

# Proyecto de Titulación

Mateo Quiguiri

3/07/2022

## Pronostico del monto de crédito quirografario del BIESS con datos abiertos de internet y técnicas de Machine Learning (ML).

### Introducción:

De acuerdo con Haselbeck (2022), predecir la demanda futura para respaldar el análisis corporativo y la toma de decisiones es una ventaja competitiva potencial en muchas áreas, entre estas la economía. No obstante, la literatura sobre la demanda de crédito es relativamente escasa en comparación con los estudios sobre la demanda de dinero.

Medeiros (2019) afirma que existe evidencia de que el uso de métodos ML con predictores no tradicionales pueden mejorar la predicción de variables macroeconómicas como la inflación.

Aunque los algoritmos de ML se están volviendo más comunes en la literatura de pronóstico, no está claro que sean superiores a los métodos tradicionales, sino que dependen de la aplicación y los datos. Haselbeck (2022)

### Objetivos y Justificación:

#### Objetivo General:

- Contrastar la utilidad de implementar nuevas fuentes de datos y métodos de Machine Learning para el pronóstico de monto de colocación de crédito quirografario.

#### Objetivos Específicos:

- Caracterizar el conjunto de variables disponible y necesario para la predicción del monto de colocación de préstamos.
- Determinar si los avances en métodos de Machine Learning hace posible mejorar la predicción del monto de colocación de crédito.

### Datos

Periodicidad de las series: **Mensual**

Periodo de estudio: **Octubre de 2010 a Septiembre de 2021**

#### Fuentes de información:

- Superintendencia de Bancos: Series del sistema financiero nacional (*Credito total*)
- Banco Central del Ecuador (BCE): Series macroeconómicas (*inflación, tasa activa, precio\_wti*)
- Google Trends: Series de Datos abiertos de búsquedas de frase de Google (*Google queries*)

fecha	credito_total	google_q1	google_q2	google_q3	google_q4	inflacion	tasa_activa	tasa_pasiva
2010-10-01	23.8794	0	0	0	0	0.254321	8.94	4.30
2010-11-01	41.5186	0	0	0	0	0.270719	8.94	4.30
2010-12-01	36.9411	0	0	0	0	0.514367	8.68	4.28
2011-01-01	50.9040	4	4	0	0	0.679188	8.59	4.55
2011-02-01	77.8330	14	5	0	14	0.554961	8.25	4.51
2011-03-01	85.0643	4	17	0	4	0.336827	8.65	4.59
2011-04-01	80.3985	8	8	0	4	0.820080	8.34	4.60
2011-05-01	80.4887	4	11	0	4	0.345850	8.34	4.60
2011-06-01	76.9657	7	15	0	4	0.039772	8.37	4.58
2011-07-01	83.0625	11	7	0	7	0.175822	8.37	4.58

#### Variables:

- “total\_credito” Monto mensual de colocacion del Banco Pichincha de PC
- “inflacion” Riesgo proveniente del mercado real
- “tasa\_activa” Riesgo del sector financiero
- “precio\_wti” Precio del petróleo WTI
- “google\_q1” Numero de busquedas de “Credito Quirografario”
- “google\_q2” Numero de busquedas de “Credito de Consumo”
- “google\_q3” Numero de busquedas de “Credito quirografario BIESS”
- “google\_q4” Numero de busquedas de “Credito consumo Banco BIESS”

#### Muestra de los datos:

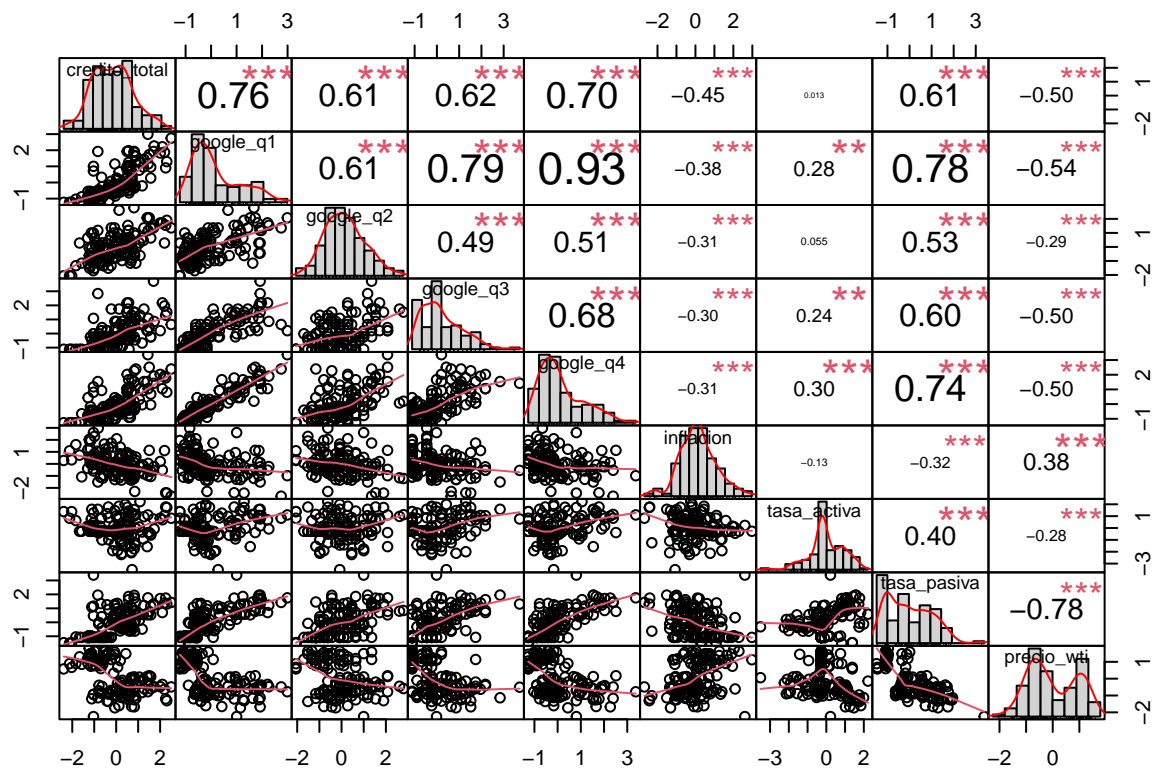
#### Normalización de las variables

Para evitar un análisis sesgado en el análisis posterior, todas las variables se han normalizado.

