **degradation rate** in case of a bad hyperparameter choice is **greater than** the **improvement rate** in case of an optimal tuning.
In general, the **impact of the hyperparameters' choice is remarkable**: thus, a great effort should be put also on the tuning process, consequently identifying the optimal configuration and obtaining the best results from each single model.
Another important point that must be highlighted is the **strong dependency of the predictions' accuracy from the metric** that is chosen to evaluate them: changing the metric, we can also draw different conclusions. This point should be appreciated looking at

MAE and adjusted MAPE's values in Table 2 and Table 3.

Finally, this kind of analysis is very **specific**: it strongly depends on the time series and its main characteristics: the results strongly depends on the time series under analysis. At the same time, it is very hard to find *a priori* a correlation between the best model in terms of accuracy and the main peculiarities of the time series, probably due to the **very complex** set of factors that drive the forecasting optimization.
Due to this consideration, a deep analysis must be made on each dataset in order to identify the best model and the optimal hyperparameters it should be tuned with. In particular, the best practice remains to test each model tuned with its optimal hyperparameters on each series. Due to the high number of possible combinations of a dataset and a model, the best idea is to **make the whole process automatic**, in order to test all the couples (dataset - model), to tune the parameters at each step and therefore to obtain the best performances.

5. Conclusions and Future Developments

The results of this analysis clearly show the strong impact that a good choice of the model, an effective hyperparameters' tuning and an automatic testing process have on the predictions' accuracy and confidence. In many cases, the results are satisfying and enable the availability of **reliable previsions**.
What emerges is the importance of investing in the Forecasting Science, trying to make even more accurate predictions. This can be enabled by trying more models, even the latest and most exploratory ones. Moreover, the implementation of **fully automated systems** for the trials of many models with many parameters configurations enables to save time and to choose the optimal model in a broader list. Nowadays, the use of **cloud-based services and cloud computing** is crucial to support the implementation and maintenance of very complex algorithms which are the most effective ones to pursue the time series forecasting objective.
From a business perspective, the great advantage of owning these accurate time series projections is remarkable. Thanks to forecasting, the planning of the company's resources in terms of time, money and materials sees a great improvement, thus enabling their optimal allocation and avoiding waste: the **improvement in terms of efficiency** is significant. On the other side, very precise sales predictions enables a good demand forecasting, which makes the company able to meet customers' requests in a better way: the **upgrade in effectiveness** is great too.
Being able to forecast the future with a sufficiently high level of confidence makes every strategic choice as if it were be based on real data. The **decision-making process** sees a huge gain, thus ensuring a strong **competitive advantage** to the company.

References

[1] Augmented dickey–fuller test.

[2] Autoregressive models, statsmodels documentation.

[3] Exponential smoothing, oracle documentation.

[4] Measuring forecast accuracy: The complete guide.

[5] A. K. Shrivas H. S. Hota, Richa Handa. Time series data prediction using sliding window based rbf neural network. *International Journal of Computational Intelligence Research*, 2017.

[6] Gracie Myers. How data analysis improve decision making, 2019.

[7] George Athanasopoulos Rob J Hyndman. *Forecasting: Principles and Practice*. PLOS One, 2018.

[8] Benjamin Letham Sean J. Taylor. Forecasting at scale. *PeerJ Preprints*, 2017.

[9] Carlo Vercellis. *Business Intelligence: Data Mining and Optimization for Decision Making*. Wiley, 2009.