Producing Economic Forecasts With Machine Learning Models

Otto Vintola* Ohto Makinen† Miro Keimiöniemi‡ Bach Pham§
Quan Tran¶ Minh Ha Le‖ Duc Chu**

December 2, 2024

Abstract

This is the work for the Data Science Project of Group 3. The goal of the project is to analyse and predict economic development in different regions of Finland, focusing especially in foreign trade. Development of GDP, foreign trade (exports), and population growth are modelled using public data which was available for a limited number of years between 2001 and 2023. Models used included linear, Bayesian probability, and Prophet models, and given the limited availability of data, the forecasts are based on utilizing previous trends of the chosen variables. When making economic forecasts, reliability and economic interpretability are both important factors and with the model choices we aim to find a balance between these factors. Based on the results, we found significant differences between the regions which could not be simply explained by any relevant feature variables, but are most likely due to a combination of a large number of factors. The forecasts are delivered through an interactive user interface that allows for a granular analysis of the regional differences and includes tools to estimate the effect of economic shocks to the future development. This project is implemented in collaboration with the OP Group as part of the Aalto University's course Data Science Project (CS-C3250).

*otto.vintola@aalto.fi; 1024221
†ohto.makinen@aalto.fi; 794879
‡miro.keimioniemi@aalto.fi; 100787595
§bachpq2003@gmail.com; 100553130
¶quan.tran@aalto.fi; 100644052
‖minhha.le@aalto.fi; 100860221
**duc.chu@aalto.fi; 100685794

Contents

1. Introduction

Economic development is one of the most studied topics in the economy and for a large bank economic prosperity of regions is an important factor to follow: banks offering credit to companies always bear the risk of default and the company not paying back, so it is in the bank's incentives to have the best possible estimate of the risks of a company. Regional differences are one factor determining this risk.

The aim of the project is to make forecasts of the development of the key economic indicators of Finland on an regional level. We have chosen population, gross domestic product (GDP) growth, and foreign trade as the key indicators determining the future success of a specific region. Making economic forecasts is complicated as they often are a sum of dozens of factors of the region (such as productivity, skilled labor force) but also factors of others regions and even other countries (cost competitiveness of key competitors, economic development of key export countries).

The dataset used was collected from public sources and the available time period was short. The lack of wide data sets increases the effect of random variation to the models and have to be taken into account when evaluating the results. Given these limitations, the results should be seen more as directional guides of the future outlook rather than exact estimates.

In section 2, we discuss the background of the project, and in section 3 we introduce the data. Section 4 discusses the models chosen for the analysis. Section 5 compares a public report of economic forecasts to our results. Section 6 is a discussion of the results, section 7 evaluation of ethical issues and section 8 self-reflection. Finally, section 9 concludes and section 10 discusses our group formation.

2. Background

Finland can be divided geographically in multiple ways: there are 4 main regions, consisting of 19 counties, which consist total of 69 sub-regions and finally, there are 302 municipalities in total [1]. The original project assignment was to make the forecasts and analysis on a municipality to level but due to the lack of key data, the project was decided to be implemented on a county/regional level. Figure 2.1 presents the geographical division of Finland into counties.

The difference between the financially most developed region Uusimaa (where capital region is located) and the least developed counties is smaller than in many other countries

[1] https://www.localfinland.fi/finnish-municipalities-and-regions

and globally. Still, highly populated Uusimaa producing 40% of the Finnish GDP yearly [1], behaves differently compared to other regions and is therefore in many cases analyzed separately in the models of the upcoming sections. Even when excluding Uusimaa, there are noticeable differences in Finland when it comes to unemployment, population growth, and economic growth. These differences are important factors when a bank assesses the riskiness of a possible business lender in a region and act as the main motivator for the work.

Producing reliable economic forecasts is a complicated task foreacasting. Growth as such originates from innovating companies aiming to either produce higher quality products or to produce the current set of products with lower resources (less workforce or material) [2]. In addition to domestic factors such as availability of workforce, and legislation, also international factors such as the market prices of key materials and demand in the global markets affect the performance and future outlook of a firm. In the work, instead of focusing on a firm-level development, we aggregate them on a regional level look at these domestic and international factors and aim to estimate the future development of the regions.

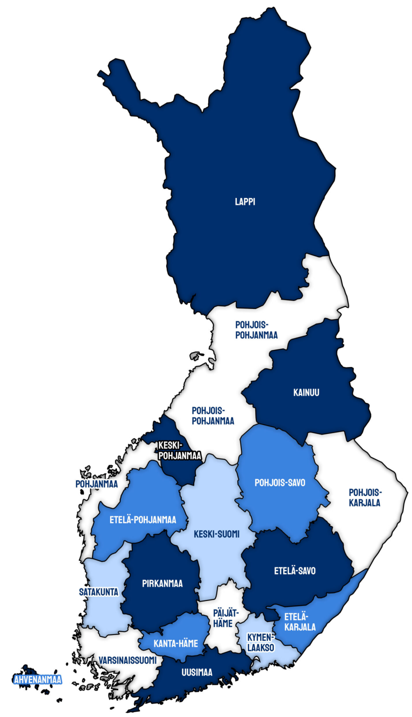


Figure 2.1: Geographical division of Finland.

3. Data

The data described in this section serves as the foundation for our analysis of the regional development of Finland. The section is structured so that it describes, the conceived datasets, the sources of the data, a part of the preprocessing steps, and the layout of the storage. So, the subsections are not in the temporal order of the project flow but rather arranged like this for logical ordering of the information. For now, taking into account that there are a few foundational datasets like the tax data, foreign trade by region, GDP per region, and company-specific data, is enough for the purposes of this section.

3.1. Sources

The main source of data for the project was Statistics Finland (Tilastokeskus) [3] which is a public institution producing official statistics of the Finnish society. Statistics Finland has a wide range of data sets regarding economic development, health and education-related statistics, and household-specific data.

One of the key factors we were interested in is the foreign trade (imports and exports) of the regions. This information was collected by Tulli (Finnish Customs) [4] and was available on a yearly level from 2015 to 2023. The values were reported at the end of the year, so in December for each region. Additionally, Verohallinto [5] and Kaupparekisteri [6] proved extremely useful for this project due to the granularity of their data. Combining the sources gives a close look at the economic state of each region and industry per year. The succeeding part discusses the datasets closely and describes each dataset and their crucial aspects.

3.2. Datasets

One of the foundational datasets was the tax data collected from Verohallinto [5]. It was crucial because it created the connection between companies and their incomes and locations. Through this dataset alone it is possible to calculate the taxable incomes of regions, which when normalized, is a measure of economic prosperity.

Furthermore, another important dataset regarding to our analysis is the company related data fetched from Kaupparekisteri [6]. The crucial aspect of this dataset is the connection from companies to their respective industries. This possibility permits the analysis to compare the sizes of industries in regions and study the impact of the top industries on regions.

Additionally, the data regarding imports and exports is fetched from Tulli [4] and it is crucial for forecasting the foreign trade of a region. In this dataset, each region has the value of imports and exports each year, coupled with features describing the change, such as cumulative change percentage over the years. This dataset was utilized as the training and validation for the foreign trade analysis.