## Estadística Descriptiva:

	Overall
	(N=132)
<b>credito_total</b>	
Mean (SD)	140 (48.6)
Median [Min, Max]	138 [23.9, 262]
<b>google_q1</b>	
Mean (SD)	25.3 (20.6)
Median [Min, Max]	18.5 [0, 87.0]
<b>google_q2</b>	
Mean (SD)	11.4 (5.57)
Median [Min, Max]	11.0 [0, 27.0]
<b>google_q3</b>	
Mean (SD)	5.08 (4.55)
Median [Min, Max]	4.00 [0, 22.0]
<b>google_q4</b>	
Mean (SD)	10.4 (8.20)
Median [Min, Max]	8.00 [0, 38.0]
<b>inflacion</b>	
Mean (SD)	0.163 (0.322)
Median [Min, Max]	0.153 [-0.704, 1.12]
<b>tasa_activa</b>	
Mean (SD)	8.30 (0.464)
Median [Min, Max]	8.17 [6.67, 9.33]
<b>tasa_pasiva</b>	
Mean (SD)	5.19 (0.608)
Median [Min, Max]	5.10 [4.28, 7.24]
<b>precio_wti</b>	
Mean (SD)	68.6 (23.1)
Median [Min, Max]	62.3 [16.5, 110]

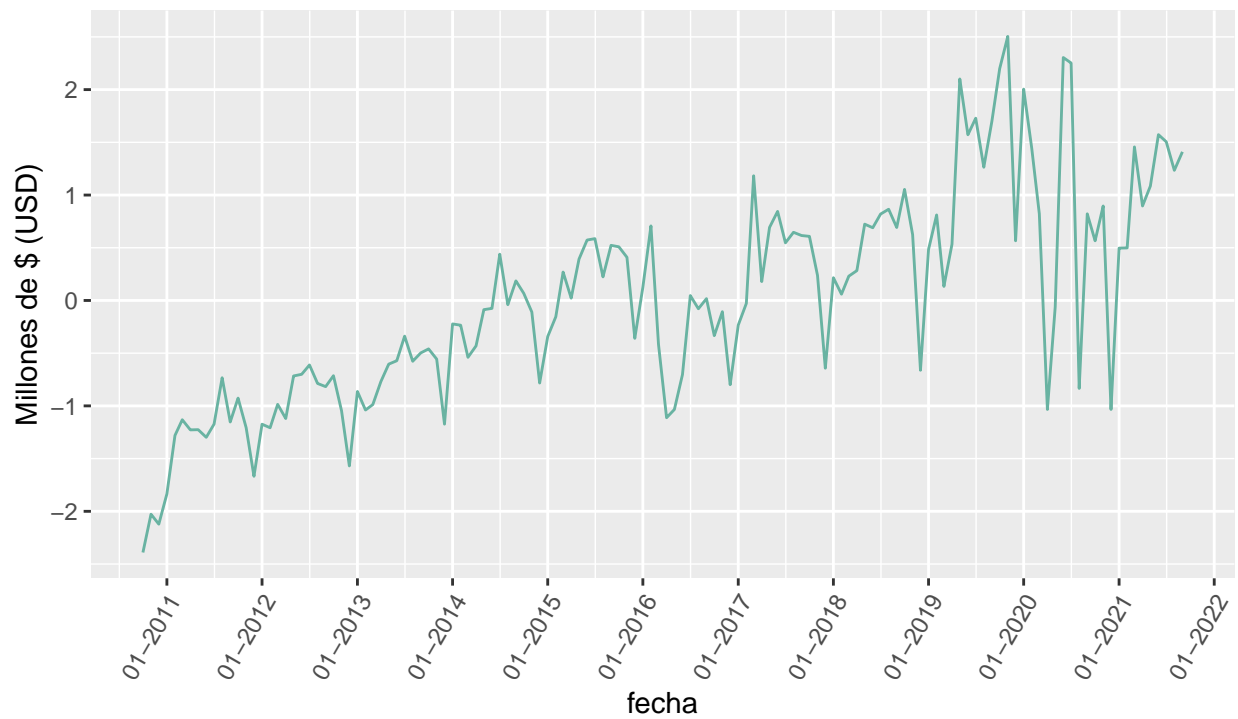
## Correlación



Variable objetivo: Monto de colocación de crédito quirografario

### Colocacion Credito Credito Quirografario BIESS

De Oct 2010 a Sep 2021



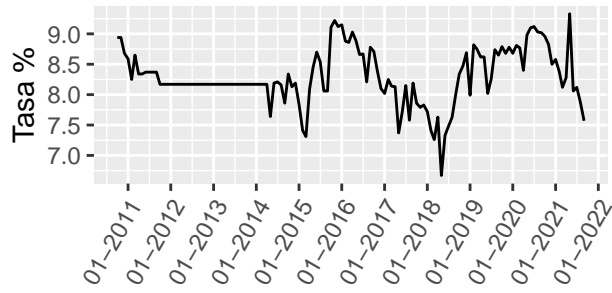
Fuente: BIESS

Como se observa en el gráfico, desde la fecha de creación del BIESS, el monto de crédito tiene una tendencia creciente y regular que se interrumpe en el año 2016 debido a cambios en las condiciones de los créditos, como los montos, plazos y tasas. También se aprecia el shock debido por el COVID a partir de 2020. Es evidente la presencia de estacionalidad y la no covarianza-estacionaridad de la serie.

