Research practice I

Final report

A nowcasting model for Medellín city

María Camila Vásquez Correa[†], Juan Carlos Duque [‡] and Jairo Alejandro Gómez[‡]
mvasqu49@eafit.edu.co, jduquec1@eafit.edu.co, jagomeze@eafit.edu.co

[†]Mathematical Engineering, Universidad EAFIT
[‡] Research group in Spatial Economics (RiSE), Universidad EAFIT

March 21, 2019

Abstract

In this article we present the first steps of a nowcasting methodology for the Medellín unemployment rate by using time-series regression to identify the generating process of the series and object detection to identify several categories of automobiles in images, in order to follow a first step to prove the hypothesis of how well the traffic of the city can reflect the employment, or in general, the economic activity inside it. Two models were evaluated: a SARIMA model to forecast the unemployment rate using historical data, and a Single-Shot Detector trained to detect 8 classes of vehicles in CCTV traffic cameras installed in Medellín.

Keywords: Nowcasting, Unemployment, Object detection, Time series.

1 Introduction

Macroeconomic variables such as unemployment rates and GPD are key for the conduct of monetary policies, macro-prudential policies and fiscal policies. The real gross domestic product (GPD) is a summary of the health of an economy (Bragoli & Modugno, 2017) and is one of the most heavily monitored indicators. Also, unemployment rate, released on a monthly basis (DANE, 2018), helps the government tracking the state of the economy and, in the Colombian case, is the input for new policies and the redesign of existent ones with the aim of reducing unemployment and improving the quality of life of Colombians.

The period used to update unemployment rate and GDP does not allow the policy makers to take actions based on fresh data: they have to wait a long time until the indicator is released to make decisions with it. Our world is changing faster each day and this requires decisions with the same velocity. This implies that the conventional methodologies, such as surveys and the traditional data collection that take two months to generate an indicator as important as the unemplotment rate or the GDP, begin to be inappropriate to support the decision making.

To overcome this problem, some authors have proposed nowcasting models to estimate economic indicators with a higher frequency, allowing their respective economies to react faster. Nowcasting is a contraction of the words *now* and *forecasting* and it is defined as "the prediction of the present, the very near future and the very recent past" (Elliott & Timmermann, 2013). These models have been applied to estimate the GDP in most of the literature.

In this work, we aim to investigate a new way of producing an indicator for the unemployment rate in Medellín city, by exploring the potential of recent developments in the field of artificial intelligence, especially in convolutional neural networks, for the design and implementation of a nowcasting model that allows the appropriate authorities to measure in real time (or at least with a much higher frequency that is currently available) the unemployment rate, a key economic variable.

The underlying hypothesis is that the unemployment rate and the traffic flow on the streets are negatively correlated. In particular, the circulation of cars, buses, motorcycles, taxis and others is can be an indicator of the unemployment rate. Using the object vehicle counts, along with historical data provided by key actors from the city, we aim to construct a methodology that provides a timely forecasting for the Medellín's unemployment rate.

2 State of the art

The nowcasting problem is not new. It has been treated by several authors in several countries in order to generate a real-time indicator for their decision makers, using diverse types of data and models to perform the forecast. Some of these studies are discussed in the following paragraphs.

In Canada, Bragoli & Modugno (2017) propose a model for the nowcasting of the GDP considering some variables from the United States. The problem is clear: the value of the gross domestic product of the current quarter is generated with a delay of two months, so the country has a bigger delay compared to other developed countries such as Japan or the United Kingdom. Their target variable is the quarterly GDP and their input series are the purchase management index, employment, manufacturing shipments, retail sales, exports and imports, among others. Many of these variables are obtained from surveys and trade indicators.

In Urasawa (2014), authors propose another model for the nowcasting of the real GDP of Japan, selecting monthly indicators such as industrial production, employment and retail sales. They follow the lead of some other authors that have models for the short-term forecast of the real GDP in various economies. For instance, they mention Lahiri & Monokroussos (2013), where authors study the effects of survey data in nowcasting the U.S GDP, and Karim *et al.* (2010), who introduce a model for nowcasting the French GDP.

It has been found that, in terms of nowcasting for macroeconomic variables, the models proposed only include GDP. Some other examples include models for the Turkish economy, which is proposed by Modugno *et al.* (2016), and considers financial data and survey data for the nowcasting; The Czech model, detailed in Rusnák (2016) and the euro-zone model detailed in Maximo & Gabriel (2008).

For the models discussed above, the data used for the nowcasting has economic nature (surveys and financial indicators), and in most models, the measured variable is GDP. There is a great opportunity for this research project to explore ways of including new types of data (processed images from the city) and to estimate new variables (unemployment rate, for instance).

The final input for our nowcasting model will be a set of vehicle countings. However, in order to get these countings, we need to identify in a prior stage the vehicles that appear in a set of images or in a video sequence, a process known in computer vision as an object detection. A few works that approach this problem in the literature are discussed in the following paragraphs.

In (Sermanet *et al.* , 2013), a framework for using Convolutional Neural Networks (CNNs) for classification, localization and detection is presented, the presented approach is useful in order to identify multiple objects in an image. Following that work, (Girshick *et al.* , 2013) presented the Regional Convolutional Neural Network (R-CNN) framework, that aims to apply high-capacity CNNs to bottom-up region proposals in order to localize and segment objects, and then classify them into their own categories, outperforming the previous framework.

Since its introduction, the R-CNN framework has been refined. Some improvements include the Fast R-CNN, presented in (Girshick, 2015), which in comparison with Spatial Pyramid Pooling network (He *et al.* , 2014), SPPnet, it has a high performance in detection of objects, being more accurate and faster. The Fast R-CNN outperforms the R-CNN by combining the SPPnet to speed up the test time.

However, the previous approaches required a fair amount of time when computing the region proposals. Aiming to tackle this problem, (Ren *et al.* , 2015) presented an algorithmic change, based on deep CNNs to find the regions, in order to drastically reduce the cost of the previous R-CNN approaches. This approach, which is called a Region Proposal Network (RPN), takes advantage on the high convolutional layers present in the detector as input and outputs a set of rectangular object proposals, each with an objectness score. This model is called the Faster R-CNN, and it was extended by the Mask R-CNN (He *et al.* , 2017).

