A nowcasting model for Medellín City

María Camila Vásquez Correa, Juan Carlos Duque, Jairo Alejandro Gómez

Research Group in Spatial Economics, School of Sciences, Universidad EAFIT

mvasqu49@eafit.edu.co, jduquec1@eafit.edu.co, jagomeze@eafit.edu.co

Introduction

The unemployment rate in Colombia helps the government tracking the state of the economy, and it is the input for new policies and the redesign of existent ones for improving the quality of life of Colombians. The monthly period used to update the unemployment rate [1] and the three months used to update the GDP do not allow the policy makers to react fast enough to changes in the market or to extreme circumstances. They have to wait a long time until the indicator is released to make decisions with it. The underlying hypothesis of this research 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 can be an indicator of the unemployment, and therefore a good feature for estimating 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 real-time forecasting for Medellín's unemployment rate.

Figure 1: Left: normal vehicular flow. Right: vehicular flow during an economic crisis.

Methodology

This research has three main stages. The first one is the characterization of unemployment in Medellín through the analysis of time-series and it aims to provide an insight of its behavior. The second one is the building of a regression model using counts from vehicles and the historic data of the unemployment rate. Finally, the third stage aims to build a regression model capable of forecasting the unemployment rate using only vehicle counts. In this stage, we focused on the definition of a pipeline for vehicle counting, trough the building of a data set for the training of pre-built object detection models. As part of this work, we explored:

- The cameras to monitor based on their coverage.
- The vehicle's categories to count.
- The object detection model to use for the counting.

Using these variables, the objective is to search for a neural network architecture that provides the necessary elements to perform the detection of vehicles in an image.