#### Variables Explicativas:

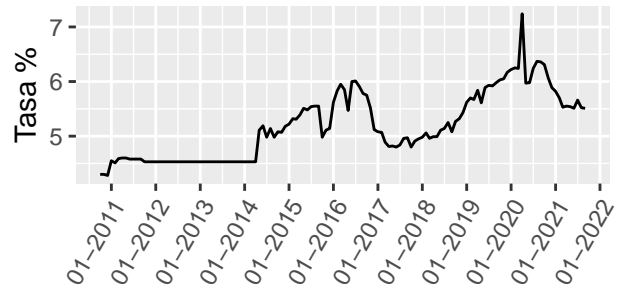
##### Indicadores Macroeconómicos

Tasa de Interes Activa



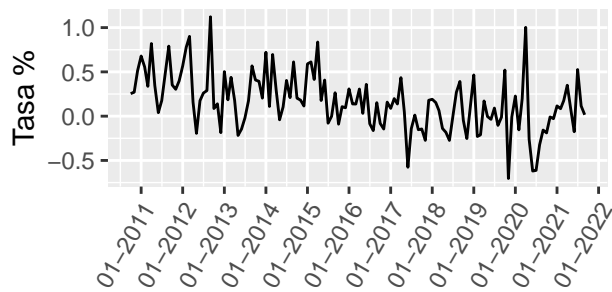
Fuente: Banco Central del Ecuador

Tasa de Interés Pasiva



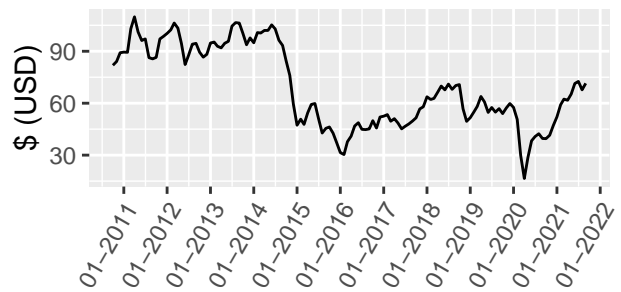
Fuente: Banco Central del Ecuador

Inflación mensual



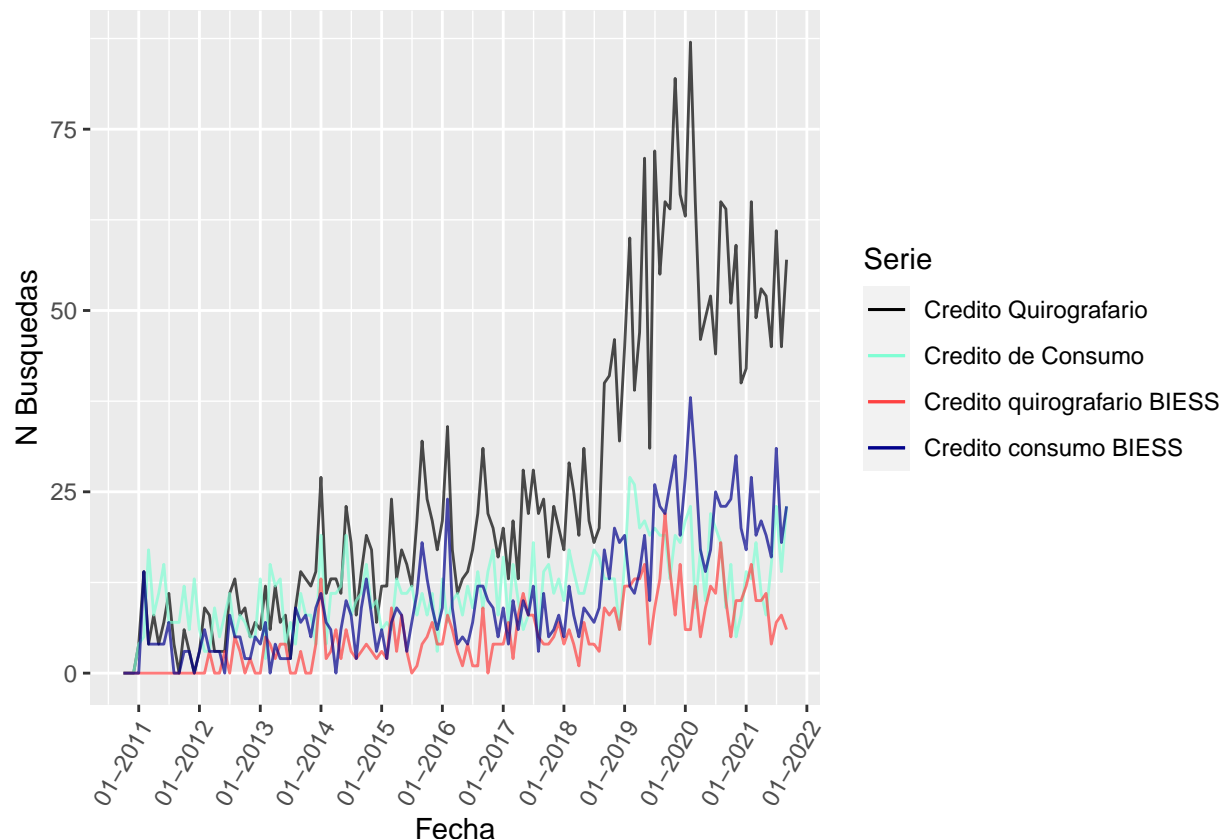
Fuente: Banco Central del Ecuador

Cotización barril WTI



Fuente: Banco Central del Ecuador

Datos abiertos de Google Queries



Como se puede observar la frase “Crédito Quirografario” es la más buscada. La serie mantiene un crecimiento relativamente estable hasta 2018 donde el la tendencia incrementa y se mantiene hasta justo antes la crisis del COVID donde se evidencia una disminución y estancamiento alrededor de las 50 búsquedas mensuales.

### Estudio comparativo experimental:

Dado el objetivo de la investigación y la disponibilidad de datos se aplicará métodos de Machine Learning, entre estos **Long Short-Term Memory (LSTM)**, **Gradient Boosting Decision Tree (GBDT)**, **XGBoost (XGB)** y **Random Forest (RF)**, pero también se aplicarán técnicas econométricas como **ARIMA** y **ARIMAX**, de manera que se pueda comparar el poder de predicción y seleccionar el modelo el que mejor se desempeñe al pronosticar la variables de interés.

Dada la condicion preeliminar del estudio, solo se mostraran los pronósticos de tres modelos, **ARIMA**, **Random Forest** y **XGBoost**.. En todos los casos, se ha utilizado el **95% de los datos para el entrenamiento** del modelo y el **5% para la evaluación**, la razón tiene relacion con el COVID debido a que es un evento reciente, y ha afectado en gran manera la variable objetivo, sería incorrecto entrenar el modelo con datos anteriores al COVID pues no capturaria el shock, lo que se reflejaría en un bajo poder de predicción.