Using a completely different approach, (Liu *et al.* , 2015) introduced the Single-Shot detector (SSD), which outperforms the previous releases with a simpler model, easy to train and straightforward to integrate with other systems. It is one of the favorite approaches for real time processing of images, due to its test time and high accuracy.

One of the most interesting applications for object detection is the detection and tracking of objects in video. Some approaches to perform this task are the You Only Look Once (YOLO) architecture (Shafiee *et al.* , 2017), available in three versions, as well as a SqueezeNet architecture (Wu *et al.* , 2016), that perform really well in real-time detection of objects.

3 Methodology

3.1 Unemployment rate characterization

The macroeconomic variable chosen to perform the nowcasting is the unemployment rate in Medellín city. In order to build the model, we first need to characterize it, observe its behavior and fin possible generating processes for the time series obtained in (DANE, 2018).

The unemployment rate is an indicator of the degree of use of human resources in the economy (DANE, 2016), it is also an indicator of the life quality of the population and its performance in society. For this reason, the unemployment rate is an important indicator to take into account in the design of public policies for Colombian cities. This rate is released by the DANE in a monthly basis, based on a mobile quarter measurement, with a two-month delay.

Going trough the methodology of the National Administrative Department of Statistics, DANE, we found the following information about the calculation of the unemployment rate in the country for the series reported for the years 2000 trough 2018 (DANE, 2016). This department performs a continuous survey in the country, called the "Large integrated household survey" (Gran Encuesta Integrada de Hogares), or GEIH, for its Spanish acronym, which provides the unemployment rate among other important information on a monthly basis.

3.1.1 Calculation of the unemployment rate

The unemployment rate, UR, is the ratio between the number of people seeking for a job and the number of people in the labor force:

$$UR = \frac{DS}{EAP} * 100 \tag{1}$$

Where EAP stands for Economically Active Population and DS is the number of people seeking for a job. This last portion of the population is extracted from the GEIH. It is composed by two additional variables:

- *Open unemployment rate (OUR)*:

  $$OUR = \frac{DSA}{EAP} * 100$$

  DSA stands for the number of people that were:

  - Unemployed in the reference week.
  - Did errands in the last month.
  - Available.

- *Hidden unemployment rate (HUR)*

  $$HUR = \frac{DSO}{EAP} * 100$$

  DSO stands for the people that were:

  - Unemployed in the reference week.
  - Did not do errands in the last month, but in the last 12 months.
  - Available.

For Medellín, which is one of Colombia's capital cities, the unemployment rate is calculated every month for the city and its Metropolitan Area. The collection period is weekly, and is designed in a way that allows the department to collect the necessary data for a good forecast of the unemployment rate, while reducing time and cost. The criteria used for choosing the households to be surveyed guarantees, based on statistics, a relative standard error of less than 5%, and also the sample design allows segments to not be repeated in the capital in three years, ensuring sufficient variability in the sample to achieve reliable results (DANE, 2016).

This survey was updated in 2000. Before that year, the unemployment rate was calculated using another survey, and it provided the data quarterly, not in a moving quarter. The left graph of Figure 1 shows the historical data for the unemployment rate from 1994 to 2000, in orange, and in blue the one from 2000 to August 2018. Due to the sample period difference, we are taking the data from 2000 for analysis purposes.

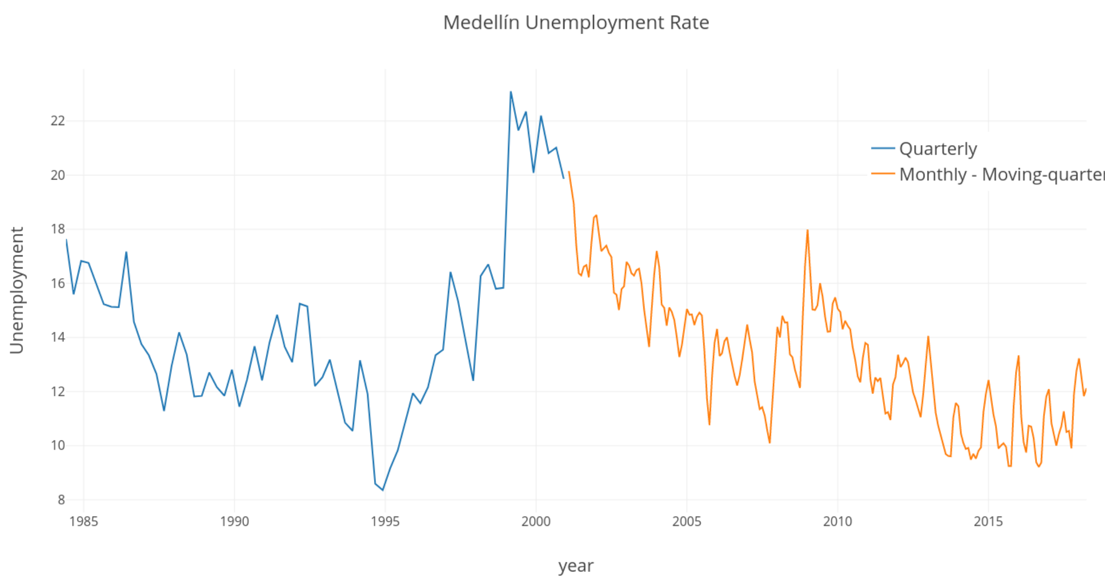


Figure 1: Time series of the unemployment rate of Medellín and its Metropolitan Area.

3.1.2 Time series analysis

In order to characterize the unemployment rate, the first step was to perform a complete time-series analysis aiming to understand its behavior and quantify the dependence of the series from past observations. This approach will lead us to to design a basic prediction model that could serve as a baseline to measure our progress in the nowcasting process, once we add the vehicle counts. Also, it could give us an idea of the behaviors that the counts series should capture, or how the counts should be taken to fully reflect the characteristics of unemployment of the city.

