



A nowcasting model for Medellín city

This research is part of the project: *economic challenges of Latin American Cities within the framework of the SDG, a disruptive vision on how to tackle them.*

*María Camila Vásquez Correa**

*PhD. Juan Carlos Duque***

*PhD. Jairo Alejandro Gómez****

** Mathematical Engineering student- mvasqu49@eafit.edu.co*

*** Research group in Spatial economics - jduque1@eafit.edu.co*

**** Research group in Spatial economics - jagomeze@eafit.edu.co*

Contents

- Unemployment:
 - Justification.
 - Characterization.
 - Time series analysis
- Object detection:
 - Categories.
 - Data collection.
 - First prototype.
 - Track segments.
- Ongoing work.
- Future work.
- References.

Nowcasting the unemployment

Provide information to generate precise and efficient policies to reduce the unemployment rate and contribute to de Sustainable Development Goal 8: *promote sustained, inclusive and sustainable economic growth, full and productive employment and decent work for all (ONU).*

Characterization (1/2)

The unemployment rate is an indicator of the degree of use of human resources in the economy and is the ratio between the number of people seeking for a job and the number of people in the labor force:

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

Where DS is the number of people seeking for a job and EAP stands for Economically Active Population (DANE, 2016).

Characterization (2/2)

For Medellín, that is one of Colombia's capital cities, the unemployment rate is calculated every month for the city and its Metropolitan Area. This indicator is released with a delay of two months.

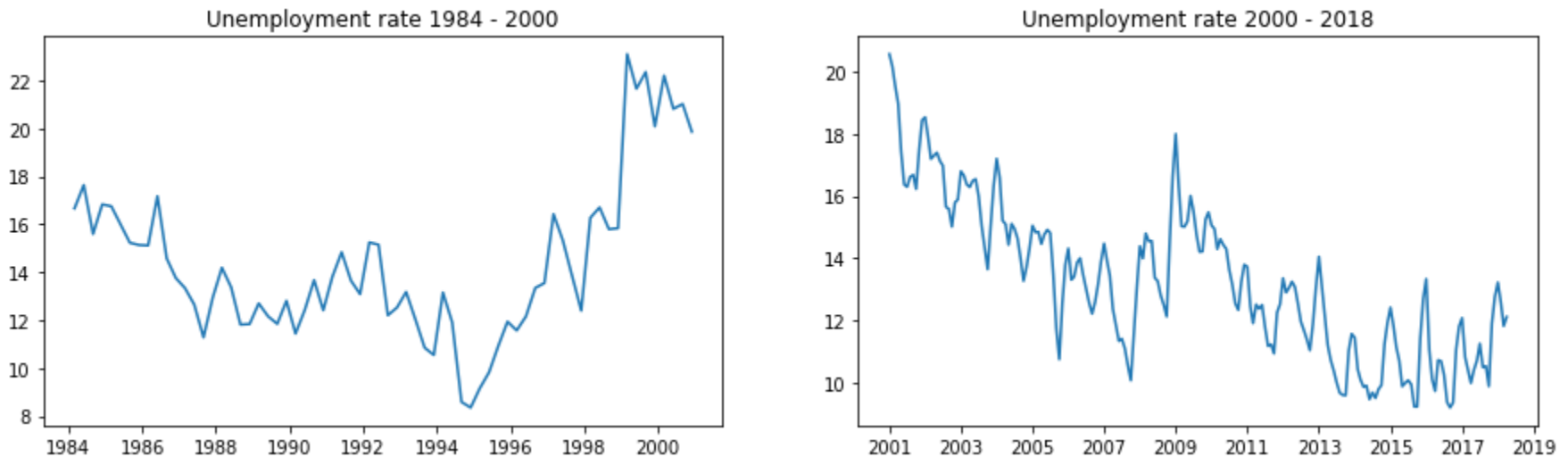


Figure 1: Unemployment rate for Medellín and its Metropolitan Area (DANE, 2018).

Time series analysis (1/3)

The series for the unemployment show evidence of non-stationarity by different tests (Dickey Fuller, KPSS), so we begin by a log transformation of the data (Box & Cox, 1964).