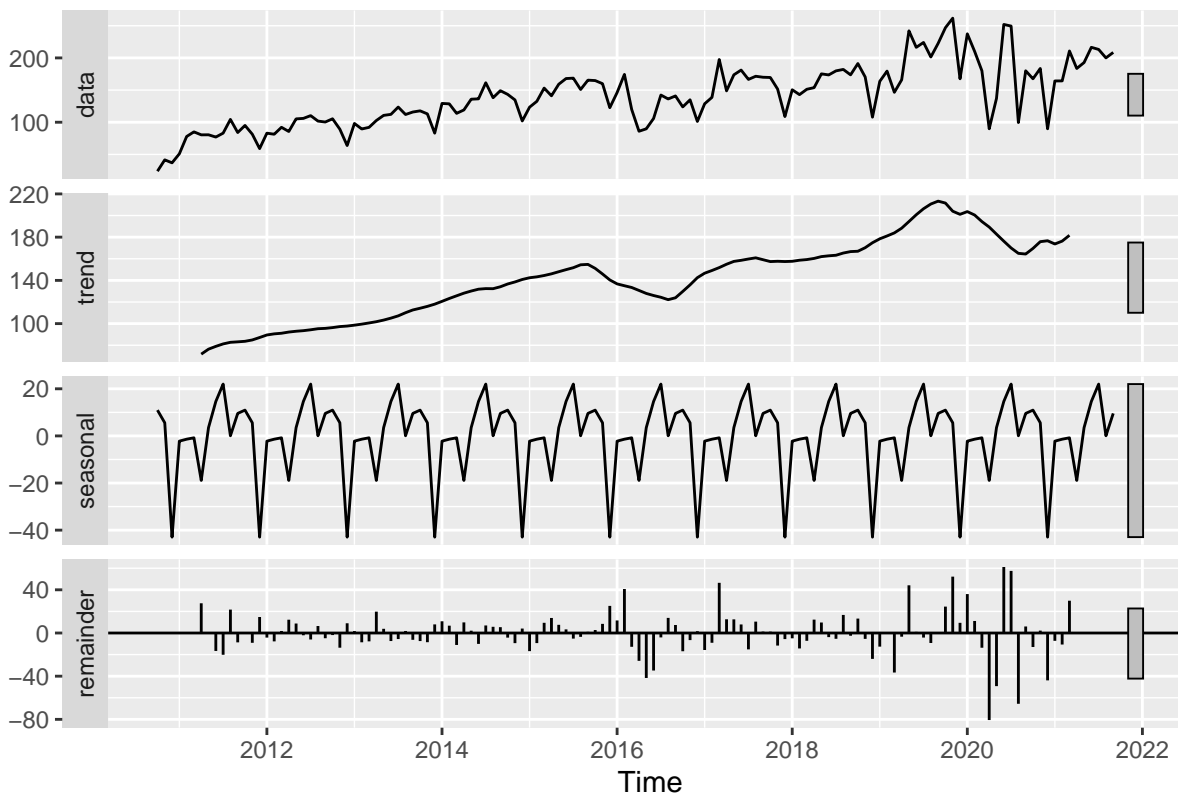
Teniendo en cuenta que los modelos ML aplicados para regresión necesitan que la variable fecha de la base sea representada como variables dummy, se ha ampliado la base, tal que la base que se utilizará para los modelos ML es la siguiente:

### Modelo SARIMA

Al descomponer la serie se puede observar la presencia de una tendencia y estacionalidad, la cual nos da una idea de que la especificación del modelo incluirá una parte estacional y que para eliminar la tendencia, será necesario diferenciar la serie.

credito_total	google_q1	google_q2	google_q3	google_q4	inflacion	tasa_activa	tasa_pasiva	precio_wti
-2.390521	-1.2273561	-2.056537	-1.11448	-1.2664245	0.2832895	1.3848000	-1.4575066	0.5764112
-2.027531	-1.2273561	-2.056537	-1.11448	-1.2664245	0.3342099	1.3848000	-1.4575066	0.6711702
-2.121730	-1.2273561	-2.056537	-1.11448	-1.2664245	1.0908061	0.8240173	-1.4904277	0.8905438
-1.834393	-1.0333890	-1.337906	-1.11448	-1.2664245	1.6026221	0.6299002	-1.0459930	0.9052552
-1.280232	-0.5484713	-1.158248	-1.11448	0.4406159	1.2168620	-0.1034311	-1.1118351	0.9013610
-1.131422	-1.0333890	0.997645	-1.11448	-0.7786986	0.5394940	0.7593116	-0.9801508	1.4893860