Lastly, there is the GDP and population dataset which is retrieved from Tilastokeskus [3]. Each region has their respective population and GDP per capita by year. To note, no normalization of GDP is needed since it is normalized by population when using values per capita. The linear models for forecasting population and GDP were trained mainly using this dataset.

Tax data:

Field	Example
Tax Year	2022
CompanyID	1936189-3
Name	Tampereen Sammonparkki
Tax Municipality	837 Tampere
Taxable Income	588,41
Total Taxes Paid	117,68
Total Advances	58,32
Tax Returns	1085,12
Residual Tax	20,11

Company data

Field	Example
CompanyID	1234567-8
Business Registration Date	1978-03-15
Business End Date	2018-03-15
Company Name	Vähämäyry Oy
Main Business Line	Sawmilling
Business Line ID	51101
Company Form	Limited Company

Figure 3.1: Example rows from tax and comapny data. Using CompanyID (Y-tunnus) as the primary and foreign key, connecting industries, regions, tax years and incomes, is possible.

3.3. Preprocessing

Generally, the data from Vero, Statistics Finland, Kaupparekisteri, or Tulli is not made available in such a format that it would be usable out-of-the-box. Therefore, basic wrangling tasks are required in order to achieve a usable set of predictor and output variables.

However, the wrangling tasks are domain and problem-specific due to the uniqueness of the problems. For instance, for some problems – like time series analysis – it may be better to linearly interpolate missing values, rather than remove them. Conversely, for regression issues, it might be better to use an imputation technique that relies on the median or mode of the dataset.

As a definite task, however, fetching data from Kaupparekisteri through the open data API requires preprocessing. The data is returned as JavaScript Object Notation (JSON) [7] which has nested dictionaries. Consequently, the data needs to be converted to a Data Frame and subsequently flattened. Additionally, filtering out irrelevant information regarding our project i.e., the date for the inception of the company.

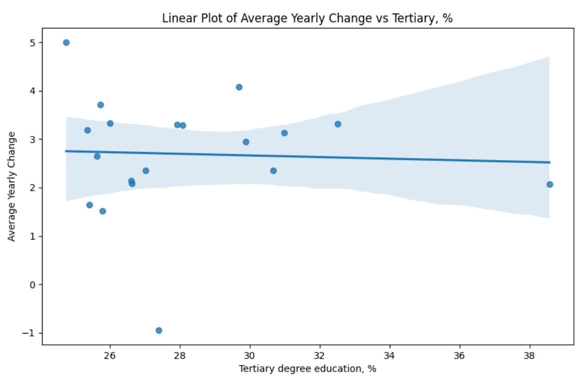
3.4. Storage

The data is stored in the shared repository which is cloned to personal workspaces during pulls. Concretely, the data is in comma-separated values(CSV) and Microsoft excel open XML format spreadsheet(XLSX) files in directories which are divided by theme. There is a folder for geographical, foreign trade, population, area and industry specific, and tax data.

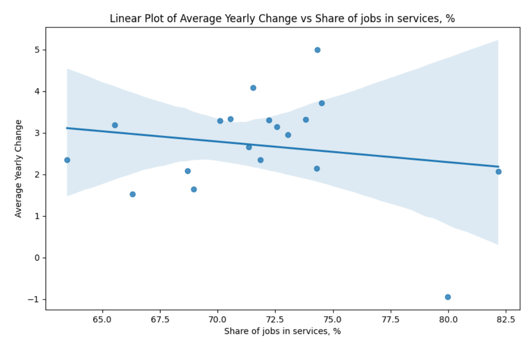
3.5. Preliminary Analysis

Before looking for viable machine learning models, we did some preliminary visual analysis of the relationship between possible features on economic growth (see 3.2). The mean of the average GDP growth of regions between years 2015 and 2021 is plotted against population statistics. We investigate whether high education could instrument innovation or share of services instrument creation of high output jobs indicating economic growth. Also, high share of working age population or urbanization (share of population in cities) could create clusters (concentrations of companies) where economic growth is faster.

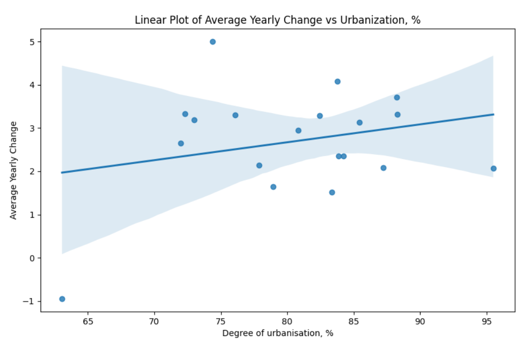
As we can see from the figure 3.2, the correlation between economic growth and these factors is weak. This highlights the previously discussed difficulty of trying estimate economic growth. Growth most likely originates from a wide set of factors of which some such as culture-related ones are not measurable. Wide range of different data sets allowed for extensive analysis beforehand, and next we will discuss the models chosen for predicting.



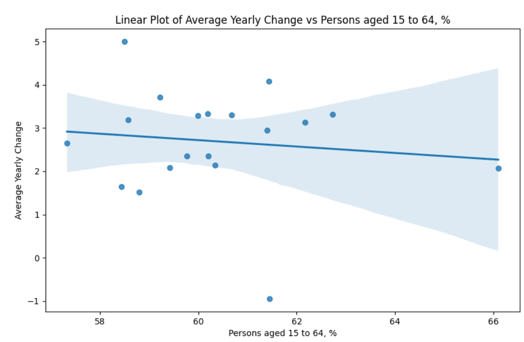
(a) Avg. growth on higher education



(b) Avg. growth on jobs in services



(c) Avg. growth on urbanization rate



(d) Avg. growth on working age population

Figure 3.2: Finding relevant features for explaining variation

4. Machine Learning Models

4.1. GDP and Population

4.1.1. Basic Linear Models

Due to the small amount of data points and the requirements for explainability, basic linear models were chosen as a baseline for modeling and predicting our key metrics. After visual inspection of the dataset, it was observed that among our key metrics, GDP and Population have a strong correlation with respect to year. So these two metric were suitable to be modeled with a simple linear regression model[2].

To explain some metrics, Lasso regression[3] was employed to predict GDP with industry data in an attempt to determine the most influential and significant industry in each region. However, due to the small amount of data points available with both industry and GDP data, the model fails to converge for some regions even with increased tolerances and was not included. Other linear models that are more robust to outliers such as Huber regression[4] and RANSAC regression[5] were also considered and tested, as shown in figure 4.1. However, due to the small amount of data available, Huber regression completely failed to model the trend of the data, and RANSAC performed nearly similar to just a simple linear model when accounting for significant events such as the 2008 market crash and the 2019 pandemic. As such, a final simple linear model was chosen to predict both GDP and Population data. All linear models used are from the Scikit-learn Python library [8].

To account for known and visible impact to the GDP metric within the dataset, mainly the 2008 financial crisis and the 2020 pandemic were considered in the model with descending impact as they are realized over time. Since the data are collected as averages in each regions per year, only major economic events have a noticable impact on the data. Unfortunately, the effects of other minor events such as a paper mill closing, as described in the initial project description, cannot be identified from the data and were not modelled.