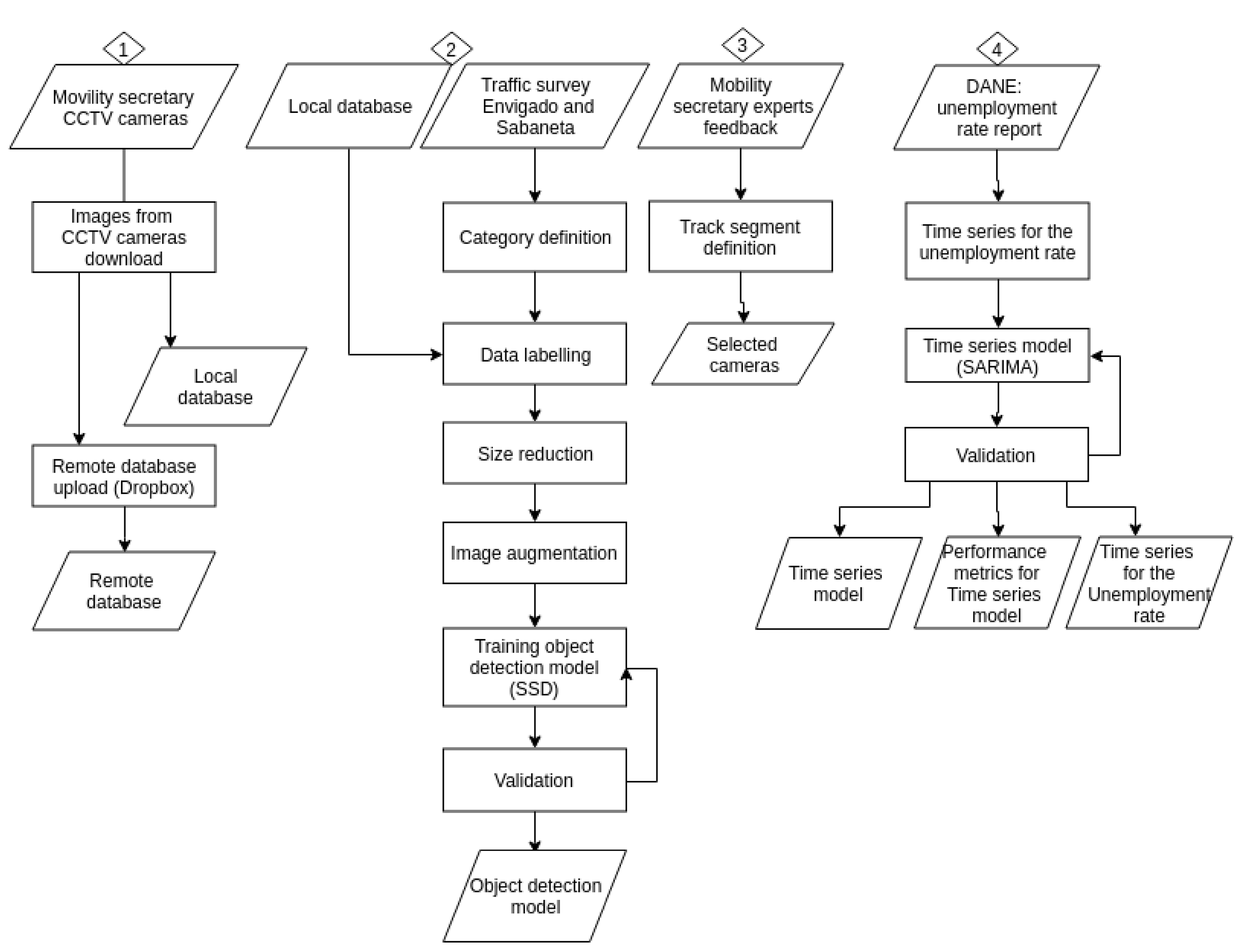

Figure 2: Flow diagram that guides this research.

Results

- **Time series analysis**

We performed a time-series analysis and found that the unemployment rate has hysteresis and seasonality. Finally, we fitted a SARIMA$(2,1,2) \times (2,0,1,12)$ to the series and obtained the results shown in Figure 3.

