**Text

Description automatically generatedA picture containing schematic

Description automatically generated**

**Nowcasting Macroeconomic Indicators using Google Trends**

**Executive summary**

The information on economic indicators is crucial for policymaking and taking decision at right time but this information is usually available with a lag. So, the need for alternative data sources is of growing importance for both supplementing Statistics Canada’s data holdings and for nowcasting economic activity. The goal of this project is to develop a methodology to predict macroeconomic factors such as Gross Domestic Product (GDP), retail trade sales and retail e-commerce sales in real time by using the real time data source, Google Trends. Google Trends provides daily, weekly, and monthly reports on the volume of Google queries related to different industries which can help to understand the business cycles and provide signals about multiple aspects of the economy that can further be used to estimate the economic factors in real time. The nowcasting of economic indicators will provide more timely information for policymaking.

1. **Introduction**

Macroeconomic factors are the key drivers of economy and timely information on macroeconomic factors helps in good policymaking. However, this information is available with a lag, for instance, the data for the present month’s GDP is generally published in the coming month/quarter which causes delay in decision-making. To overcome this issue of delayed information gave rise to nowcasting approach and this approach is fetching the interest of economists and researchers nowadays as this modern approach provides the information on economic indicators in real-time. Traditional macroeconomic indicators have some lag, and to fill this gap of information, Google Trends have been widely used as it may help in predicting the present [1]. The volume of queries on different industries may be correlated with the current level of economic activities in respective industry and may help to predict the subsequent data release [1].

Many researchers have used Google trends for nowcasting the economic activity. Google Trends provide information of business cycles and economic activities in economy and the salient features of these business cycles can be captured with few unknown factors using dynamic factor analysis models [2]. These models are applicable to high-dimensional data and can reduce the dimensionality of economic systems. DFM became the mainstream tool for nowcasting GDP growth over the time. Later on, new techniques emerged, and researchers have started to use machine learning algorithms for nowcasting economic factors. Woloszko [3] proposed a weekly tracker to estimate GDP in 46 Organisation for Economic Co-operation and Development (OECD) countries and G20 countries (excluding European Union). The proposed OECD tracker is based on a machine learning algorithm that estimates the relationship between Google Trends variables and GDP growth.

Dauphin et al. [4] have also used Google Trends data to estimate GDP growth, they provide comparative analysis of different nowcasting approaches such as Auto-Regressive models, Dynamic Factor Model (DFM) and some machine learning algorithms like Regularized Regression models, Random Forest, Support Vector Machine (SVM) and Neural Networks, and state that there is no one-size fits all model as different models are suitable for different datasets. Richardson et al. [5] used machine learning algorithms to nowcast GDP growth in New Zealand and their results show that machine learning algorithms boosted trees, SVM and neural networks outperformed the traditional autoregressive models for their study. The aforementioned studies indicate that traditional econometrics models and machine learning models both can be used for the nowcasting economic factors and success, and accuracy of the model may vary for different datasets. Therefore, a comparative study between traditional and modern machine learning algorithms may be more appropriate to fit a model on data in hand.

**References**

[1] H. Choi, H. Varian, Predicting the present with Google Trends, *Economic record*, *88 (2012)*, 2-9.

[2] Stock, J.H. and Watson, M.W., 2016. Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics. In *Handbook of macroeconomics* (Vol. 2, pp. 415-525). Elsevier.

[3] Woloszko, N. (2020). Tracking activity in real time with Google Trends, OECD Economics Department Working Papers, No. 1634, OECD Publishing, Paris.

[4] Dauphin, M.J.F., Dybczak, M.K., Maneely, M., Sanjani, M.T., Suphaphiphat, M.N., Wang, Y. and Zhang, H., 2022. *Nowcasting GDP-A Scalable Approach Using DFM, Machine Learning and Novel Data, Applied to European Economies*. International Monetary Fund.

[5] Richardson, A., van Florenstein Mulder, T. and Vehbi, T., 2021. Nowcasting GDP using machine-learning algorithms: A real-time assessment. *International Journal of Forecasting*, *37*(2), pp.941-948.

1. **AIMS AND OBJECTIVES**

The aim of the project is to develop a methodology to predict macroeconomic indicators such as GDP and monthly retail sales with real-time data source, **Google Trends**. Keywords from the google would be identified which will serve as the predictors for nowcasting the GDP and monthly retails sales.

**Key Objectives**

The key objectives are as mentioned below:

* Nowcasting the macroeconomic indicators like GDP and monthly retail sales using real time data.
* Prediction of GDP will be at national level and will be industry wise.
* GDP nowcasting will be quarterly.
* Nowcasting for retail sales and digital sales will be monthly at national level and as per the industries.

Keywords will be searched from Google Trends

website: <https://trends.google.com/trends/?geo=CA>, and will serve as the predictors for nowcasting the macroeconomic indicators.