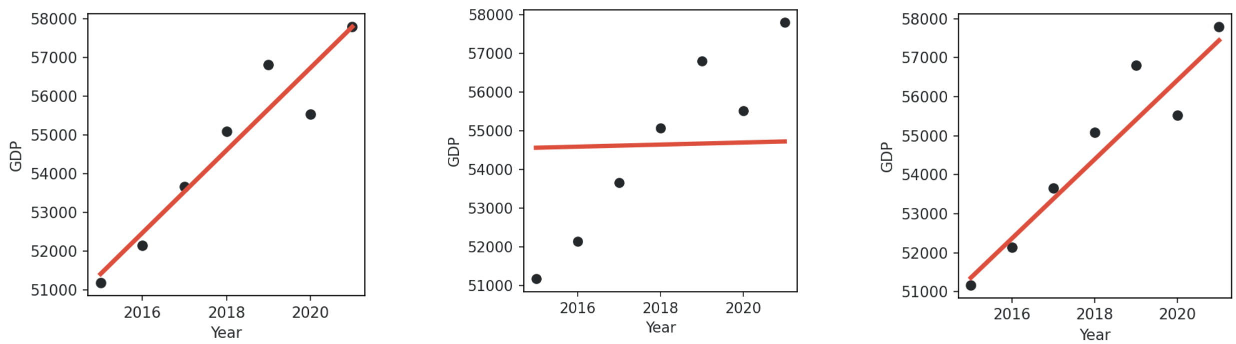
The final linear model for GDP was fitted with the GDP figures from the year 2000 up until 2021, excluding the year 2009 and 2020 for reasons mentioned above. Then, to account for and simulate the effects of these major financial events, the slope of the model was applied to the year before the event happens, and the difference between the predicted

[2] `https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html`

[3] `https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.Lasso.html`

[4] `https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.HuberRegressor.html`

[5] `https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.RANSACRegressor.html`



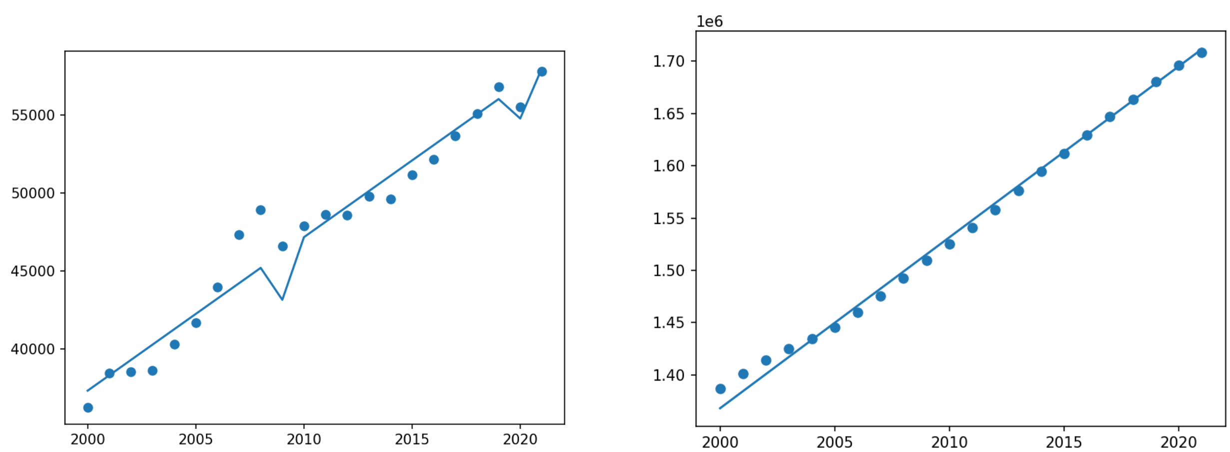
(a) Linear Regression (b) Huber Regression (c) RANSAC Regression

Figure 4.1: Comparison of different linear models on the GDP dataset in the Uusimaa region.

GDP and actual GDP was calculated as a fraction. This fraction is then multiplied with the result of the linear model as a modifier when an major financial event would be simulated. This process was done separately for both the 2008 financial crisis and the 2019 pandemic. The final model follows the formula:

$$y_i = (\beta_0 + \beta_1 year_i) \times \theta_f^{\texttt{finacial\_crisis\_effect}_i} \times \theta_c^{\texttt{covid\_effect}_i} \tag{4.1}$$

where θ_f and θ_c are precalculated modifiers for the financial crisis and pandemic respectively. And $\texttt{finacial\_crisis\_effect}_i, \texttt{covid\_effect}_i \in [0, 1]$ are variables indicating the severity of the event. They take the value of 0 if there are no major event, 1 if the event are similar to the financial crisis or the pandemic, and in between 0 and 1 to indicate recovery or a minor financial event. These final models are depicted in figure 4.2.



(a) Final GDP linear model. (b) Final Population linear model.

Figure 4.2: Final linear models for GDP and Population on Uusimaa.

4.2. Foreign Trade

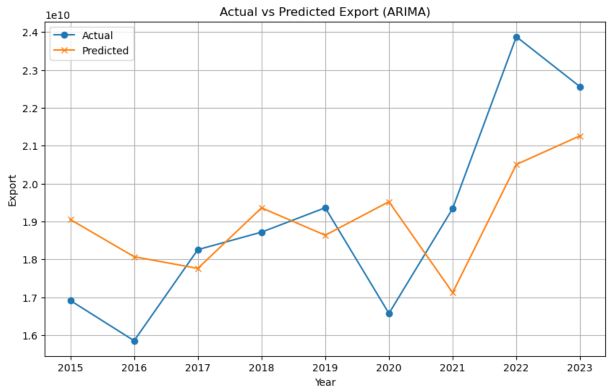
4.2.1. Frequentist Time Series Models

Time series models such as Autoregressive (AR), Moving Average (MA), and Autoregressive integrated moving average (ARIMA) are popular statistical models to analyze and forecast time series given adequate historical data [9]. These regression methods are able to capture trends and other patterns in the data. As they are univariate time series models, each region will have one time series model. Uusimaa, which is identified as the outlier and the most developed region (as further discussed in the Background section), is used in this example. As such, the foreign import data of Uusimaa [4] was used to identify its most appropriate time series model. Initially, an ACF test with a 0.05 significance value is used to reveal that the data is not stationary, which suggests the data might require differencing [6]. Subsequently, the Python 'pmdarima auto arima' library allows for efficient model selection by performing stepwise search to minimize AIC (Akaike information criterion) [7]. The best model was found to be ARIMA(0,0,1) with intercept.

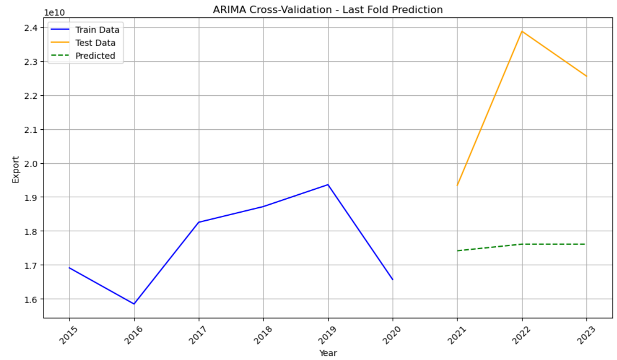
It can be seen that with an ARIMA(p, d, q), we currently have $p = 0$, which means the autoregressive component is not used. This suggests that the trend component of the time series is negligible or does not significantly contribute to its behavior, meaning no prior values significantly affect the current value in the model [9]. In other word, there is no significant partial autocorrelation between the current and past values. The $d = 0$ also indicates that no differencing is required, as the series is already stationary. A stationary series has a constant mean and variance over time, without trends or seasonality, so no adjustments are necessary to stabilize it [9]. However, this also contradicts the initial ACF test. The $q = 1$ implies only the previous error term ("shock") affects the model [9]. This makes it a moving average model of order 1, where only the most recent error impacts the current value. This also suggests that the time series has a short-term randomness component, with no significant correlation in the values after the first lag, rather than long-term correlation [9]. So in the end, the best model is a MA(1) time series model.