Following the proposed methodology in (Wei, 2006), we are searching certain patterns in the series that allow us to identify whether or not they are stationary, if they are auto correlated, have moving average, present seasonal pattern and if they have an stable long-term variance or they present heteroscedasticity. After identifying these characteristics, we will be able to find a possible generating process of the series, and propose a first forecasting model for it.

3.2 Vehicle counting

For the vehicle counting in the images we chose a Single Shot Detector (SSD) over a Faster Regional Convolutional Neural Network (RNN) for the object detection step, due to the fact that it has a better performance in terms of test time, and it has a good accuracy. Also, there are some SSD architectures already trained to identify characteristics from vehicles, and detect categories such as cars, buses and motorcycles in images that could serve as a first prototype approach to perform the counts of the vehicles ciculating Medellín.

SSD are designed for object detection in real-time. Faster R-CNN uses a region proposal network to create boundary boxes and utilizes those boxes to classify objects. While it is considered the start-of-the-art in accuracy, the whole process runs at 7 frames per second. Far below what a real-time processing needs. SSD speeds up the process by eliminating the need of the region proposal network. To recover the drop in accuracy, SSD applies a few improvements including multi-scale features and default boxes. These improvements allow SSD to match the Faster R-CNN's accuracy using lower resolution images, which further pushes the speed higher. According to the following comparison, it achieves the real-time processing speed and even beats the accuracy of the Faster R-CNN (Liu *et al.* , 2015).

In order to make use of the available Single Shot Detectors, a Transfer Learning (Pan & Yang, 2010) approach is proposed. It makes use of previously learned features by the network to help it identify new categories. To perform this step, it is necessary to consider the new categories of interest, build a dataset that contains a representative number of examples per class, and follow the transfer learning process, that takes the training of the network in a certain step, called a checkpoint, and introduces the new training and test data.

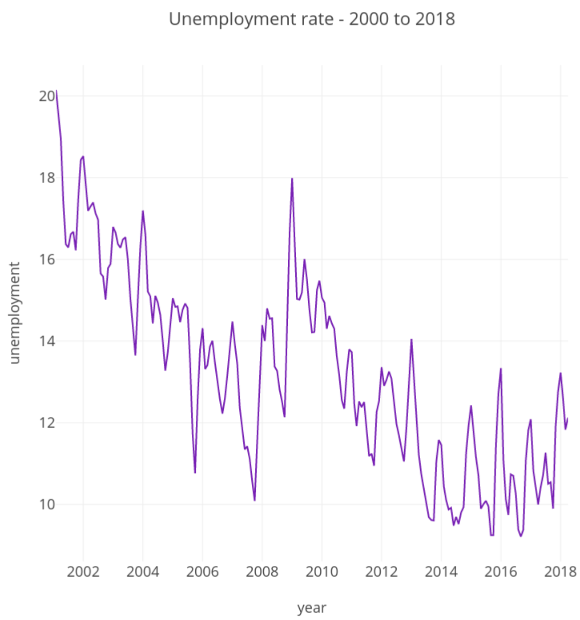
4 Results

4.1 Time series model

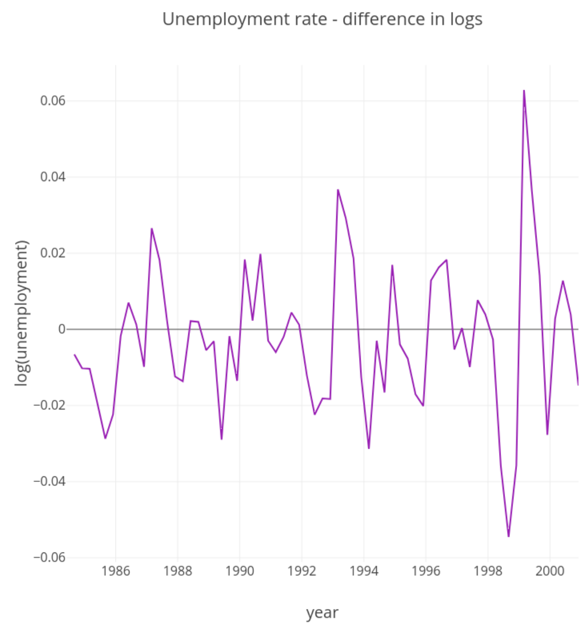
For this analysis we take the most recent data, the monthly unemployment rate, measured since 2000 to 2018. The first step in identifying the generating process of the series, is to check its stationarity. In Figure 2 is shown the graphs for the levels series (2a) and its respective difference in logs (2b), according to the Box-Cox transformation (Box & Cox, 1964).

Looking at the levels graph, and evidencing its clear trend, the first approach is to difference the series and observe if the differentiated one has also a trend. In this case, the differentiated series may be stationary. In order to analyze the stationarity, we apply the Dickey-Fuller test for unit roots, see (Dickey & Fuller, 1979), to check whether the series are stationary or not. We performed two versions of this test:

- Applying the **augmented Dickey Fuller test** for unit roots, and observing the critical values, the p-value for the original series is 0.32 and for the differentiated series the p-value is



(a) Unemployment rate series in levels.



(b) Unemployment rate, difference in logs.

Figure 2: Unemployment Rate

0.03, meaning that the differentiated series, with 95% confidence, does not have a unit root and one can proceed to fit a model.

- Applying the **Dickey Fuller GLS test** (Cheung & Lai, 1995), that has a better performance than other tests, with an appropriate number of lags, the p-value for the original series is 0.64 and for the differentiated series the p-value is 0.03, that conserves the results from the previous test.

For additional criteria, the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test (Kwiatkowski *et al.* , 1992) can be applied. This test behaves differently from the Dickey Fuller test, since it has a null hypothesis of stationarity. The p-value for the original series is 0.01, that rejects the null hypothesis, thus, the series are not stationarity. And for the differentiated series, the p-value is 0.43, that does not have evidence to prove that the series contains a unit root, and therefore, is not stationary. Figure 3 illustrates the autocorrelation and partial autocorrelation graphs of the series, which show evidence of auto-regressive component and moving average, along with a clear evidence of hysteresis, illustrated also by the results of the stationarity tests.