1. **DATA SET**

Data set for this project are open ended and the short description about them is as mentioned below:

1. **Gross Domestic Product (GDP) at basic prices monthly:** This dataset is a comma separated file containing the information about the monthly GDP. This file contains data from 1997 and do have some missing values, thus will require data wrangling.
2. **Gross Domestic Product (GDP) at basic prices quarterly:** This dataset is comma separated file containing the information about the GDP quarterly. This file contains data from 1997 and do have some missing values, thus will require data wrangling.
3. **Retail trade sales by province and territory:** This dataset contains information about the retail sales as per the province and territory. This data file is also comma separated and will require data wrangling.
4. **Retail trade sales by industry:** This is a comma separated data set containing the information about the retail sales trades as per the industry. Data wrangling is required in this dataset as well.
5. **Retail sales, price, and volume:** This is a comma separated data set containing monthly retail sales, price, and volume data. This data set will need some data wrangling.
6. **Retail E-commerce sales:** This is a comma separated dataset containing the information about the retail e-commerce sales. This is the data for digital sales and will require some data wrangling as well.

Our focus will be on the data starting from 2004 as we have Google trends available from that period and this will provide us huge data for our nowcasting.

In addition to this, we will be accessing Google Trends website to get real time data for the macroeconomic indicators predictions.

1. **METHODOLOGY**

The detailed description on the methodology of the whole project is as mentioned below.

1. **DATA WRANGLING**

The dataset that we have mentioned above for the GDP contains the information about Canada and is a complete data starting from the year 1997. Also, it has some missing values. We will be cleaning the data in such a way that we have information starting from 2004 and will replace the missing values with the desired values.

Also, the dataset for retail sales and digital sales will be filtered so that it contains information starting from the year 2004.

1. **MODEL FITTING**

* The data set is a time series data and will require the nowcasting using the real time data from google trends. We will be searching for keywords from the Google Trends and will be using them as the predictors in predicting the macroeconomic indicators.
* Some of the keywords are Economic crisis, loans, GPS, unemployment, affordable housing, economy news, agriculture, and forestry.
* The information is divided in **category and subcategory**, and we will be accessing each identified keyword in nowcasting our indicators.
* Since the data is time series, we will be fitting different models for an instance**Dynamic Factor Model(DFM)** which will serve as a dimension reduction method and can be helpful in analysing the time series data. Also, we are aiming at using **Auto-regressive Integrated moving Average (ARIMA)**models to understand the data and predict the indicators. This will serve as a good model for deeper understanding and accurate predictions. Apart from this, we will be implementing **LASSO** in coordination with **ARIMA family models**.
* We will be implementing the above-mentioned models and will be performing the comparative analysis. This will provide us the accurate predictions and will let us choose the best model for our nowcasting.

**IF TIME PERMITS**

* If time permits, we will also implement neural network and try to predict the macroeconomic indicators which could provide high efficiency.
* Forecasting of GDP as per different province and at all industry levels.
* Predicting GDP on monthly basis.

1. **VISUALISATIONS**

* Once the model fitting is achieved, predictions of GDP (quarterly basis), monthly retail revenue and monthly digital sales will be presented in a tabular format.
* This will be followed by bar graphs and line charts for the

predictions.

* A dashboard will be created which will give the information about the nowcasting of the indicators. (If time permits)