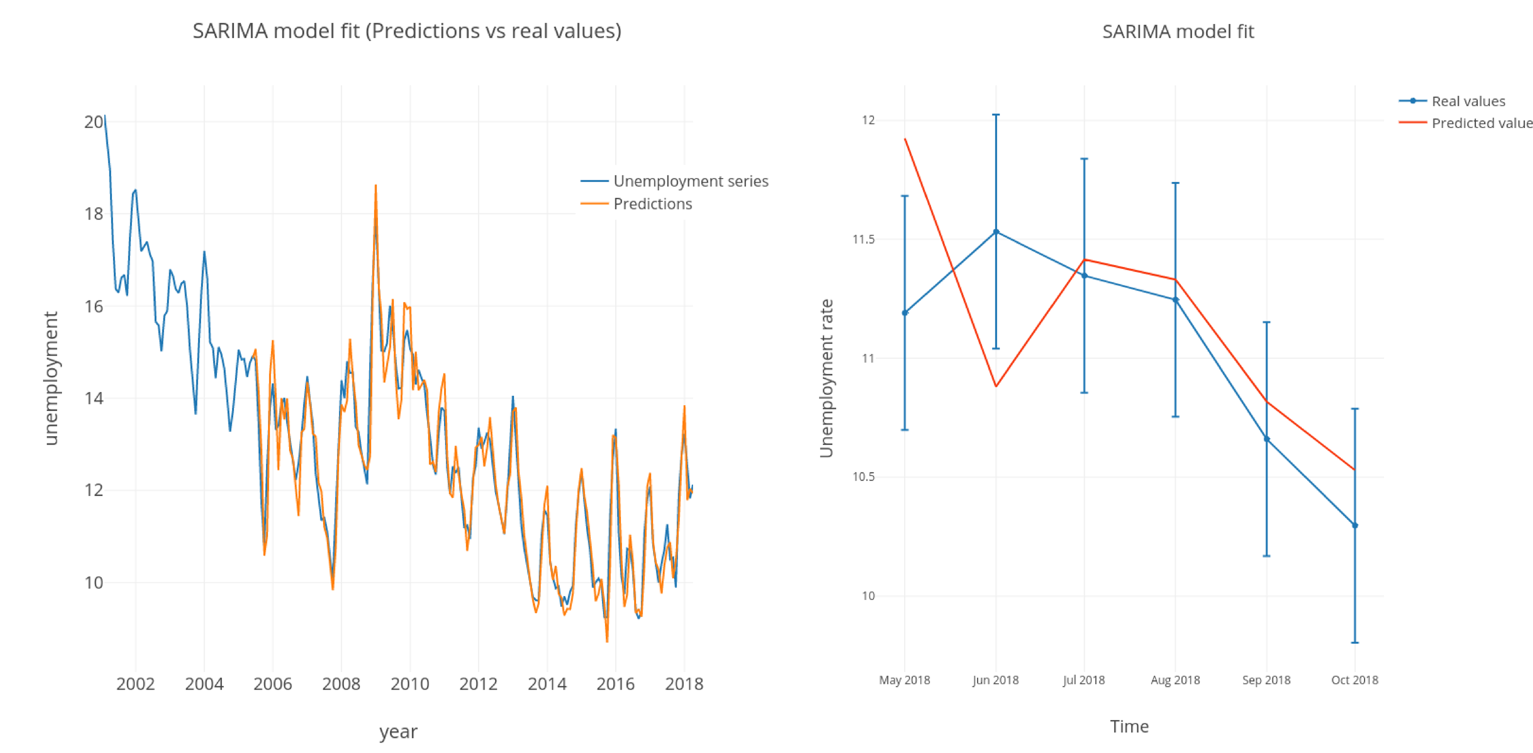

Figure 3: Predictions of the unemployment rate using a SARIMA model.

- **Object detection model**

Taking images from CCTV cameras installed in Medellín, we trained a Single Shot Detector [2], using different sets of hyper-parameters, to perform the detection of selected categories of vehicles in chosen camera images based on the registered cars in the Metropolitan Area of Valle del Aburrá ([3], [4], [5]). We extended the training data-set to 2000 labelled images for training, and 200 images for validation. Also, we performed image augmentation in the training set, in order to generate more examples for training, some of the transformations applied are shown in Figure 5.