Hysteresis is defined as a condition in which a system state does not depend only on the external conditions at the moment considered, but also on the evolution followed until reaching said state (González, 2012). In several studies, it has been proved that the unemployment rate for Medellín city has clear signs of this phenomenon. The hysteresis of the series reflects in an absence of an effect of monetary policy and a regionally differentiated impact of investment and exports on this rate are palpable (González, 2012).

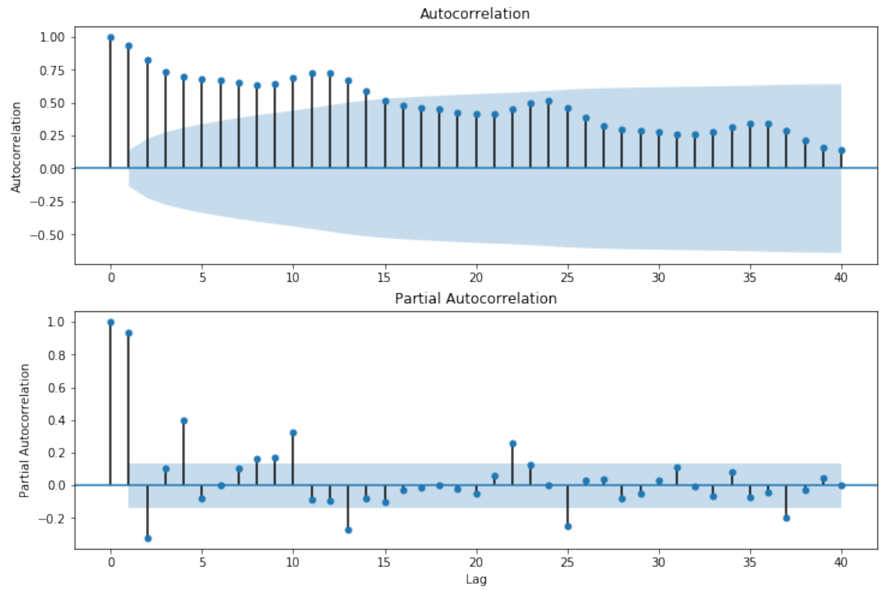


Figure 3: Autocorrealtion and partial autocorrelation graph for the series.

Given the behavior of the city economy across different periods, we can intuitively consider a seasonality hypothesis, in which unemployment rises or falls depending on the needs of the market in a certain period of the year. Figure 4 shows the unemployment rate series for each month, showing strong differences between each other. Following this idea, we could consider a pattern for the raising or decreasing of unemployment based on the current month of the year.

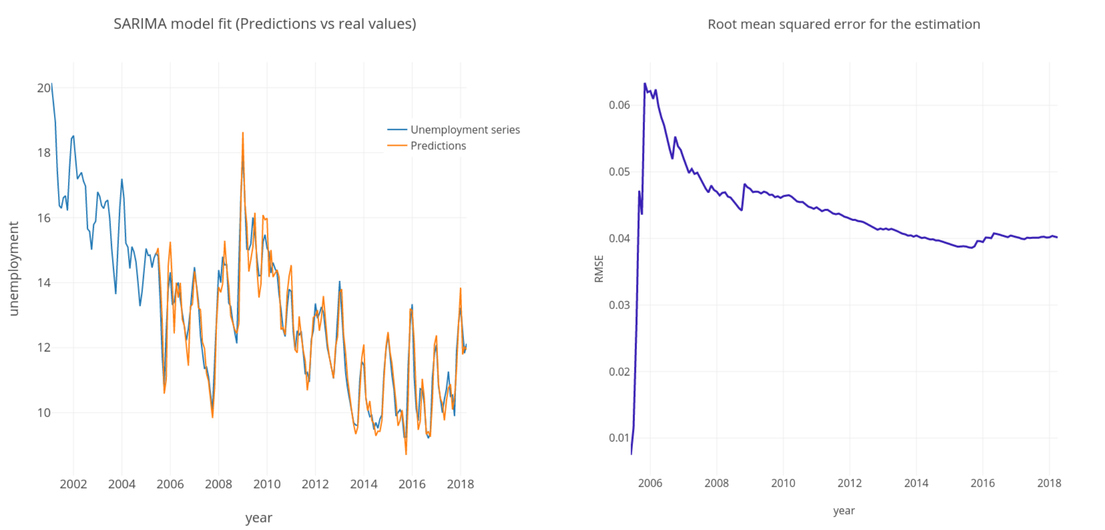
Figure 4: Unemployent rate for each month.

With these resources, (the non-stationarity of order one for the series, the autocorrelation and moving average component and the seasonal pattern), we proceed to fit a seasonal, auto regressive, integrated, moving average SARIMA model for the series Wei (2006). Before trying the combinations of parameters that could suit our model, the estimation results for different values of p, d, q and s are shown in Table 1. The best model for each s was chosen using also the Akaike's information criteria (AIC) (Akaike, 1974).

Table 1: Results for the model estimation.

Model	**AIC**
SARIMA(2,1,2)x(2,0,1,12)	-778.54
SARIMA(1,1,2)x(2,0,2,6)	-761.74
SARIMA(1,1,0)x(1,0,2,3)	-727.11
ARIMA(2,1,2)	-665.27