Decomposition plot of the Credit Volume Time-Series

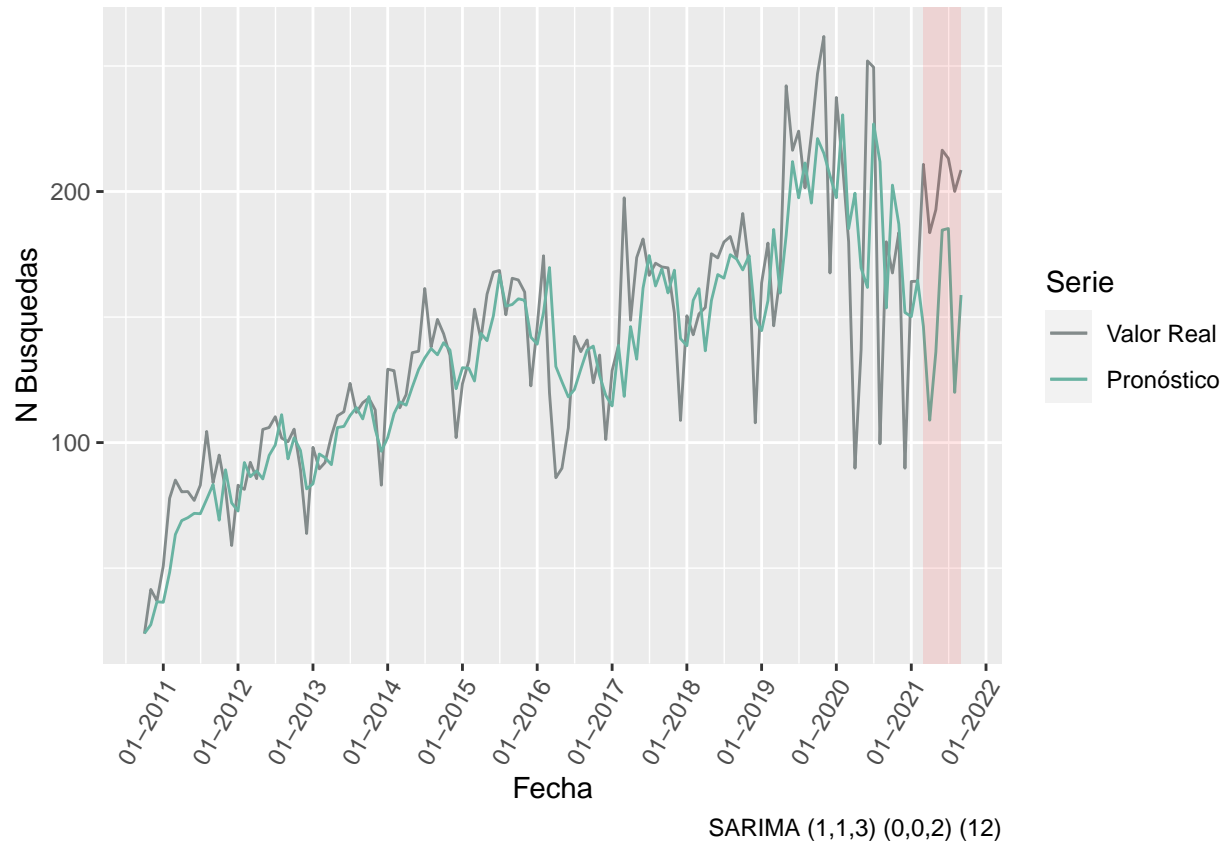


Con la ayuda de la función `autoARIMA` del paquete *forecast*, que automáticamente busca la mejor especificación estimando varios modelos y comparando su calidad mediante criterios de información como AIC, el modelo adecuado es el **SARIMA (1,1,3) (0,0,2) (12)** (La función considera la necesidad o no de aplicar primeras o segundas diferencias de las series)

**Ajuste del modelo** A continuación se observa el ajuste del modelo, tanto en el conjunto de entrenamiento como de prueba.

```
## Don't know how to automatically pick scale for object of type ts. Defaulting to continuous.
```