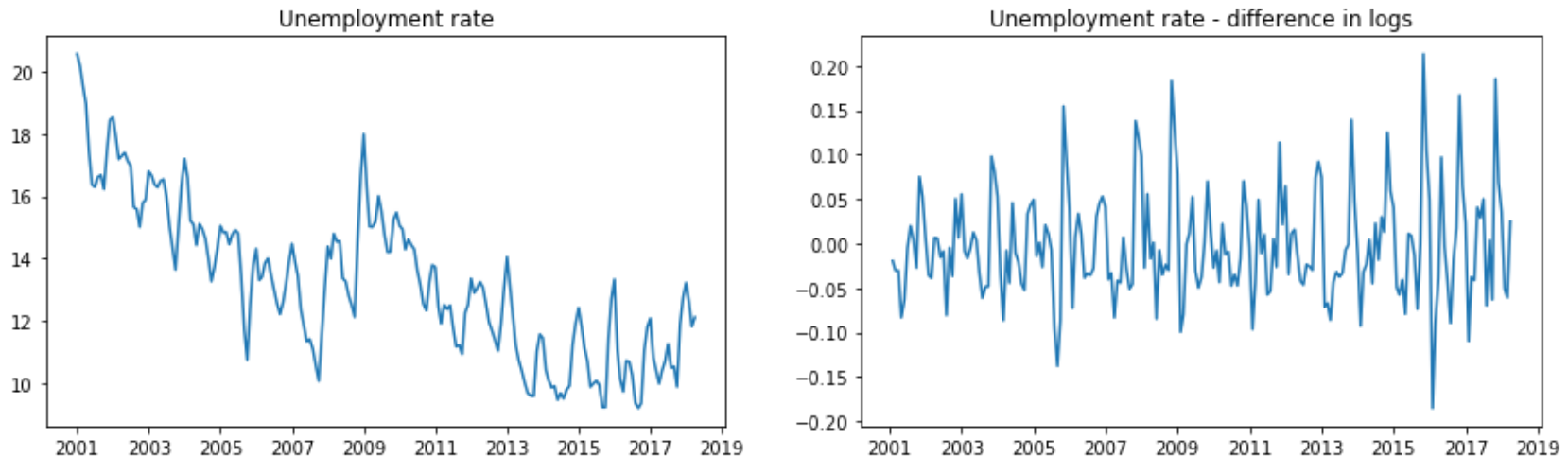


Figure 2: Left: Unemployment rate. Right: Differentiated unemployment rate, Box-Cox Logarithm transformation.

Time series analysis (2/3)

The seasonal behavior of the city's economy is also shown in the data.

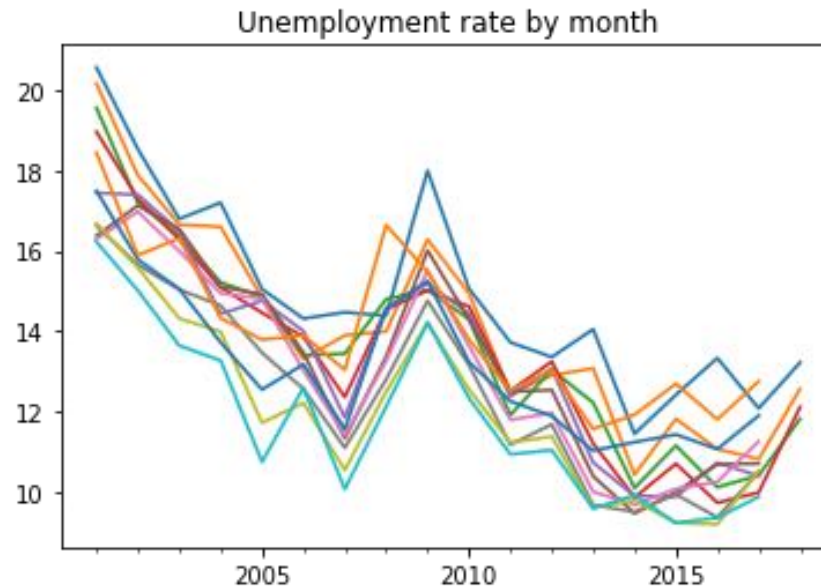


Figure 3: Unemployment rate for each month.

Time series analysis (3/3)

The model fitted was a SARIMA(2,1,2)x(2,0,2,12), chosen by the Akaike's information criteria, that shows evidence of hysteresis and seasonality.