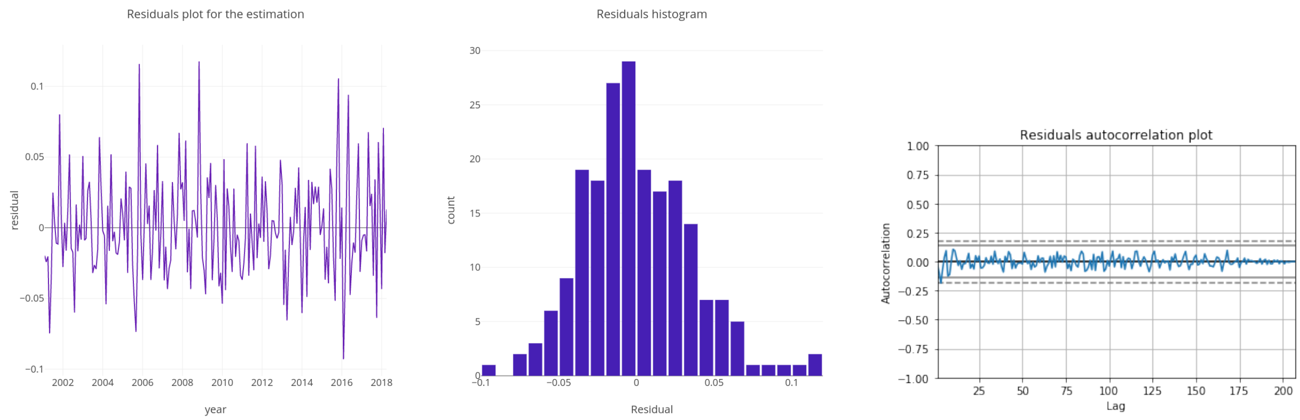
We choose the SARIMA$(2, 1, 2) \times (2, 0, 1, 12)$ model, even though, when making the regression, the second lag for the autoregressive seasonal component is only significant with a 10%. In Figure 5a, are shown the results of fitting the model to a portion of the data an predicting the next values from it. Also, Figure 5b is shows the root mean squared error for these predictions. By seeing the scale, one can notice that the error is really low, and it decreases as the prediction values are added to the training data.



(a) Predictions for the model on a rolling window. (b) Root mean squared error for the predictions.

Figure 5: Predictions of the unemployment rate using a SARIMA$(2, 1, 2) \times (2, 0, 1, 12)$ model.

Figure 6a shows the behavior of the residuals for the estimation. It can be observed that it resembles white noise, which gives us confidence in the model and its predictions. Figure 6b presents the histogram of the residuals of the estimation, and it has the overall shape of a normal distribution. Also, in Figure 6c we include the autocorrelation graph for the residuals, with dotted bands at 95% and 99%, and there is no evidence of autocorrelation. Also, the mean for them is almost zero, and they have a low variance, so we can conclude that the residuals are white noise, and the model should work in further estimation without the need of analysis for heterocedasticity.



(a) Residuals plot. (b) Residulas histogram. (c) Residuals correlation plot.

Figure 6: Residual analysis for the model fitting.

From this analysis, we found some interesting characteristics of the unemployment: it depends on its previous values, as well as in the particular behavior of the cities economy, and the policies designed for its control will not be reflected in the short term. Also, the data collection and processing made by our approach should consider these trends and behaviors, and the seasonality that includes a whole year in order to capture the whole phenomenon of the unemployment rate. This is the first forecasting model proposed, following a classical approach, that we intend to improve as the research continues.

4.2 Vehicle detection and counting

4.2.1 Information to be extracted

For the cameras situated in strategic places of the city, our objective is to identify the number of vehicles of different types circulating the city in various periods of time. For this purpose, we selected some important information to be extracted from the images of the cameras of the city: the type of vehicles we want to detect, count and generate a time-series of countings with. The selection of these categories of interest is:

- Motorcycles of any type, bicycles and three wheeler.
- Car, all terrain vehicle (ATV).
- Van.
- Taxi.
- Camper.
- Truck, tractor, dump truck.
- Bus, minibus.
- Industrial machinery.

Counting these vehicles will allow us to have a time series that provide the information about the automobiles in the city with different sample periods (dayly, weekly and monthly), that will finally be the input for the nowcasting model we want to propose. These categories were defined by exploring the existing and registered automobiles in the cities of Envigado, Copacabana and Sabaneta, which are part of the Metropolitan area of Medellín. The data can be found in (Datos Abiertos, 2018a), (Datos Abiertos, 2017) and (Datos Abiertos, 2018b). The *industrial machinery* category is formed by the following subcategories (appearing in (Datos Abiertos, 2018a)):

- Froklitft.
- Farm equipment.
- Backhoe.
- Tractor.
- Excavator.
- Charger.
- Front loader.
- Semitrailer.

The ambulance category is discarded due to the fact that no economic information is provided by it, and it will be considered within the *van* or *car* category, according to the case. Also, we merged some categories that could be considered small, such as bicycles and three wheelers with motorcycles, for analysis purposes. The Single Shot Detector will be trained to detect all of these categories in images from real cameras of Medellín.

4.2.2 Track segments to be monitored

In order to obtain the data described in the section 4.2.1, we selected some strategic streets from the city, that provide the most information about the current activity. The selection was made based on the recommendations made by the Mobility Secretary of Medellín and one external expert in Mobility. The selected road tracks include:

1. San Juan Avenue: Strategic relevance due to the high flow of buses, and its proximity to a commercial point in the city, which generates high-traffic flow from industrial sectors, wood, leather, metalworking clusters and its proximity to the utility company of Medellín (EPM).
2. North freeway: Relevant for its connection with the Colombian coast and the capital of the country.
3. South freeway: Relevant for its connection with other municipalities in the metropolitan area and the South of the country.
4. Iguaná: It routes part of the vehicular flow that comes from the West.
5. 80th Avenue: Located on the Western side of the city. Captures the flow of vehicles from North to South.
6. 33 Avenue: Located on the Western side of the city. Captures the flow of vehicles from East to West.

These sectors contribute to analyze the economic flow of the city, the inputs and outputs from other regions of the country and the traffic. However, we also included other important corridors to analyze the intra-urban road traffic. These segments reflect the internal vehicular traffic, and they could be useful to capture the commerce of the city itself and people mobility for job-related reasons:

- Las Palmas Avenue.
- Avenue El Poblado.
- 30th Avenue.
- Avenue Guayabal.

4.2.3 Data collection

The data that we will use for the detection and further counting of the vehicles on the road, come from the cameras of the city located on the segments defined in the previous section. This cameras are available for the public in Mobility Secretary of Medellín secretaría de movilidad de Medellín, with a snapshot of each camera per minute. In order to obtain the historical videos from this cameras, we contacted the Mobility Secretary of Medellín, and have several meetings with them, which allowed us to define:

- The segments to be monitored.
- The cameras needed to monitor those segments.
- The procedure to follow in order to obtain the videos.