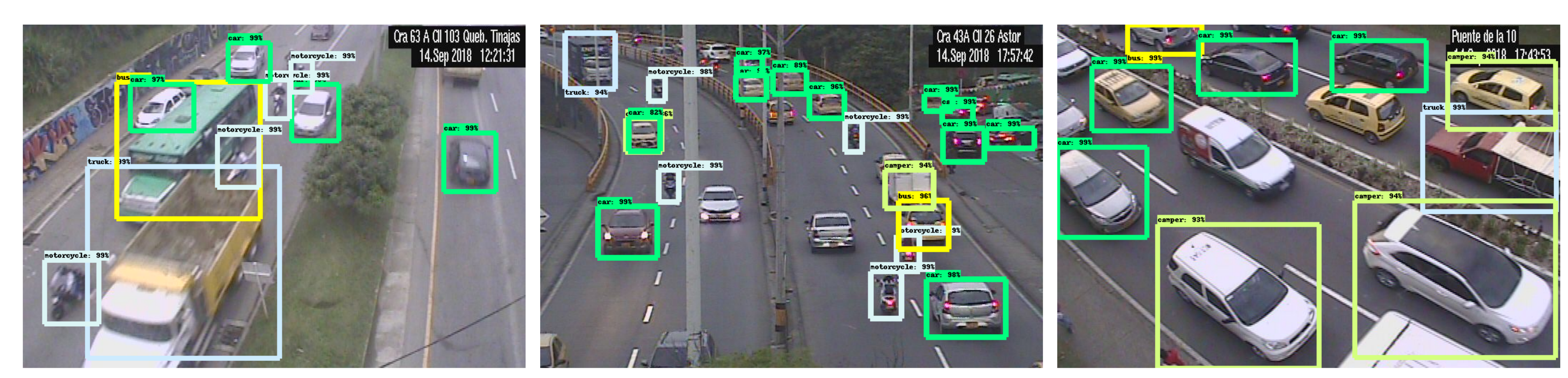

Figure 4: Vehicle detection using a pre-trained Single Shot Detector

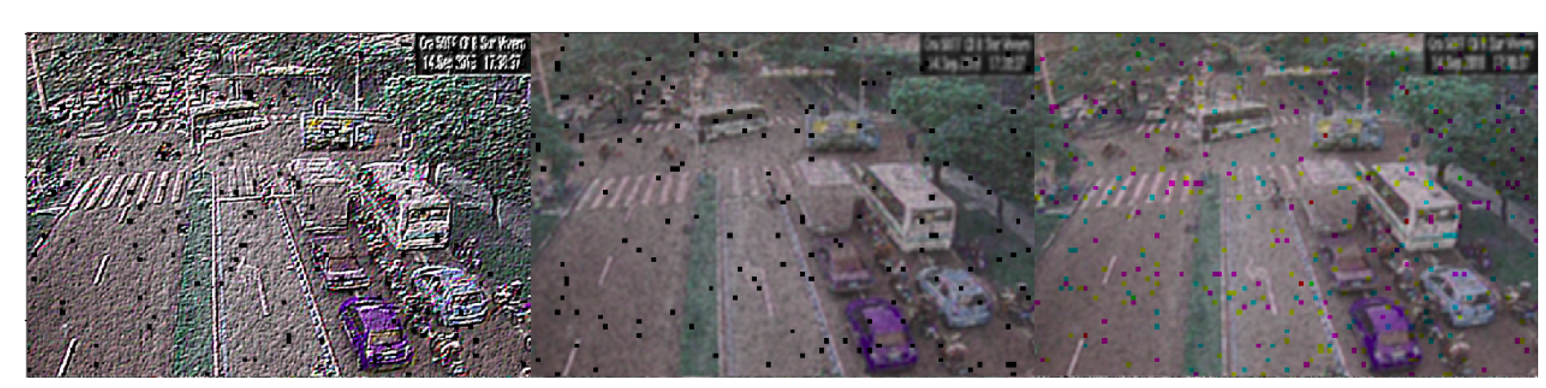

Figure 5: Some transformations applied to the orginінal images.

The models are currently being trained in Apolo Scientific Computing Center.

Conclusions and future work

We are still training the SSD object detector model aiming to improve its performance on the selected classes. We have noticed that given the similarities of some of the categories and the diversity in the orientation of the traffic cameras in Medellín, we need to label more images and improve our transfer learning strategy. Once, we are able to get reliable vehicle counts, we will follow the steps shown in Figure 6.