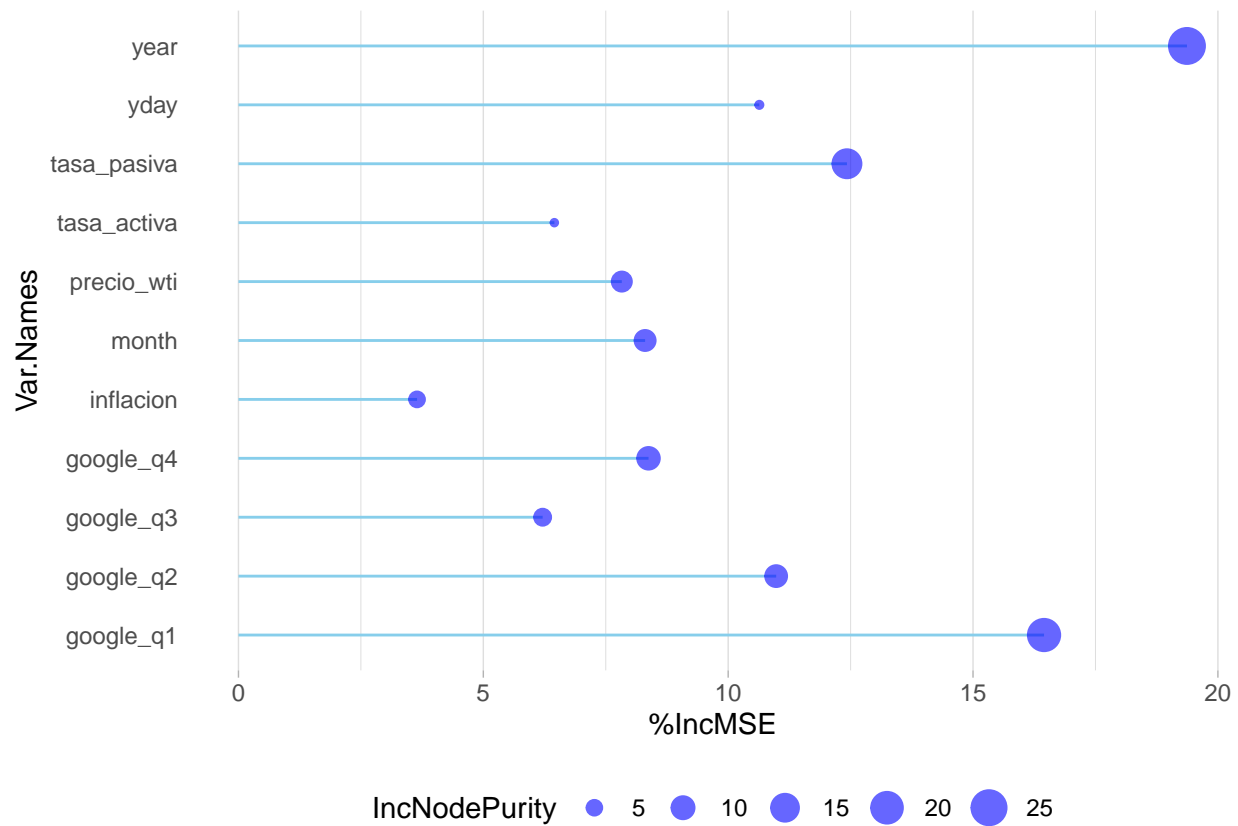




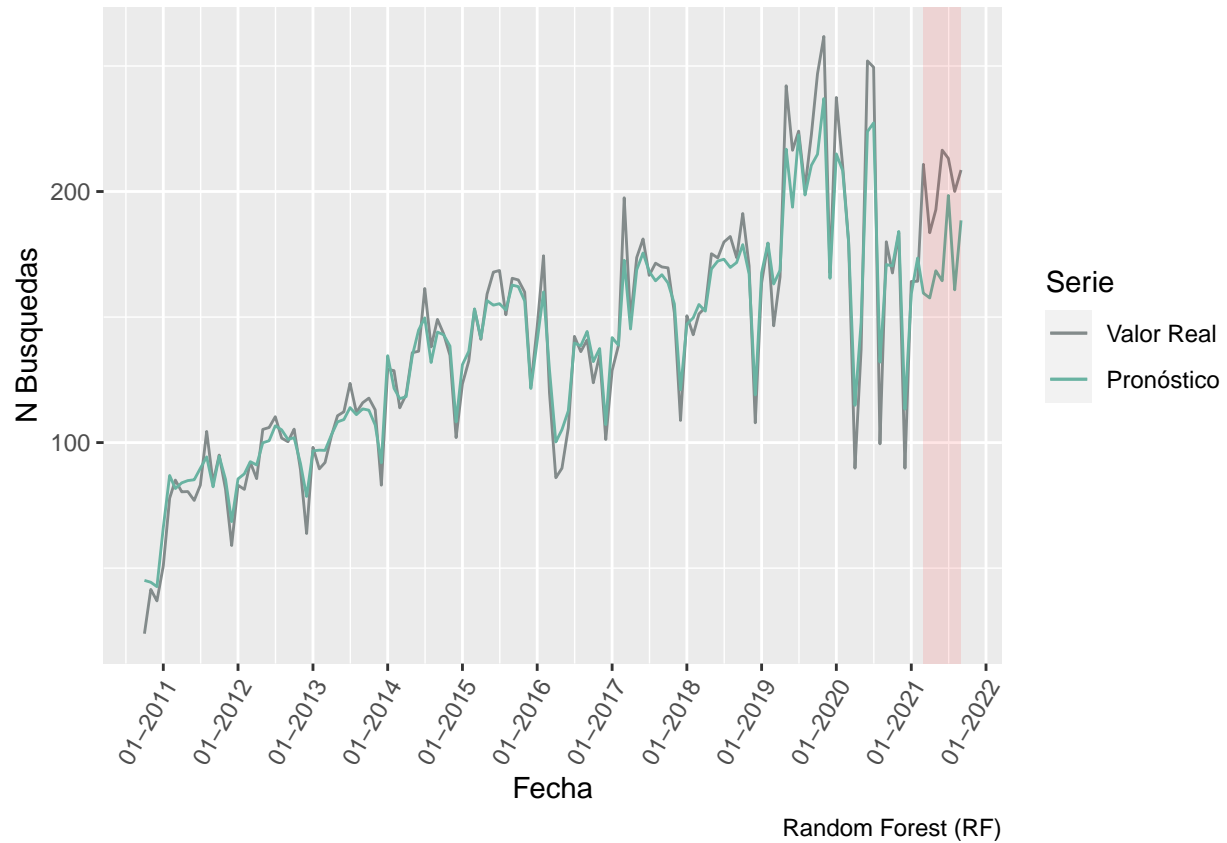
Dada la naturaleza de la serie, la aplicación de los modelos autoregresivos parecen no tener un buen ajuste a pesar de aplicar transformaciones para estabilizar la media, por lo que para mejorar el ajuste de estos modelos se propone la aplicación de transformaciones Box Cox que estabilicen la varianza, posteriormente se compararan nuevamente los resultados.

### Modelo Random Forest RF

**Importancia de las variables** Después de construir el modelo podemos sacar un índice de importancia de las variables, de acuerdo a que tanta influencia tienen sobre la variable dependiente.



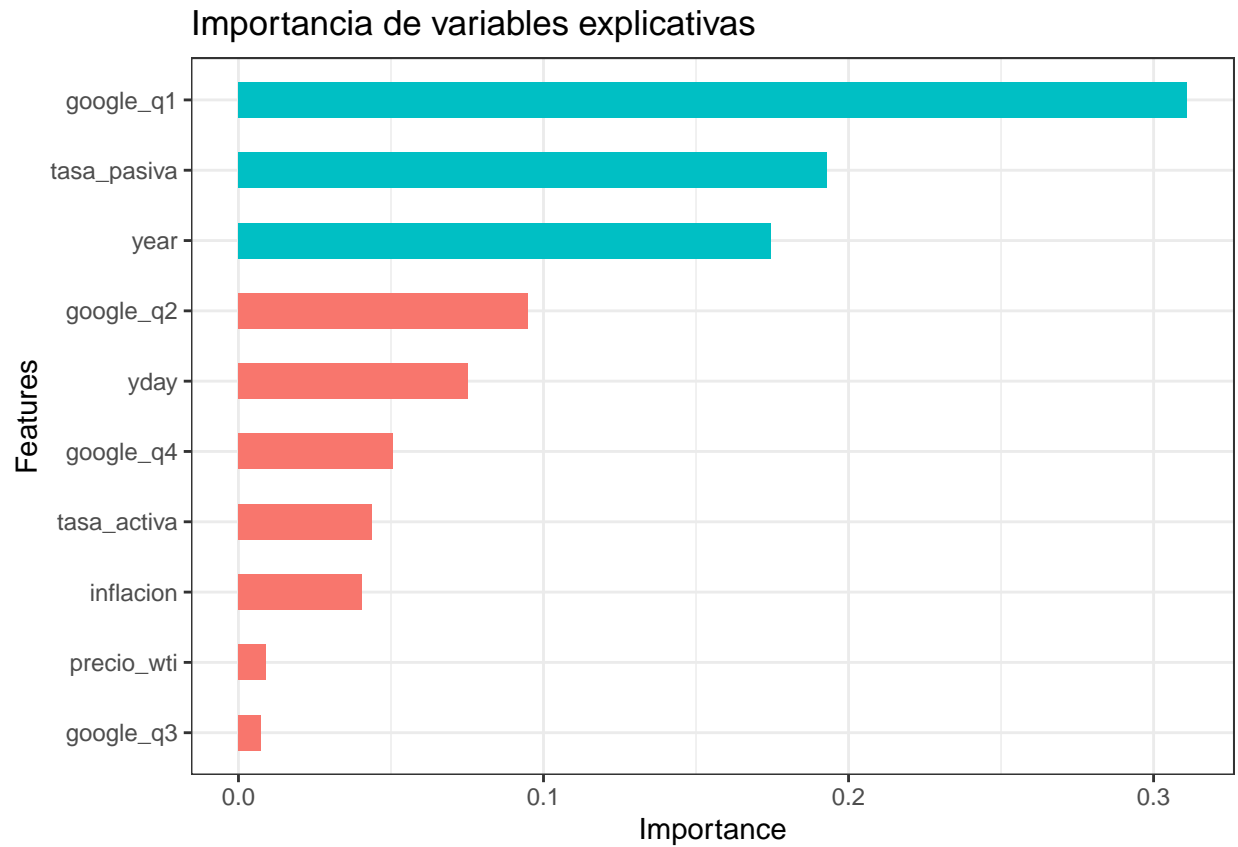
**Ajuste del modelo** En el siguiente grafico podemos observar el ajuste del modelo, tanto en el conjunto de entrenamiento como de prueba.