While this process occurs, we are constantly downloading the available images in the web page, making a request to the server per minute and storing the image of all of the 81 available cameras, indexed by date and time of recording. Figure 7 shows some examples of the data downloaded so far, about 4GB per day. An effort was made to access the videos and not only the snapshots of CCTV cameras owned by the Mobility Secretary. Various meetings were held with staff f, and at the end, the information they could provide us only had the last 7 days for each camera. Unfortunately, such time span not enough for the analysis. Therefore, it was decided to continue downloading images every minute.

Figure 7: Sample images taken from the cameras located in Medellín.

According to the track segments defined, 26 out of a total of 81 cameras were chosen to inspect and analyze the traffic flow of the city. Also, the remaining images are used to train the network and perform accuracy and speed tests. We still need to evaluate in practice if these cameras provide enough spatial diversity to estimate the underlying road traffic across Medellín.

4.2.4 First prototypes

With the data that we are constantly obtaining, we intend to generate a time-series for the countings in each category. Towards this purpose, we implemented an early object detector, using the MobileNet SSD model (Liu *et al.* , 2015), pre-trained using the COCO dataset, and re-trained with 300 hand-labelled images from our dataset. This model is trained to detect the eight defined categories for this project.

Figure 8 presents the evolution of the loss function of the detector during training. The training was stopped when a certain convergence was reached, and its inference graph was exported to make predictions like the ones presented in Figure 9, that shows the detections of the model for a batch of 6 images, each of them a snapshot from one CCTV camera. Although the model is still far from perfect, the first set of series with vehicle counts was generated hourly during an entire day for a single camera, in order to have a first insight of the behavior of the vehicles in an specific sector (for this experiment, the camera was located at El Poblado Avenue).

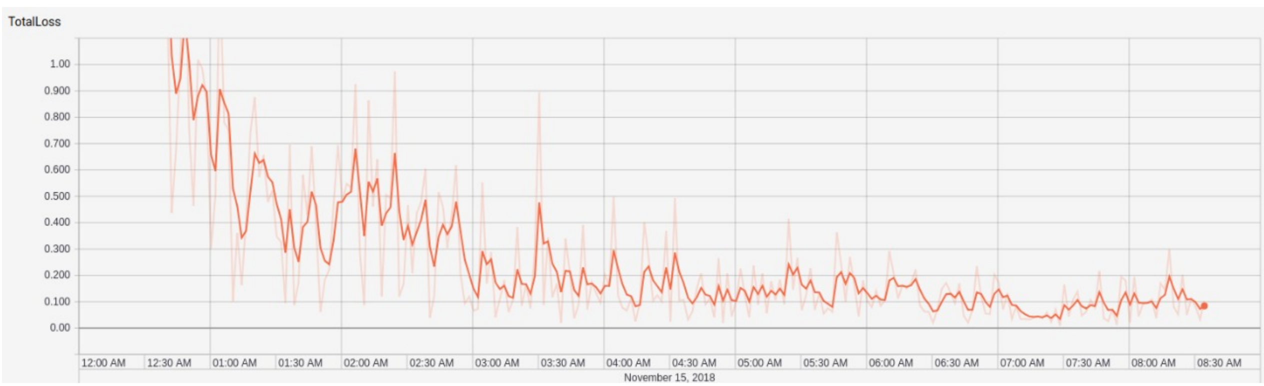


Figure 8: Loss curve of the detector during training.

The results are in agreement with what is expected for this sector of the city, showing a high vehicular activity during the first hours of the morning, especially of cars, just like in the afternoon hours, both times near the peak hours. Using these counts, we want to see if the category diversity helps us capture more information about the vehicle flow and the general activity of the city, or if less categories are enough to perform this task. However, we lack of information on how well the categories were identified, due to the fact that the network needs more examples and a re-training to get a better performance.