The trained model was used to predict Uusimaa's foreign export data, which spans from 2015 to 2023, corresponding to 9 annual observations. The coefficient of determination, R^2, was used as a metric to evaluate the proportion of variance explained by the model. An in-sample prediction gave a R^2 score of 0.35, indicating limited explanatory power. To better understand the model's utility for future forecasting, we also evaluated its performance using a time series cross-validation approach [8]. Ideally, the validation would involve a rolling

[6] https://www.statsmodels.org/dev/generated/statsmodels.tsa.stattools.adfuller.html
[7] https://alkaline-ml.com/pmdarima/modules/generated/pmdarima.arima.auto_arima.html
[8] $https : //scikit - learn.org/stable/modules/generated/sklearn.model_selection.TimeSeriesSplit.html$



(a) In-sample Prediction of MA(1) model on Uusimaa Foreign Export



(b) Cross-validation of MA(1) model on Uusimaa Foreign Export

window approach, where the model forecasts one year at a time, incrementally including previously predicted timeframes in training [10]. This would allow us to calculate the R^2 score at each step, providing a clearer picture of performance over time. However, due to the small sample size, a minimum split of 2 was used, as R^2 score is not suitable for fewer than 2 observations in each split. Additionally, while using the cross-validation function, there were too few observations to estimate starting parameters for ARMA and trend, and all parameters except for variances was set to zeros. This gave a negative R^2, which suggested that it performs worse than simply predicting the mean.

In conclusion, due to the poor performance of the model and the limited sample size, the time series model was not used in the final visualization and highlighted the need for alternative approach to forecast foreign trade.

4.2.2. Bayesian Forecasting and Inference

This section explores means to extract insights from the international import and export of every sub-region of Finland. Instead of modeling the absolute import and export values, we aimed to forecast their growth rates over the years. This approach offers various benefits. Firstly, growth rates often produce a stationary series, which simplifies modeling and meets the assumptions of many time series techniques. This way, we could also reduce autocorrelation in the series, which could lead to more accurate and stable models, especially for ARMA-type models. Moreover, predicting growth rates can sometimes yield more accurate forecasts because they account for underlying trends without the bias of absolute scale, especially for large quantities such as imports and exports.

In this section, we attempted to provide the answers to the following three research questions with Bayesian statistics:

1. Are growth rates consistent across major regions?

2. Did COVID-19 & the Russian invasion of Ukraine impact the growth rates?

3. Can growth rates be forecasted accurately?

Are growth rates consistent across major regions?

To address this question, we fitted a hierarchical model with unequal variance and hierarchical priors for the mean and the variance. Since the model's fit was good, we inspected the posterior distribution of the random effects parameters as follows

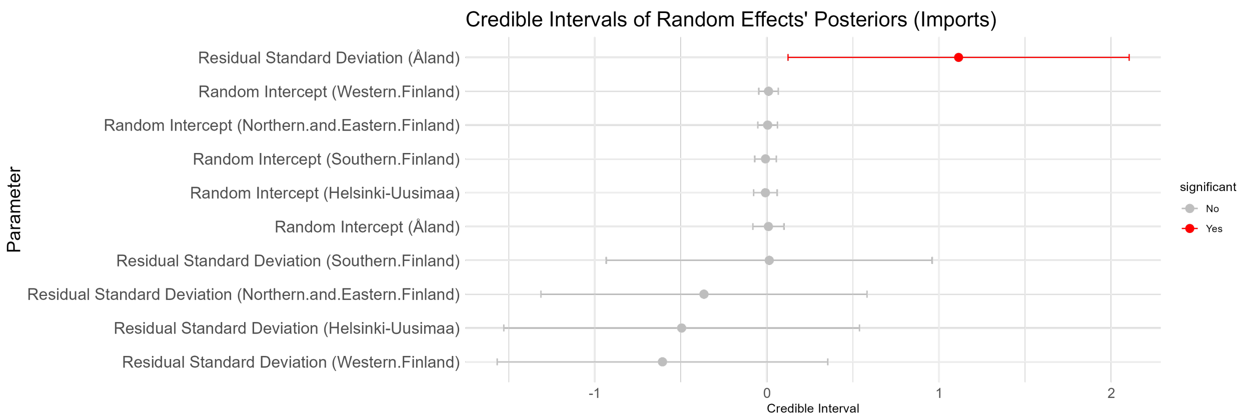


Figure 4.4: 95% Credible Intervals of Random Effect Parameters for Imports

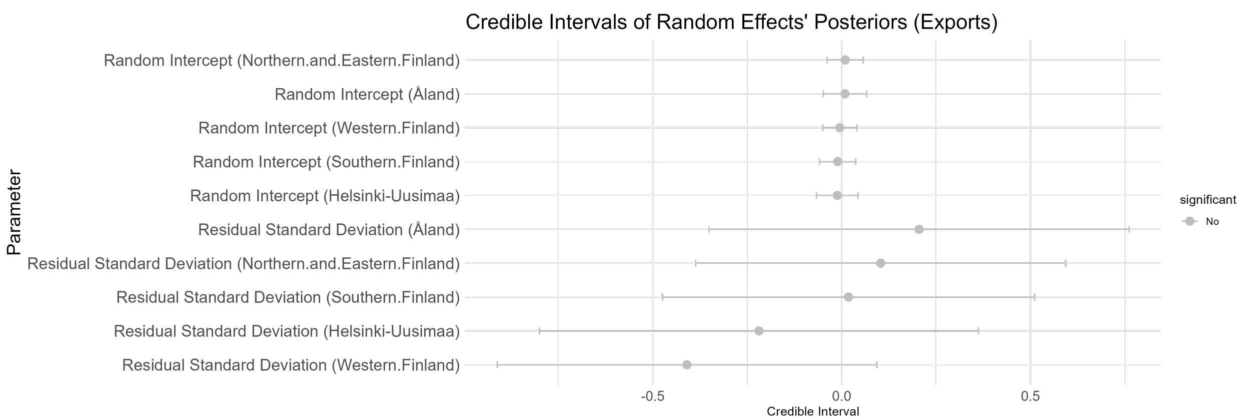


Figure 4.5: 95% Credible Intervals of Random Effect Parameters for Exports

In this hierarchical model, the residual SD represented the variability of the response variable (growth_rate) within each group (here, major_region). It quantified how much individual observations within a region deviate from that region's estimated mean growth rate. Here, the residual SD for Åland imports growth rates was significantly higher than the global residual SD, indicating that there was more variability in the growth rates of imports within that region. This could be explained by the unique geographical and political characteristics of Åland that may contribute to the heterogeneity.