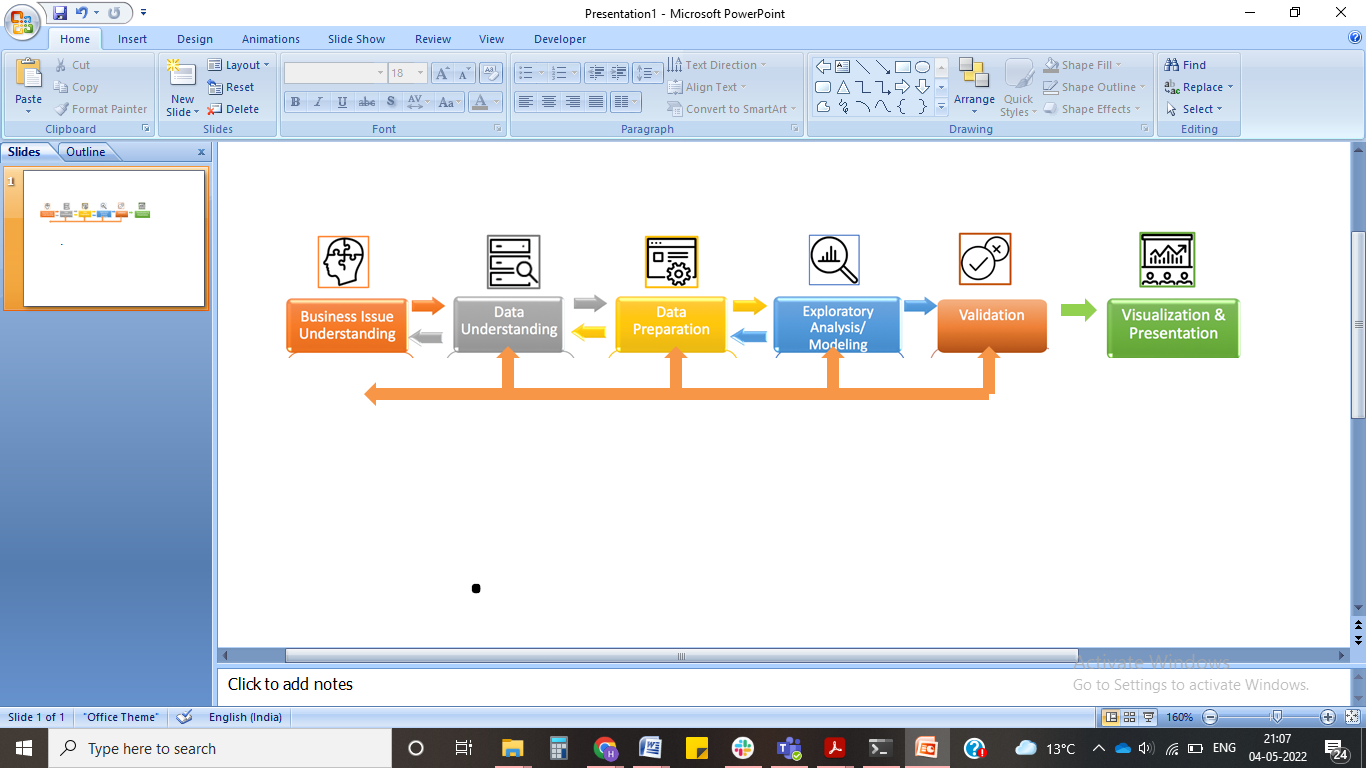
1. **FINAL PRESENTATIONS**

* A final presentation deck will be created which will highlight the research results and predictions.
* This will be structured and concise
* Also, a report will be created that documents the methodology and findings of the research results.

We will be aiming at using Python for all our coding, model fitting and analysis. Also, Git Hub will be the channel where we will be committing our code.

1. **DELIVERABLES AND SCHEDULE/ TIMELINE**

For our project we would be having following deliverables. These outlined deliverables have been elaborated below with the pre-assumed timeline.



**Timeline of Project**

We will start the project as soon as the first meeting kicks off. The total time taken for the final submission of the project would be in **7 weeks**, the final presentation would be delivered on **June 22, 2022**.

|  | **Phase** | | |  | | **Days** |
| --- | --- | --- | --- | --- | --- | --- |
| **Phase 0: Analysis** | | | | | | **2.00** |
|  | **Research & Planning** | | | | | 2.00 |
|  |  |  | **Kickoff** | | |  |
|  |  |  |  | | *Discussion with project partners to define scope of work, expectation settings, high level discussion on the project components.* |  |
|  |  |  |  | |  |  |
|  |  |  | **Literature Review** | | |  |
|  |  |  |  | | *Survey and discussion of literature in the given problem of study* |  |
| **Phase 1: Data Understanding** | | | | | | **8.00** |
|  | **Data Filtering** | | | | | 1.00 |
|  |  |  |  | |  |  |
|  | **Data Cleaning** | | | | | 7.00 |
|  |  |  |  | | *We will clean the dataset and extract the useful factors from the link provided.*  *The missing values would be handled using interpolation of the data.*  *Standarization and Transformation might be applied as required.*  Follow-up meeting for review. |  |
| **Phase 2: Exploratory Analysis and Modeling** | |  |  | |  | **14.00** |
|  | **Exploratory Analysis** |  |  | |  | 2.00 |
|  |  | Digging deep into the data structure  Understanding the statistics measures  Understanding the data distribution  Follow-up meeting for review | | | |  |
|  | **Modeling and Validation** | | | | | 12.00 |
|  |  | Exploring Time series with machine learning models  Fit multiple models  Testing and validating the models and selecting the best one  Follow-up meeting for review | | |  |  |
| **Phase 3: Visualization/Dashboard** | | | | | | **13.00** |
|  | **Visualization** | | |  | | **3.00** |
|  |  | Plot type selection  Follow-up meeting or chat for review | | | |  |
|  | **Dashboard** | | |  | | **10.00** |
|  |  | Layout Designing and Styling  Follow-up meeting for review  Structure the design  Follow-up meeting for review  Visualization Compliances | | | |  |
|  |  | Functional Testing  Follow-up meeting for review | | | |  |
|  |  | Visualization Deployment | | | |  |
| **Phase 4: Presentation** | | | | | | **6.00** |
|  | **Documentation and Presentation** | | | | | **6.00** |
|  |  | Written Documentation | | | |  |
|  |  | Prepare Presentation  Follow-up meeting for review  Final Changes  Final follow up and submission | | | |  |
|  |  |  |  | | ***Project Effort*** | **0.00** |