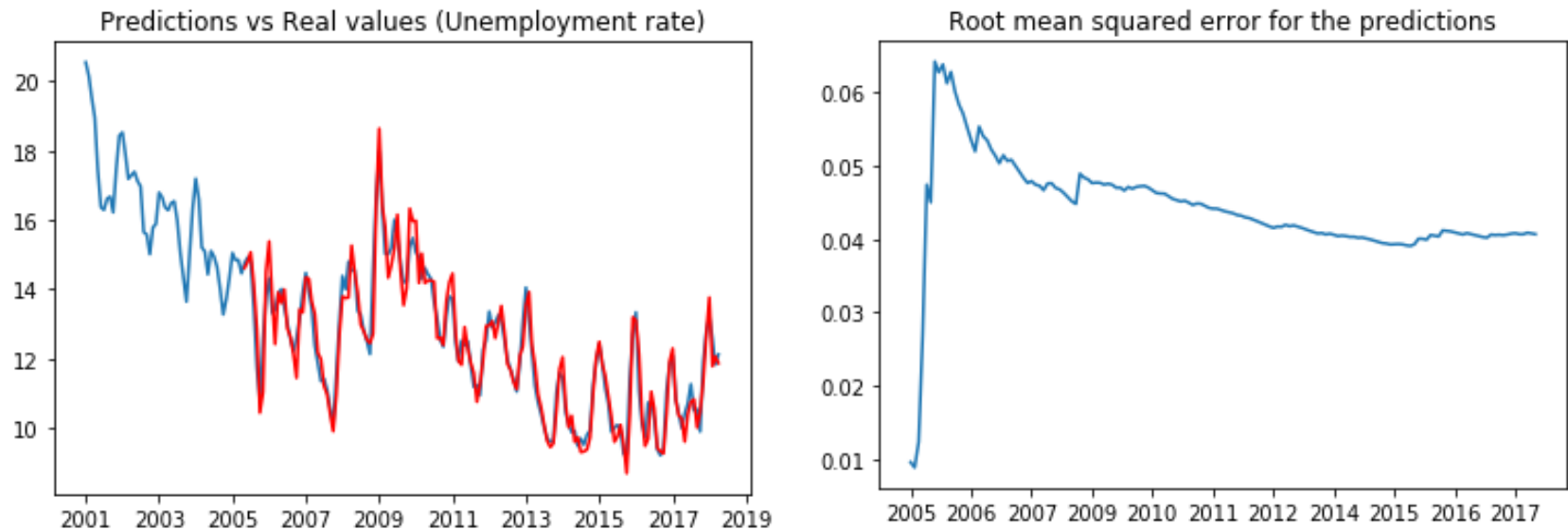


Figure 4: Results of the fitted model. Left: Predictions with the training. Right: Root mean squared error (RMSE).