Did COVID-19 & the Russian invasion of Ukraine impact the growth rates?

To assess the impact of significant events on growth rates, we introduced Bayesian Difference-in-Differences Regression models using dummy variables $\texttt{covid\_effect} = 1_{\texttt{year} \in \{2021,2022\}}$ and $\texttt{war\_effect} = 1_{\texttt{year} \in \{2022,2023\}}$:

- COVID-19 Impact Model: $y_i = \beta_0 + \beta_1 \texttt{year}_i + \beta_c \texttt{covid\_effect}_i + \epsilon_i$
- War Impact Model: $y_i = \beta_{0'} + \beta_{1'} \texttt{year}_i + \beta_r \texttt{war\_effect}_i + \epsilon_i$

where β_c and β_r represented the impact of the COVID-19 and Ukrainian war on the growth rate, respectively. It turned out that the impacts of these events were not statistically significant on the growth rate of imports and exports for every region as the 95%-credible intervals of β_c and β_r did include zero.

We also tested out Google's Bayesian structural time-series (BSTS) with its `Causal Impact` package [11]. However, the models did not fit well enough to make any inference.

Can growth rates be forecasted accurately?

For each sub-region and target variable, we fitted 25 Bayesian models using the `brms` package [12], including:

- Four combinations of Linear Regression, Gaussian Processes, and Splines.
- Nine ARMA(p, q), with $(p, q) \in \{1, 2, 3\} \times \{1, 2, 3\}$.
- Twelve AR(p') and MA(q'), with $p', q' \in \{1, 2, 3, 4, 5, 6\}$.

In total, we trained $25 \times 2 \times 19 = 950$ models. These models were coupled with sets of weakly-informative priors of the parameters, namely $\mathcal{N}(0, 1)$, $\mathcal{N}(0, 10)$, and Cauchy$(0, 2.5)$. Additionally, we also employed the frequentist Holt-Winters model and Naive(n) models, which used rolling means with a window of size n ($n \in \{3, 4, 5, 6\}$) to predict future values, serving as references and baselines. All of these models were trained using data from 2016 to 2021 and evaluated based on the R^2 metrics calculated on the data in 2022 and 2023. The model with the highest R^2 value for each region and metric was selected as the best-performing model. This simple validation strategy was adopted since we do not have adequate computational resources to carry out more sophisticated approaches such as cross-validation for this number of Bayesian models. Likewise, since we needed to save as much data for training as possible, we could not estimate the performance of the best models. The best we could do was to make the assumption that the expected performance would be comparable to the validation result.

The following is a visual validation of two Bayesian models MA(4) and MA(5) out of 25 models forecasting the growth rates of imports of the Satakunta region (see Figure 4.6). This is followed by the selected models for the Keski-Suomi region (see Figure 4.7).

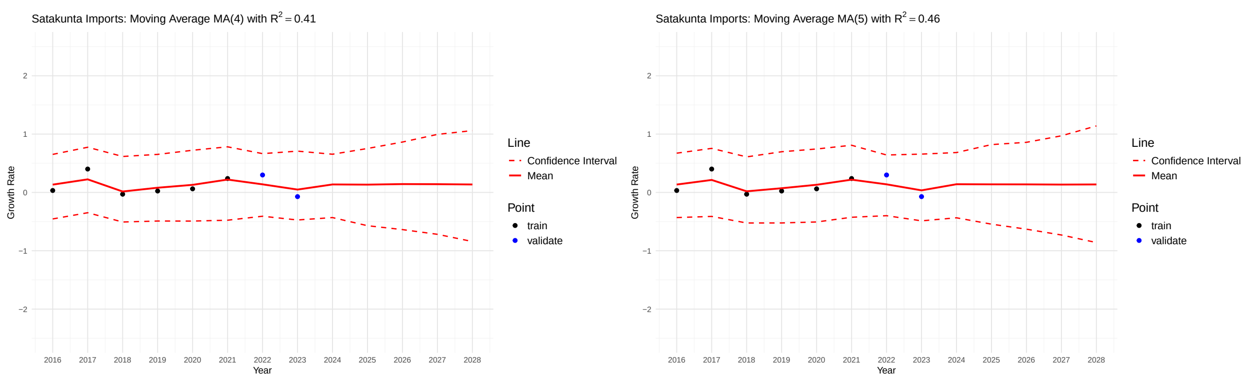


Figure 4.6: The solid line is the mean of the predictive posterior distribution. The dash lines are the 95% credible intervals. Black points are observations used for training models, while blue ones are used for validation.

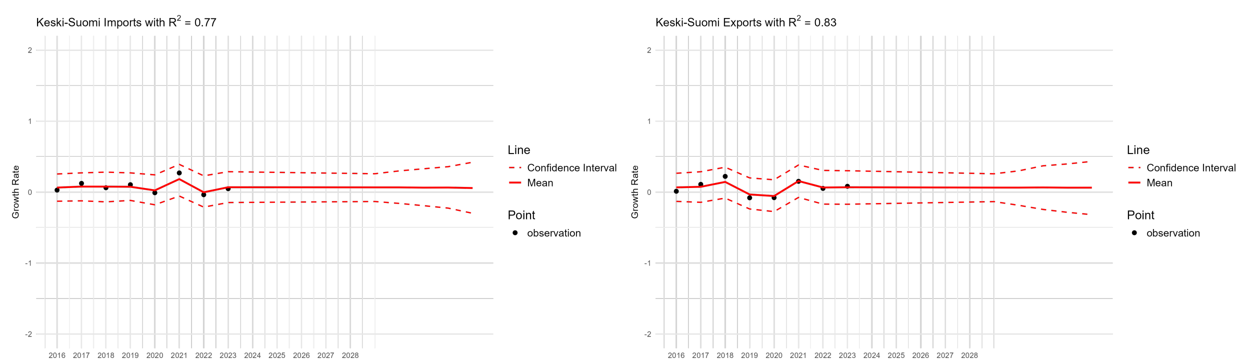


Figure 4.7: Best performing models for the growth rates of imports and exports of Keski-Suomi region.

4.3. Prophet Models for Industry and Growth Forecasting

The prophet model [13] is a regression model for time series forecasting developed by Meta. The model $y(t)$ consists of three main parts,

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t, \tag{4.2}$$

$$g(t) = \frac{C}{1 + \exp(-k(t - m))} \tag{4.3}$$

where $g(t)$ represents the non-periodic fluctuations in the data, while $s(t)$ is the periodic variance, such as annual or monthly occurrences, and $h(t)$ represents the effects of irregular schedules of holidays. The ϵ_t models the changes which are not captured by the aforementioned components.