Aquí se puede observar que el modelo RF se comporta mejor que un modelo autoregresivo.

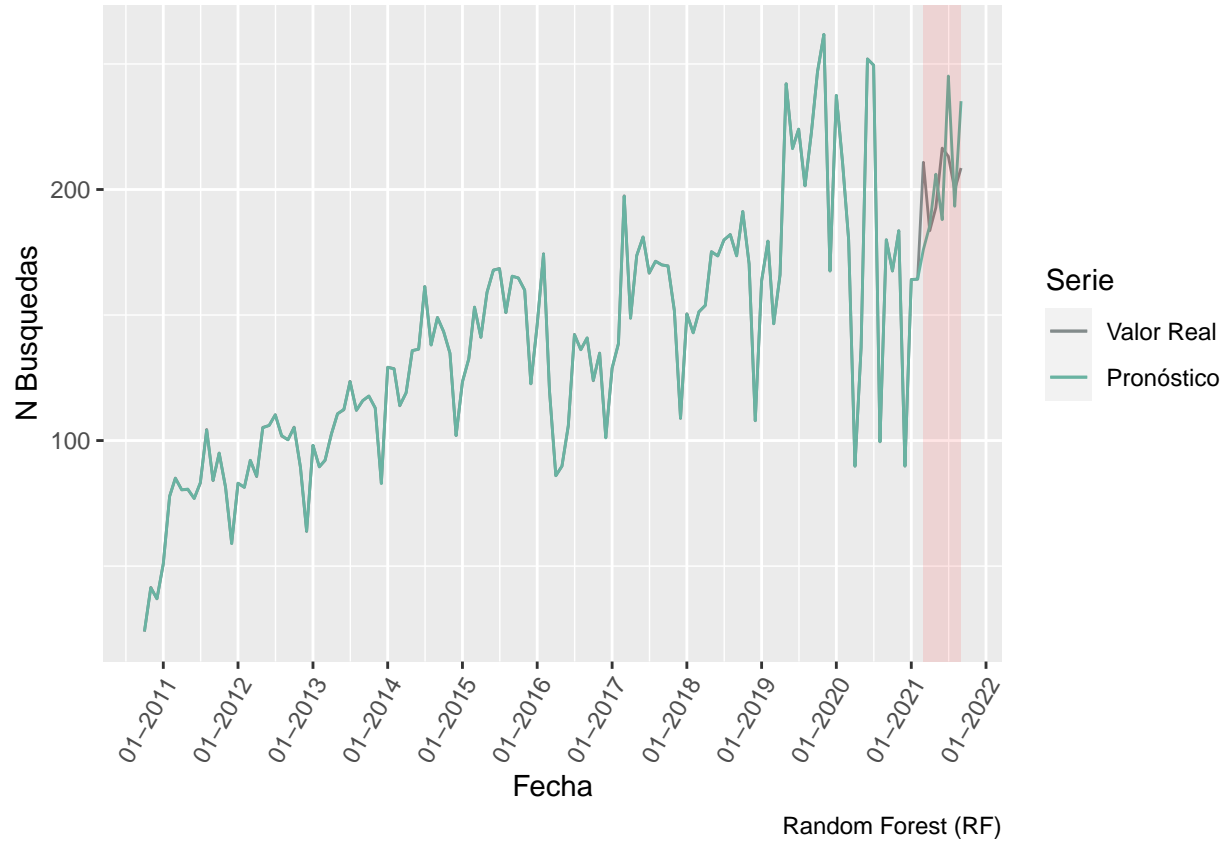
### Modelo Xtreme Gradient Boostin XGB

**Importancia de las variables** Después de construir el modelo podemos sacar un índice de importancia de las variables, de acuerdo a que tanta influencia tienen sobre la variable dependiente.