Categories

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

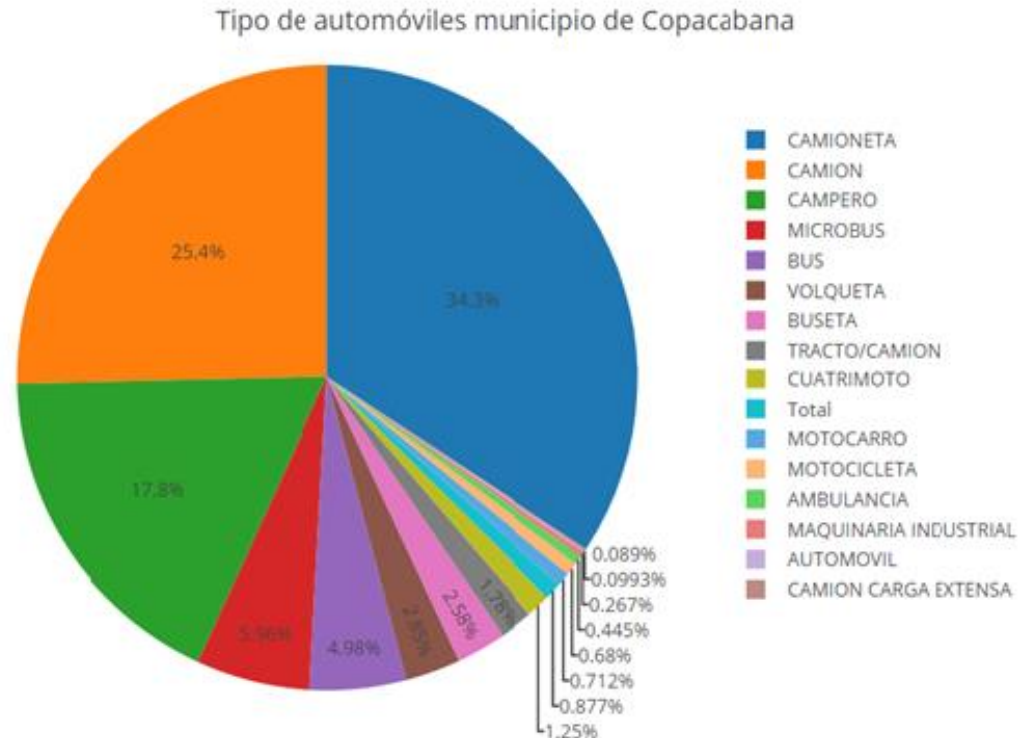


Figure 5: Counts for category in Copacabana. (Datos Abiertos, 2018 a), (Datos Abiertos, 2018 b) and (Datos Abiertos, 2017).

Data collection

Images from traffic cameras are acquired and stored every minute.



Figure 6: Images from a few CCTV cameras installed in Medellín.

<https://www.medellin.gov.co/simm/mapas/index.html?map=camarasCctv>

First prototype

Preliminary results of a vehicle detector model (SSD Mobile NET).

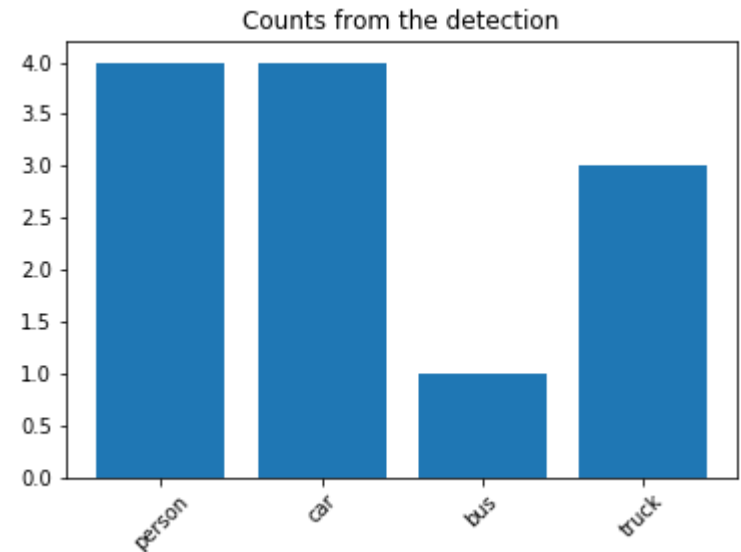
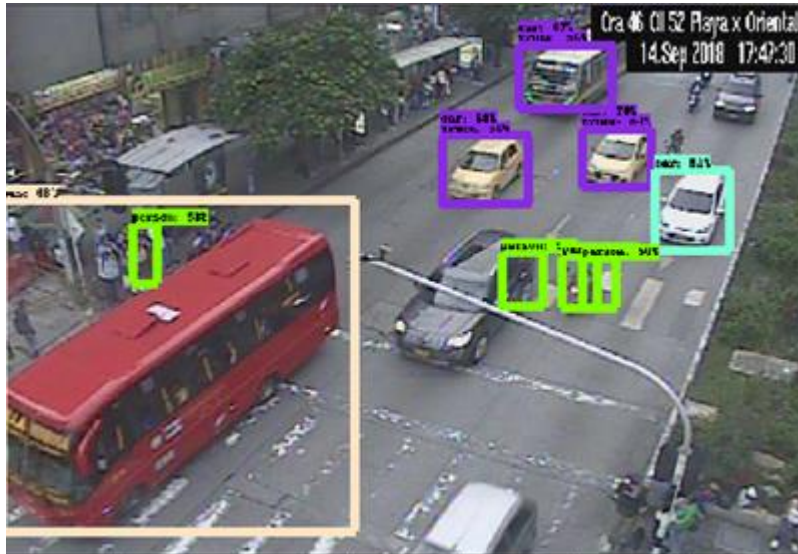


Figure 7: Vehicle detection and counting from an image of traffic located in Medellín.