The prophet model is used as a tool to cluster the regions in terms of their economic forecast. Initially, the GDP data is used to create the relative change for each year per region y_t with,

$$y_t = \frac{GDP_t - GDP_{t-1}}{GDP_{t-1}}. \tag{4.4}$$

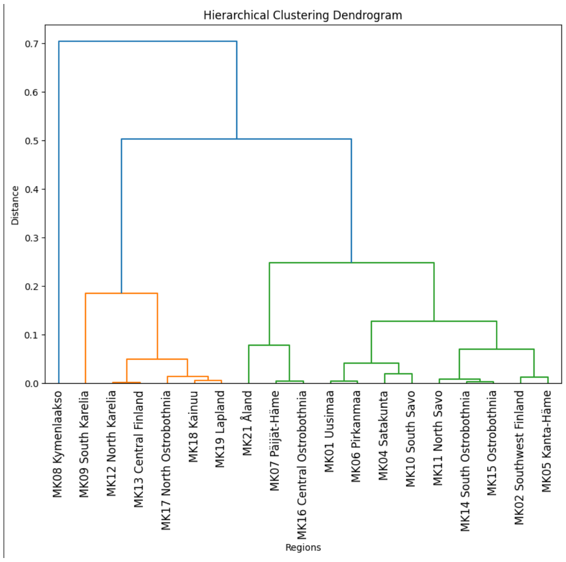
Even though the response variable is a time series variable itself, examining the autocorrelation function (ACF) [14] for each region's response variable[9], revealed no statistically significant autocorrelation. Afterwards, the Prophet model is fitted to each region's relative change presented in 4.4 in order to research the direction rather than position. However, the estimates are not of interest. Instead, the growth coefficient k—which governs the steepness—from $g(t)$ in 4.3 is extracted and used for Ward hierarchical clustering [15], which is defined by the Euclidean distance $\|X_i - X_j\|^2$. The aforementioned procedure is equivalent to comparing the slope of the growth rates throughout the years. Figure 4.8a shows the clusters formed with Ward‘s hierarchical clustering, while Figure 4.9 shows the forecasted values of y_t with Prophet for each region as a function of time.

The reason for using Prophet for this task, was the interpretability, ease-of-use, and predictive power of the model. The interpretability, especially, becomes evident when examining the clusters, which have a clear interpretation from Figure 4.9. Aside from Kymenlaakso—which is an outlier—the regions are separated into, either, growing or stagnant, or declining ones. Kainuu-Lapland are growing or stagnant, while Uusimaa-Åland are declining.

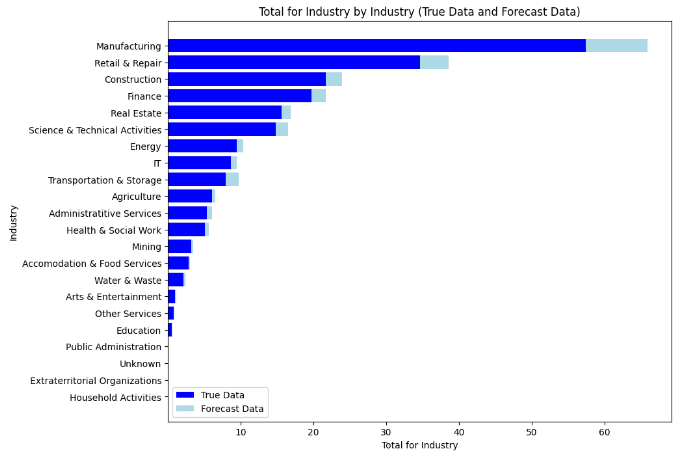
To forecast the top future industries, NeuralProphet is employed[10]. As a consequence, the GDP and population of each region is also included as the features. The choice for this was due to the hypothesis that GDP and population affect the outcome of the industry

[9]The ACFs are not presented due to the page limit.

[10]NeuralProphet's equation is not presented, again, due to the page limit.



(a) Ward's hierarchical clustering.



(b) Industry classification results.

Figure 4.8: Results of clustering and industry classification.

specific forecast heavily. Moreover, observing the training mean squared error (MSE) was slightly higher with Prophet than with its neural counterpart—namely, the average MSE across regions with NeuralProphet was around 0.06, while for Prophet it was 0.08. However, making informed decisions with a small dataset with respect to the observations, is rather tricky, because the big picture might not be well represented. Thus, the decision to use NeuralProphet for this task was also fueled by curiosity.

Forecasting the top industries in the future, is done by fitting an independent NeuralProphet model on each region and industry, where industries represent percentage shares. Subsequently, making predictions regarding the future relative value of that industry and then sorting them. Figure 4.8b shows the sum of normalized values of the regions of each year, where the retrieved data—before 2023—is highlighted in dark blue, while forecasted data is in light blue.

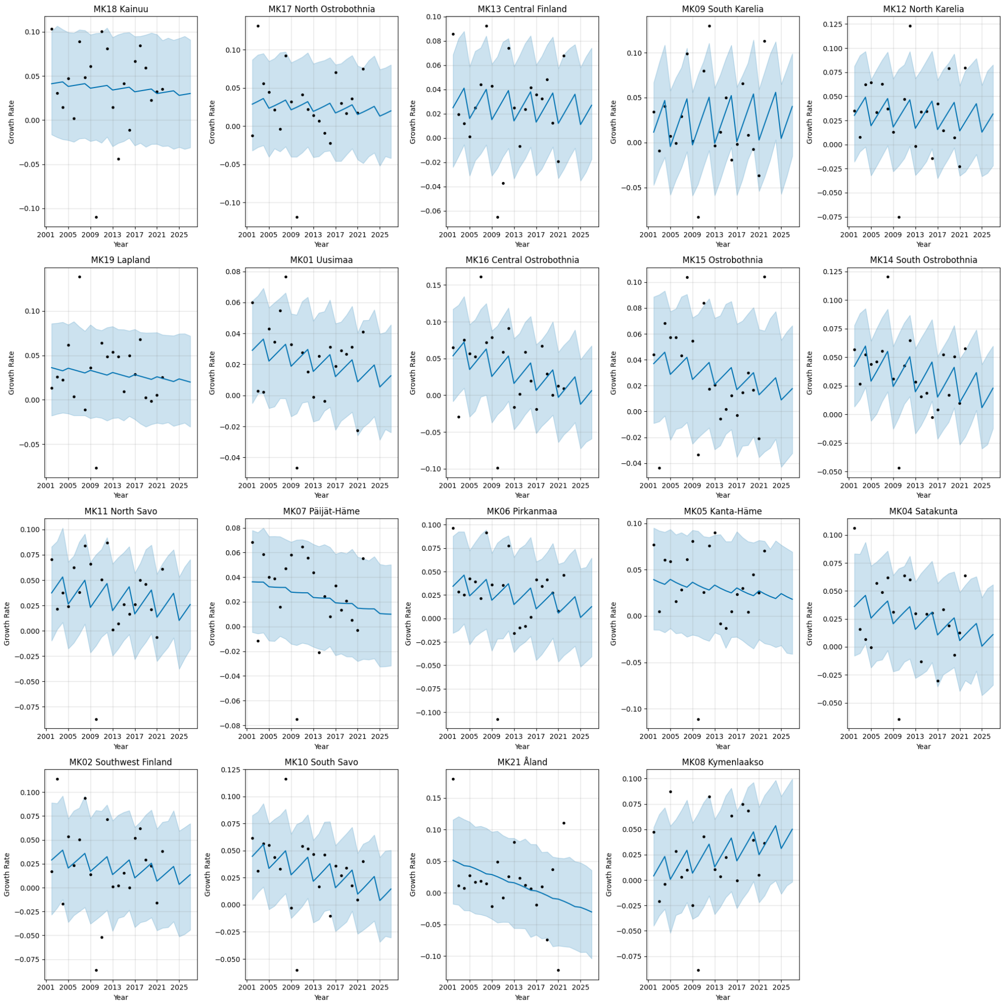


Figure 4.9: Forecasts for the growth rate of regions.