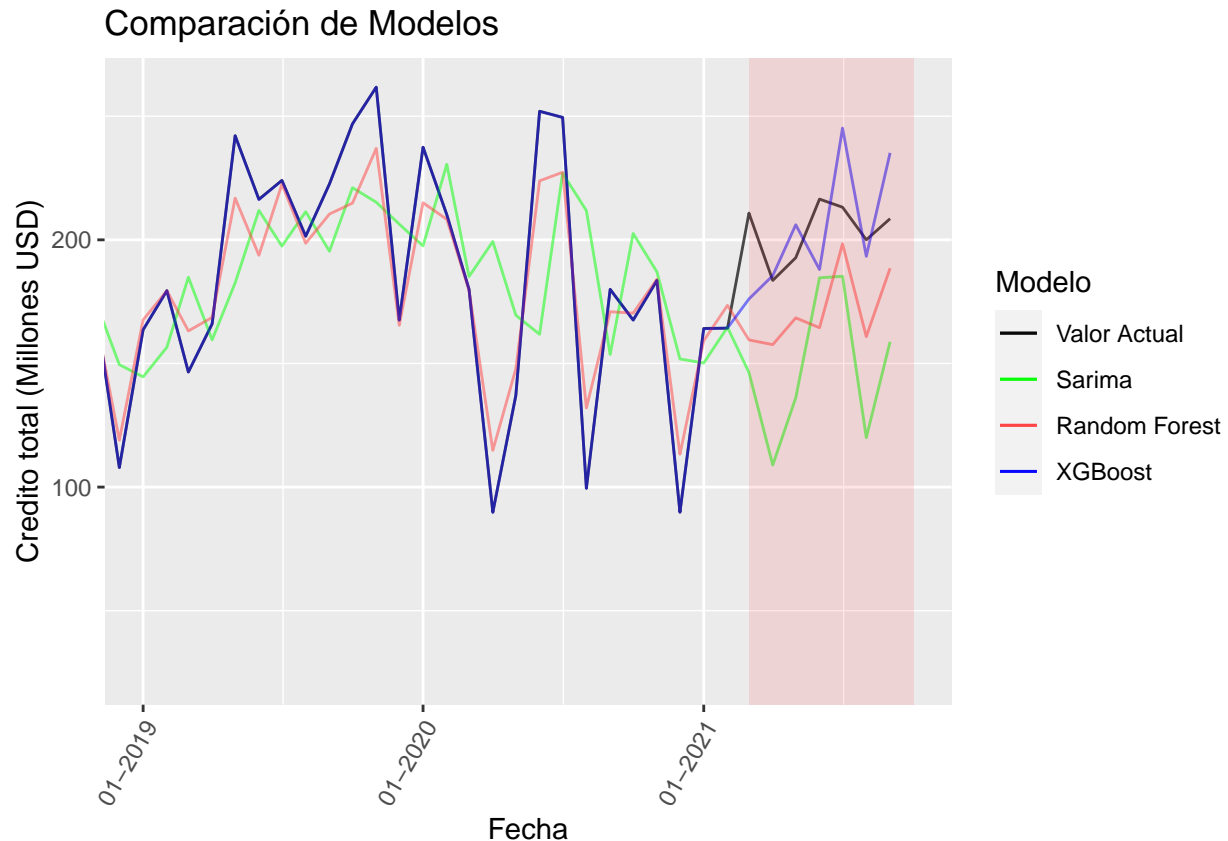
**Ajuste del modelo** En el siguiente grafico podemos observar el ajuste del modelo, tanto en el conjunto de entrenamiento como de prueba.

MODELO	MAPE	MSE	RMSE
SARIMA	93.7123	1.4291	1.1954
RANDOM FOREST	51.5909	0.5299	0.7279
XGBOOST	29.9703	0.2405	0.4904



### Comparación de modelos de acuerdo a MAPE, MSE y RMSE en el conjunto de evaluación

Es aquí donde finalmente se compara el poder de predicción de los modelos en el conjunto de evaluación y se identifica al mejor modelo



De acuerdo a la evidencia se puede concluir que en el análisis preeliminar el modelo que ofrece un mejor pronóstico del crédito quirografario es el modelo XGBoost, teniendo un ajuste excelente en con el conjunto de entrenamiento y superior a los otros dos modelos en el conjunto de prueba, de acuerdo a las los indicadores de calidad de pronóstico MAPE, MSE, RMSE.

### Conclusión:

En el presente trabajo se ha seleccionado de manera detallada las fuentes de datos y las variables relevantes para la predicción del monto de colocación del préstamo quirografario. Posteriormente se analizó los datos y se usaron en una comparación experimental de varios métodos para pronosticar la variable objetivo con y sin variables predictoras. La importancia que los datos abiertos de las búsquedas de google han tenido en el pronóstico puede explicar el mejor desempeño que los métodos de Machine Learning han tenido sobre modelos clásicos autoregresivos, sobre todo el modelo XGBoost que es conocido por ser el modelo preferido para pronósticos en varias plataformas como Kaggle.

### Bibliografía:

- Hassani, H., & Silva, E. S. (2015). Forecasting with Big Data: A Review. In *Annals of Data Science* (Vol. 2, Issue 1, pp. 5–19). Springer Science and Business Media Deutschland GmbH. <https://doi.org/10.1007/s40745-015-0029-9>
- Varian, H. R. (2014). Big data: New tricks for econometrics. *Journal of Economic Perspectives*, 28(2), 3–28. <https://doi.org/10.1257/jep.28.2.3>
- SCHOOLEY, D. K., & WORDEN, D. D. (2010). Fueling the Credit Crisis: Who Uses Consumer Credit and What Drives Debt Burden? *Business Economics*, 45(4), 266–276. <http://www.jstor.org/stable/23491789>

- Gattin-turkalj, K., Ljubaj, I., Martinis, A., & Mrkalj, M. (2007). Estimating Credit Demand in Croatia Draft version. Croatian National Bank Paper, (April), 1–36.
- Asiamah TA, Steel WF, Ackah C. Determinants of credit demand and credit constraints among households in Ghana. *Heliyon*. 2021 Oct;7(10):e08162. DOI: 10.1016/j.heliyon.2021.e08162. PMID: 34765759; PMCID: PMC8569385.
- Stavinova, E., Timoshina, A., & Chunaev, P. (2021). Forecasting the volume of mortgage loans with open Internet data in the period of noticeable changes in the Russian mortgage market. *Procedia Computer Science*, 193, 266–275. <https://doi.org/https://doi.org/10.1016/j.procs.2021.10.027>
- Reynaga, N. C. (2018). La demanda de crédito de las personas: el RCC conoce a la ENAHO. *Monetaria*, VI(1), 107–145.
- Maldonado, L., & Vera, L. (2011). Los determinantes de la demanda de crédito de los hogares: un modelo de vectores de corrección de errores para Venezuela. *Nueva Economía*, 19(November), 13–45. Retrieved from <https://www.researchgate.net/publication/317786081>
- Haselbeck, F., Killinger, J., Menrad, K., Hannus, T., & Grimm, D. G. (2022). Machine Learning Outperforms Classical Forecasting on Horticultural Sales Predictions. *Machine Learning with Applications*, 7, 100239. <https://doi.org/https://doi.org/10.1016/j.mlwa.2021.100239>
- Cunha Medeiros, Marcelo and Vasconcelos, Gabriel and Veiga, Alvaro and Zilberman, Eduardo, Forecasting Inflation in a Data-Rich Environment: The Benefits of Machine Learning Methods (April 30, 2019). Available at SSRN: <https://ssrn.com/abstract=3155480> or <http://dx.doi.org/10.2139/ssrn.3155480>