Track segments

Following an expert criteria, the following track segments were chosen to be monitored:

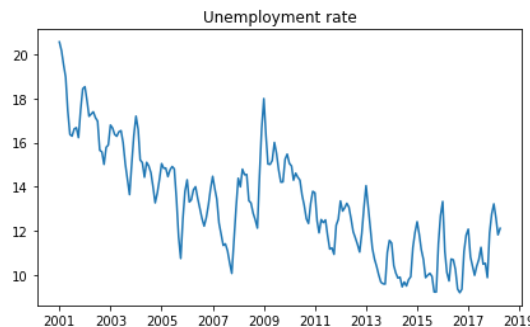
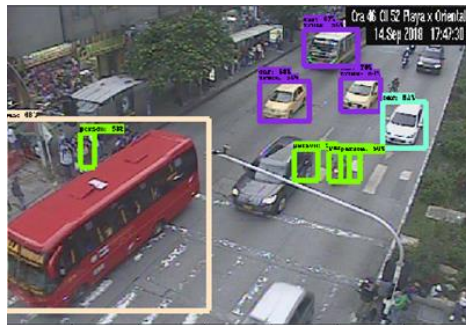
- South Freway.
- North Freeway.
- San Juan Avenue.
- Iguaná.
- 80th Avenue.
- 33 Avenue.
- Via Las Palmas.
- Avenue El Poblado.
- 30th Avenue.
- Avenue Guayabal.
- East Avenue.

Ongoing work

- Labelling the selected categories.
- Collecting data from historical videos from Medellin.
- Defining strategies to compare the counting series with the historical unemployment series.

Future work (1/2)

Forecasting unemployment along with the data from the counts of the vehicles. Is this information really significant?



New
forecasting
model.

Options:

- Machine learning.
- VAR model

Figure 8: Using counts and the historical unemployment rate a new model is trained.

Future Work (2/2)

Forecasting unemployment using counts only

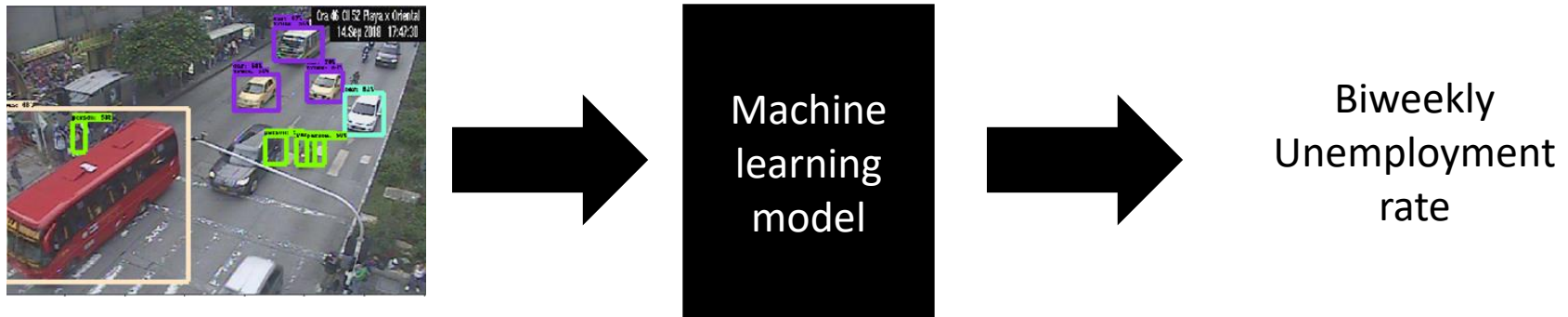


Figure 9: Using counts, a machine learning model is trained to produce a biweekly forecast of the unemployment rate.

References

- Box, G, & Cox, David. 1964. An Analysis of Transformations (with Discussion). 26(01), 211–252
- 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-Sabaneta/bewj-2mvn>.

Thank you for your attention!

Any questions?

María Camila Vásquez Correa - mvasqu49@eafit.edu.co

PhD. Juan Carlos Duque - jduquec1@eafit.edu.co

PhD. Jairo Alejandro Gómez - jagomeze@eafit.edu.co