5. User Application

OP's requirements for the project specified an "interactive, visual (map) tool allowing users to explore and compare regional differences and make insight about regional development" with the requirement of the user being able to choose two or more regions and compare them. We used Flutter [16] to implement a GUI, shown in Figure 5.1, which shows the data and predictions on a regional level with 19 separately selectable regions (maakunta).

The user can filter the information displayed on the map on an intuitive color range. The color (from red to green) for each region is then determined with respect to the average of the metric over all regions (or 0 for growth rates) if no region is selected. In the case of population, for example, white signals that the population of the region is near $\frac{TotalPopulation}{19}$.

The shade of green (red) communicates the distance to the mean in the positive (negative) direction with a darker shade being further away.

If a region is selected, it is highlighted in blue and the colors of the other regions are adjusted with respect to the selected region instead of the overall average for insightful visual comparison. The user can select multiple regions, which are indicated by a lighter blue overlays, but these do not adjust the initial reference point, which is highlighted in darker blue. Blue outlines highlight either the selected cluster (Similarity cluster) or those regions whose top industries contain the chosen industry (Filter by industry), both of which can also be chosen from the "Filters" section. In addition showing data on the map, the UI provides a table and graph options for comparing regions on a selected year (chosen with a slider) or during the whole available period. The "+" on top of the year indicator can be used to select an event that is simulated at the chosen point. The full demo can also be accessed using this link, which directs to the OP group presentation.

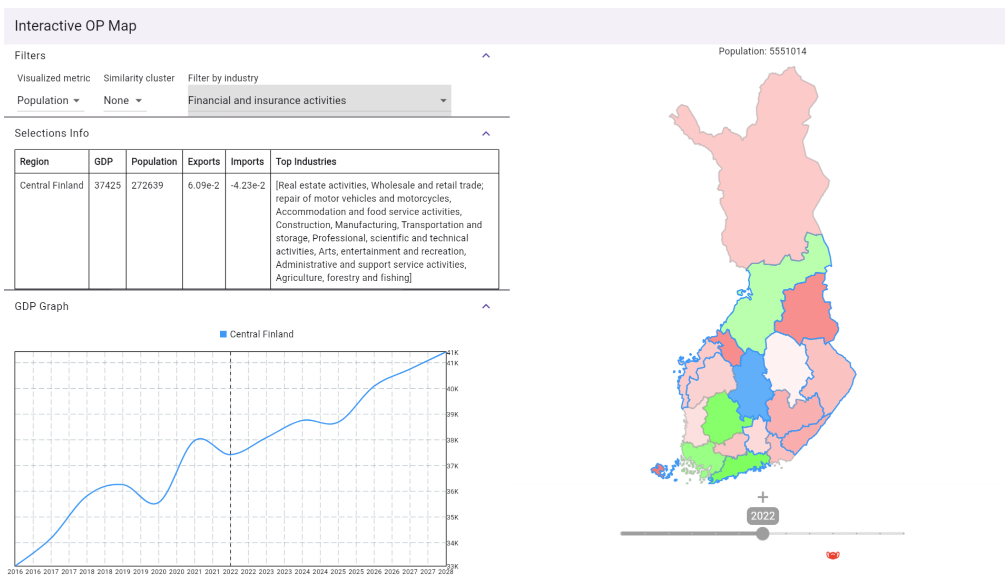


Figure 5.1: Interactive OP Map

6. Evaluation of Results

6.1. Insights from the GUI

The GUI turned out to be genuinely intuitive to use and to interpret and its output can be highly insightful with respect to the caveats on the amount of data and its quality as discussed in sections 3. Data and 4. Machine Learning Models. Below is a list of some of the most interesting observations made from using the GUI:

- Lapland's GDP per capita overtakes that of South Karelia by 2027
- South Savo's GDP per capita overtakes that of North Ostrobothnia by 2028
- North Ostrobothnia has the lowest predicted GDP growth in absolute terms, whereas Uusimaa has the highest
- The economically strongest regions with respect to GDP per capita are Uusimaa, Åland and Ostrobothnia in that order
- Population distribution will continue to concentrate even further in Uusimaa, Pirkanmaa, Southwest Finland and North Ostrobothnia (where Helsinki, Tampere, Turku and Oulu are likely the largest attractors respectively), which are the only regions along with Åland experiencing population growth now and in the future
- Uusimaa will experience the largest population increase whereas Kymenlaakso and South Savo will experience the largest rate of population decline
- Imports growth is the metric with the highest variance over time
- Kainuu will have the largest export growth percentage until 2026, where it will be overtaken by Åland
- In a market crash, Kainuu would take the biggest hit along with Kanta-Häme and Pirkanmaa
- In a pandemic, Åland's GDP per capita suffers the most, whereas Kanta-Häme, Satakunta and Pirkanmaa are barely affected at all

6.2. Comparison to Ministry Report

To evaluate the trustworthiness of our estimates, we compare them to public growth estimates produced in the spring 2024 by the Ministry of Economic Affairs and Employment of Finland [17]. According to the ministry, Finland's economy is in an economic downturn.

Finland has faced challenges such as energy crises and high inflation that have negatively affected the consumption of households. The Ministry's report collects qualitative data in the form of surveys of the outlook of regions from key regional developers. The interviewed gave an estimate of the economic outlook of the upcoming 12 months on the scale from "significantly better" to "significantly worse" compared to the previous 12 months. In most parts of Finland the interviewed experts estimated the outlook to be better than before or at least neutral. Based on the report, we would be at the bottom of the economic downturn in 2024 and growth would accelerate during the next 12 months.

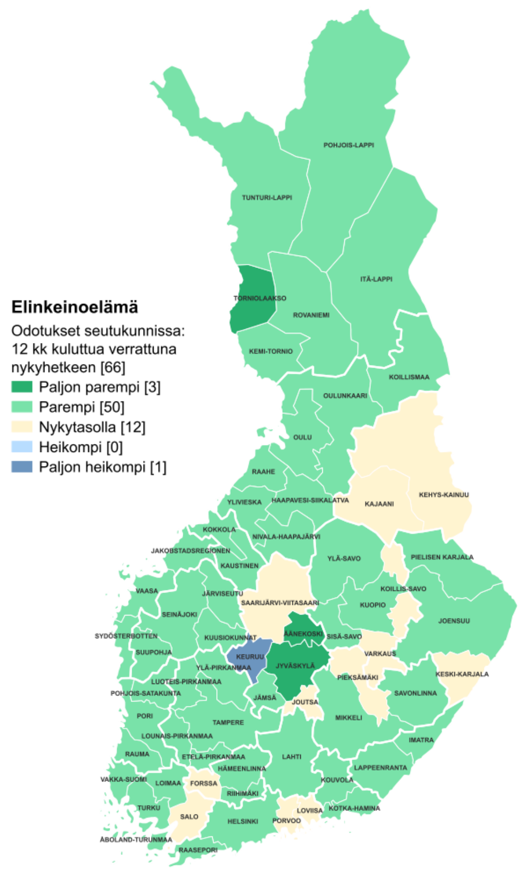


Figure 6.1: Ministry report on economic outlook

When comparing to the forecasts of our data driven approaches, the professionals of the report estimate similar trends in the development of GDP. Here it is good to note that our data is limited to year 2021, whereas the report is collected in the spring 2024 and our estimate of growth is relying more on long term growth trajectory than year specific variation.