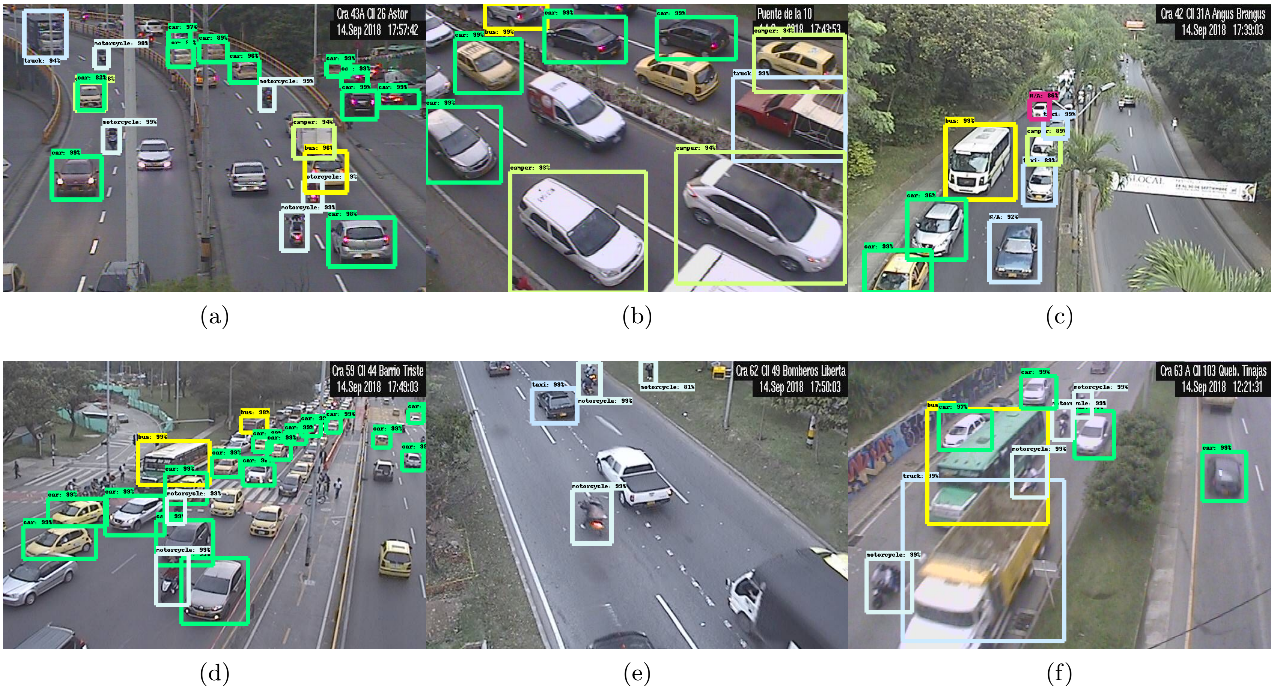


Figure 9: Detection results for the re-trained Single Shot Detector on (a), (b), (c), (d), (e) and (f).

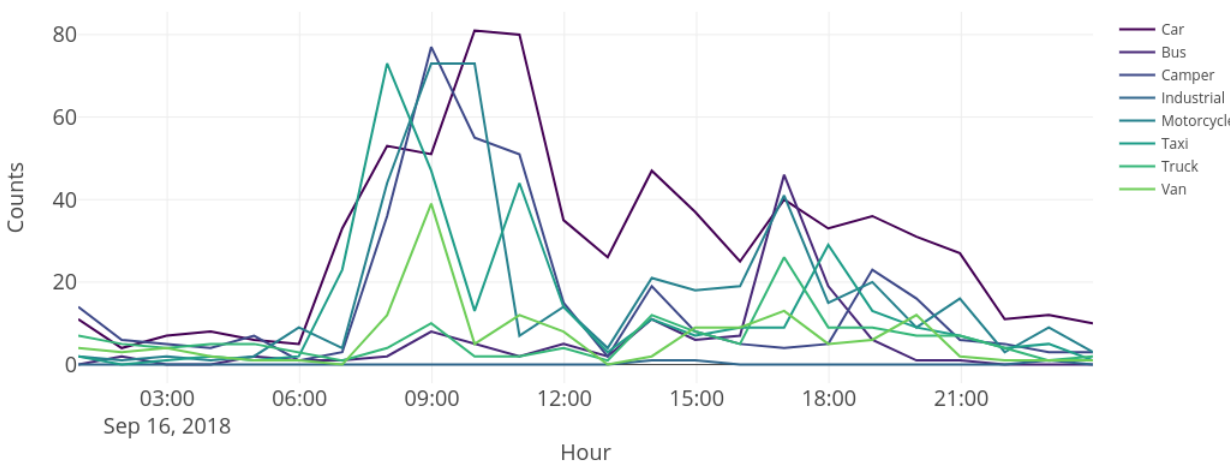


Figure 10: Hourly counts of vehicles per category extracted from a single camera during an entire day using the re-trained Single-Shot Detector.

5 Conclusions and future research

The series for the unemployment are complex and have dynamics that should be taken into account when predicting them with another data source, such as the hysteresis and the seasonality found in the time-series analysis. The generating process found is a good first step for the forecasting, and could be used as an input for the nowcasting by vehicle counts, by measuring the correlation between this series and the traffic in the city, and proposing a new forecasting model, via time-series or machine learning, whose inputs could include the historic unemployment rate and the series for the counts.

A Single Shot Detector was trained to obtain counts from cameras of the city, and, even when there is still a huge opportunity for improvement as more images get labeled, it is a first step to start generating the series of the counts by using the historical images from the public cameras of Medellín. Some future work in the detector is to refine the test step, label more training data, and compare the selected model performance with other ones to select the best one. We would also want to see the performance of other approaches like the Regions Convolutional Neural Networks (R-CNN) compared to the SSD, to evaluate the chosing of the architecture for the project.

For future research, we expect to generate new forecasting models for the unemployment rate that include the series for the counts of the vehicles from the city, one that gets fed up with the historical unemployment, and one that could predict the unemployment using only the vehicle counts, without the historical unemployment rate data. The reason for the incremental steps is first, have an insight of the generating process of the unemployment, its characteristics and the phenomenons that can affect it. Second, before starting to build a forecasting model with vehicle counts, we want to analyze whether or not the series of the vehicle counts and the unemployment rate are correlated and provide useful information. And finally, the goal is to generate the nowcasting model only with the counts, knowing they provide the information of the unemployment and capture its main characteristics.

Acknowledgements

To the researchers at RiSE group, who provided support for the development of this project. Also, to the computational biology group, and Santiago Hincapié-Potes, for their help in this practice.

References

Akaike, Hirotugu. 1974. A new look at the statistical model identification. *IEEE transactions on automatic control*, **19**(6), 716–723.

Box, G, & Cox, David. 1964. An Analysis of Transformations (with Discussion). **26**(01), 211–252.

Bragoli, Daniela, & Modugno, Michele. 2017. A now-casting model for Canada: Do U.S. variables matter? *International Journal of Forecasting*, **33**(4), 786 – 800. `"http://www.sciencedirect.com/science/article/pii/S0169207017300341"`.

Cheung, Yin-Wong, & Lai, Kon S. 1995. PRACTITIONERS CORNER: Lag Order and Critical Values of a Modified Dickey-Fuller Test. *Oxford Bulletin of Economics and Statistics*,

57(3), 411–419. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1468-0084.1995.mp57003008.x.

DANE, Departamento Administrativo Nacional de Estadística. 2016 (april). *Metodología General Gran Encuesta Integrada de Hogares - GEIH*. https://www.dane.gov.co/files/investigaciones/fichas/metodologia_GEIH-01_V9_2.pdf.

DANE, Departamento Administrativo Nacional de Estadística. 2018 (Jun). *Gran encuesta integrada de hogares (GEIH) Mercado laboral*. https://www.dane.gov.co/index.php/estadisticas-por-tema/mercado-laboral/empleo-y-desempleo.