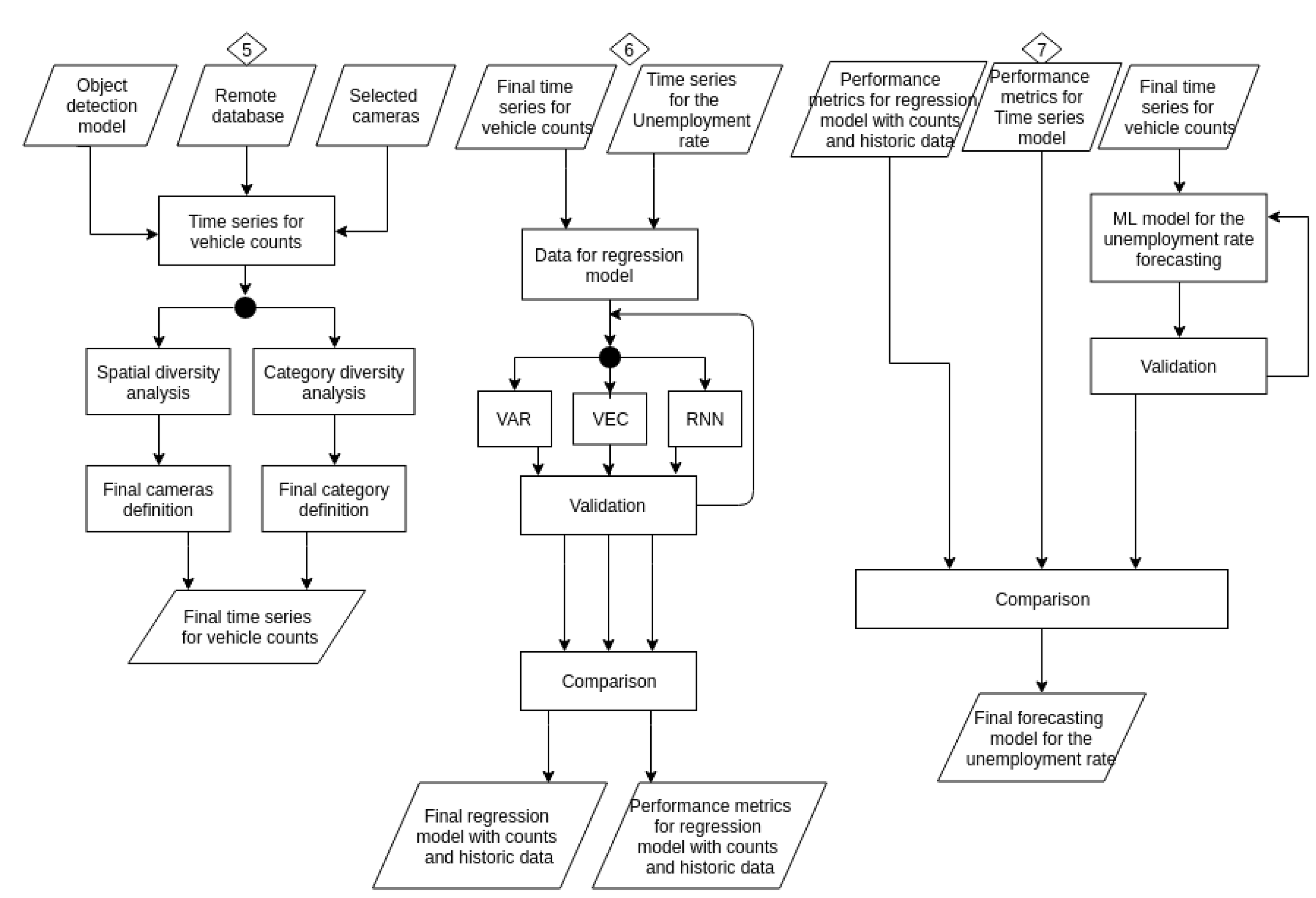

Figure 6: Flow diagram for the next research stage.

References

[1] Departamento Administrativo Nacional de Estadística DANE. Gran encuesta integrada de hogares (geih) mercado laboral. `https://goo.gl/b43Wpb`, Jun 2018.

[2] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C. Berg. SSD: Single shot MultiBox detector. In *Computer Vision – ECCV 2016*, pages 21–37. Springer International Publishing, 2016.

[3] Gobierno digital Colombia Datos Abiertos. Distribución del parque automotor inscrito 2016(ii). `https://goo.gl/KmGLt5`, Ago 2017. Visited November 17 2018.

[4] Gobierno digital Colombia Datos Abiertos. Parque automotor municipio de Envigado hasta 31-07-2018. `https://goo.gl/CQqqrX`, Jul 2018. Visited November 17 2018.

[5] Gobierno digital Colombia Datos Abiertos. Registro histórico del parque automotor de Sabaneta. `https://goo.gl/5hNK1Y`, Oct 2018. Visited November 17 2018.