The report does not discuss development of population which rarely varies significantly on the short term (except in case of wars or pandemics). When comparing the foreign trade figures we notice a difference to our estimates: our Bayesian model expects the trade to improve in 2024 and the growth to be flat after that, but the ministry report expects grwoth in 2025. This is most likely due to recession continuing longer than expected which is already included in the report published in the spring 2024 but not in our 2023 data.

7. Ethical Issues

Typical ethical issues with data science projects is the collection and handling of data. If individual level data is utilized, it is important that there is the consent of subject and that the individuals are not identifiable from the data. With the chosen topic and approach, there are no ethical related to the ownership of data as all data is collected from public sources and is aggregated over municipalities or regions.

Other possible ethical issues could include exploitative or manipulative methods of processing data to impact the outcome. OP Group as an external collaborator could in principle have other commercial goals, but given that as OP operates all around Finland and data can be used to evaluate the risk level and future outlook of different regions, it is difficult to argue that there would be motives for produced manipulated results. To avoid risk of any misuse, we have aimed for high-level of transparency by listing all data sources and thoroughly discussing data processsing and modelling steps taken.

8. Self-reflection

When it comes to the learning outcomes of the project, especially group work, communication and leading not only others but yourself were considered the most important aspects. We had to learn to set our own deadlines and goals, make prioritization due to time constraints, and delegate tasks within the group. These skills improved significantly during the project: in the early stages of the project a lot of times was used to explore the data and research options which was valuable, but at some point we had to make a decision on how to proceed. The plan can always be adjusted when new information is obtained, but having at least some kind of a timeline makes it easier to keep up with deadlines and finish on time. An example challenge met during the project was the lack of data to make reliable estimates and working time series models.

In addition to learning project management skills, the group felt that we also improved our technical skills when it comes to handling data, running machine learning models, and producing visual output in an informative format. The possibility to be able to utilize methods learned in other courses was one of the most motivating parts of the project.

9. Conclusions

This project was implemented together with OP Group with the goal of making reliable forecasts of the future development of foreign trade and other key economic factors. Economic performance and outlook are some of the most followed topics as they have significant effect on the everyday life of citizens and companies. Economic growth creates jobs and manufacture products and services, and consistently improve the quality of life. Therefore, a lot of effort is devoted to making unbiased and reliable estimates of the future growth.

Public institutions such as Finnish Customs and Statistics Finland were used as the data sources. The quality of data was good but its limited amount most likely negatively affected the reliability of the results produced. After a detailed analysis of the available data, we decided to focus on making forecasts relying on previous trends of the studied variable. We found varying trends between regions confirming for example trends such as urbanization leading to more positive population growth in regions with large cities and a positive trend in total GDP across regions (all regions grow). As the goal was to make these differences easily comparable, we constructed an intuitive interface.

The user interface is an interactive web app where the user can compare the future outlook of different regions, study the regional differences, and estimate the effect of exogenous shocks such as financial crisis on the economic performance. The user interface offers estimates of three key economic factors: foreign trade, population, and GDP. Changing colors of the regions help visually illustrate the regional differences and in addition to the forecasts of a few chosen variables, the user can compare the current differences between the regions through other features as well.

10. Group Construction

The project group 3 consisted of 7 members with the following division of workload: Bach Pham, Minh Ha Le, Quan Tran, and Otto Vintola focused on producing the machine learning models utilized in the project. They divided their workload to studying time series, Bayesian, linear, and Prophet models. In addition, Otto Vintola was in charge of data collection and implementation of the backend. Miro Keimiöniemi and Duc Chu were responsible for designing and implementing the interactive user interface. Ohto Mäkinen focused on identifying key economic factors to predict and utilize in the models and was responsible of organizing and finishing the project document.

References

[1] S. Laakso, 2021. [Online]. Available: https://uudenmaanliitto.fi/wp-content/uploads/2021/12/Uudenmaan-elinkeinojen-tilannekuva-ja-tulevaisuuden-nakymat.pdf

[2] M. Ollikainen and M. Pohjola, "Talouskasvu ja kestävä kehitys," 2013.

[3] Statistics Finland. (2024) Statistics finland. Accessed: 2024-12-01. [Online]. Available: https://stat.fi

[4] Finnish Customs (Tulli). (2024) Finnish customs (tulli). Accessed: 2024-12-01. [Online]. Available: https://tulli.fi/etusivu

[5] Finnish Tax Administration (Vero). (2024) Finnish tax administration (vero). Accessed: 2024-12-01. [Online]. Available: https://www.vero.fi/

[6] Finnish Patent and Registration Office (PRH). (2024) Finnish trade register (kaupparekisteri). Accessed: 2024-12-01. [Online]. Available: https://www.prh.fi/fi/kaupparekisteri.html

[7] F. Pezoa, J. L. Reutter, F. Suarez, M. Ugarte, and D. Vrgoč, "Foundations of json schema," in *Proceedings of the 25th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 2016, pp. 263–273.

[8] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg *et al.*, "Scikit-learn: Machine learning in python," *Journal of machine learning research*, vol. 12, no. Oct, pp. 2825–2830, 2011.

[9] G. Tunnicliffe Wilson, "Time series analysis: Forecasting and control,5th edition, by george e. p. box, gwilym m. jenkins, gregory c. reinsel and greta m. ljung, 2015. published by john wiley and sons inc., hoboken, new jersey, pp. 712. isbn: 978-1-118-67502-1," *Journal of Time Series Analysis*, vol. 37, pp. n/a–n/a, 03 2016.

[10] D. Mcmillan, *Predicting Stock Returns: Implications for Asset Pricing*, 01 2018.

[11] K. H. Brodersen, F. Gallusser, J. Koehler, N. Remy, and S. L. Scott, "Inferring causal impact using bayesian structural time-series models," *Annals of Applied Statistics*, vol. 9, pp. 247–274, 2015.

[12] P.-C. Bürkner, "brms: An r package for bayesian multilevel models using stan," *Journal of statistical software*, vol. 80, pp. 1–28, 2017.

[13] S. J. Taylor and B. Letham, "Forecasting at scale," *PeerJ Preprints*, vol. 5, p. e3190v2, 2017. [Online]. Available: https://doi.org/10.7287/peerj.preprints.3190v2

[14] Y. Dodge, *Autocorrelation.* New York, NY: Springer New York, 2008, pp. 24–26. [Online]. Available: https://doi.org/10.1007/978-0-387-32833-1_17

[15] J. H. Ward, "Hierarchical grouping to optimize an objective function," *Journal of the American Statistical Association*, vol. 58, no. 301, pp. 236–244, 1963.

[16] "Flutter - build for any screen," 2024. [Online]. Available: https://flutter.dev

[17] S. Leveälahti, J. Nieminen, and S. Tolonen, "Alueelliset kehitysnäkymät syksyllä 2024 - valto," 2024. [Online]. Available: https://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/165885/TEM_2024_44.pdf