Datos Abiertos, Gobierno digital Colombia. 2017 (Ago). *Distribucion Del Parque Automotor Inscrito2016(II)*. https://www.datos.gov.co/Transporte/Distribucion-Del-Parque-Automotor-Inscrito2016-II-/nd4h-ytqu.

Datos Abiertos, Gobierno digital Colombia. 2018a (Jul). *Parque Automotor Municipio de Envigado hasta 31-07-2018*. https://www.datos.gov.co/Transporte/Parque-Automotor-Municipio-de-Envigado-hasta-31-07/fqwq-ik2w/data.

Datos Abiertos, Gobierno digital Colombia. 2018b (Oct). *Registro histórico del parque automotor de Sabaneta*. https://www.datos.gov.co/Transporte/Registro-hist-rico-del-parque-automotor-de-Sabanet/bewj-2mvn.

Dickey, David A, & Fuller, Wayne A. 1979. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association*, **74**(366a), 427–431.

Elliott, G., & Timmermann, A. 2013. *Handbook of Economic Forecasting*. Handbook of Economic Forecasting. Elsevier Science. https://books.google.com.co/books?id=-_dSuakNHWYC.

Girshick, Ross B. 2015. Fast R-CNN. *CoRR*, **abs/1504.08083**. http://arxiv.org/abs/1504.08083.

Girshick, Ross B., Donahue, Jeff, Darrell, Trevor, & Malik, Jitendra. 2013. Rich feature hierarchies for accurate object detection and semantic segmentation. *CoRR*, **abs/1311.2524**. http://arxiv.org/abs/1311.2524.

González, Angélica Beltrán. 2012. *Histéresis en el desempleo: una revisión de los estudios para Colombia*. Universidad Colegio Mayor Nuestra Señora del Rosario. http://repository.urosario.edu.co/bitstream/handle/10336/3947/52057350-2012.pdf;jsessionid=CF10FC014D8D1CBE40F3918925EEAFAF?sequence=1.

He, Kaiming, Zhang, Xiangyu, Ren, Shaoqing, & Sun, Jian. 2014. Spatial pyramid pooling in deep convolutional networks for visual recognition. *Pages 346–361 of: European conference on computer vision*. Springer.

He, Kaiming, Gkioxari, Georgia, Dollár, Piotr, & Girshick, Ross B. 2017. Mask R-CNN. *CoRR*, **abs/1703.06870**. http://arxiv.org/abs/1703.06870.

Karim, Barhoumi, Olivier, Darné, & Laurent, Ferrara. 2010. Are disaggregate data useful for factor analysis in forecasting French GDP? *Journal of Forecasting*, **29**(1-2), 132–144. https://onlinelibrary.wiley.com/doi/abs/10.1002/for.1162.

Kwiatkowski, Denis, Phillips, Peter C.B., Schmidt, Peter, & Shin, Yongcheol. 1992. Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics*, **54**(1), 159 – 178. http://www.sciencedirect.com/science/article/pii/030440769290104Y.

Lahiri, Kajal, & Monokroussos, George. 2013. Nowcasting US GDP: The role of ISM business surveys. *International Journal of Forecasting*, **29**(4), 644 – 658. http://www.sciencedirect.com/science/article/pii/S0169207012000374.

Liu, Wei, Anguelov, Dragomir, Erhan, Dumitru, Szegedy, Christian, Reed, Scott E., Fu, Cheng-Yang, & Berg, Alexander C. 2015. SSD: Single Shot MultiBox Detector. *CoRR*, **abs/1512.02325**. http://arxiv.org/abs/1512.02325.

Maximo, Camacho, & Gabriel, Perez-Quiros. 2008. Introducing the euro-sting: Short-term indicator of euro area growth. *Journal of Applied Econometrics*, **25**(4), 663–694. https://onlinelibrary.wiley.com/doi/abs/10.1002/jae.1174.

Modugno, Michele, Soybilgen, Barış, & Yazgan, Ege. 2016. Nowcasting Turkish GDP and news decomposition. *International Journal of Forecasting*, **32**(4), 1369 – 1384. http://www.sciencedirect.com/science/article/pii/S0169207016300693.

Pan, S. J., & Yang, Q. 2010. A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering*, **22**(10), 1345–1359.

Ren, Shaoqing, He, Kaiming, Girshick, Ross B., & Sun, Jian. 2015. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *CoRR*, **abs/1506.01497**. http://arxiv.org/abs/1506.01497.

Rusnák, Marek. 2016. Nowcasting Czech GDP in real time. *Economic Modelling*, **54**, 26 – 39. http://www.sciencedirect.com/science/article/pii/S0264999315004034.

secretaría de movilidad de Medellín. 2018. *Mapas SIMM*. https://www.medellin.gov.co/simm/mapas/index.html?map=camarasCctv. Accessed: 2018-10-03.

Sermanet, Pierre, Eigen, David, Zhang, Xiang, Mathieu, Michaël, Fergus, Rob, & LeCun, Yann. 2013. Overfeat: Integrated recognition, localization and detection using convolutional networks. *arXiv preprint arXiv:1312.6229*.

Shafiee, Mohammad Javad, Chywl, Brendan, Li, Francis, & Wong, Alexander. 2017. Fast YOLO: A Fast You Only Look Once System for Real-time Embedded Object Detection in Video. *CoRR*, **abs/1709.05943**. http://arxiv.org/abs/1709.05943.

Urasawa, Satoshi. 2014. Real-time GDP forecasting for Japan: A dynamic factor model approach. *Journal of the Japanese and International Economies*, **34**, 116 – 134. http://www.sciencedirect.com/science/article/pii/S0889158314000422.

Wei, William WS. 2006. Time series analysis. *In: The Oxford Handbook of Quantitative Methods in Psychology: Vol. 2*.

Wu, Bichen, Iandola, Forrest N., Jin, Peter H., & Keutzer, Kurt. 2016. SqueezeDet: Unified, Small, Low Power Fully Convolutional Neural Networks for Real-Time Object Detection for Autonomous Driving. *CoRR*, **abs/1612.01051**. http://arxiv.org/abs/1612.01051.