

# **MODELO DE NOWCASTING PARA PRONOSTICAR LA TASA DE DESEMPLEO DE COLOMBIA UTILIZANDO GOOGLE TRENDS**

**Modalidad: Exploratorio**

**LINA MARÍA TRESPALACIOS CÁRDENAS**

**Trabajo de grado para optar al título de  
Ingeniera Financiera**

**Andrés Felipe García Suaza**

**Doctor en economía**



**UNIVERSIDAD EIA**

**INGENIERÍA FINANCIERA  
ENVIGADO  
2021**

## **AGRADECIMIENTOS**

Quiero agradecerle primeramente a Dios por permitirme llevar a cabo el desarrollo del presente trabajo de grado, por cuidar siempre de mí y de mi familia, por bendecirme y porque a pesar de las calamidades que viví durante el presente año, me lleno de fortalezas y fuerzas para continuar con el desarrollo de este.

Quiero agradecerles también a mis padres por el gran apoyo que me dieron, a mi amigo y docente Alejandro Olaya por todo el acompañamiento que me brindó durante todo el desarrollo de mi trabajo, por sus grandes aportes y por su incondicional amistad.

## CONTENIDO

INTRODUCCIÓN .....	10
1. PRELIMINARES .....	11
1.1 Planteamiento del problema .....	11
1.2 Objetivos del proyecto .....	11
1.2.1 Objetivo General.....	11
1.2.2 Objetivos Específicos .....	12
1.3 Marco de referencia.....	12
1.3.1 El Nowcasting como herramienta de pronóstico .....	16
1.3.2 El desempleo .....	18
2. METODOLOGÍA.....	20
3. PRESENTACIÓN Y DISCUSIÓN DE RESULTADOS.....	27
4. CONCLUSIONES Y CONSIDERACIONES FINALES.....	42
5. REFERENCIAS .....	43

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

## LISTA DE TABLAS

Tabla 1. Palabras Claves Para El Pronóstico De La Tasa De Desempleo De Ee. Uu .....	23
Tabla 2. Lista De Palabras Claves .....	24
Tabla 3. Palabras Claves Y Su Horizonte Temporal .....	25
Tabla 4. Posibles Escenarios.....	26
Tabla 5. Escenarios Para El Modelo De Nowcasting .....	26
Tabla 6. Escenario 1 .....	28
Tabla 7. Escenario 2.....	28
Tabla 8. Escenario 3.....	29
Tabla 9. Escenario 4.....	30
Tabla 10. Escenario 5.....	30
Tabla 11. Escenario 6.....	31
Tabla 12. Criterios Estadísticos Por Escenarios.....	32
Tabla 13. Pendientes.....	34
Tabla 14. Predicción Escenario 1.....	35
Tabla 15. Predicción Escenario 2.....	36
Tabla 16. Predicción Escenario 3, Con Aleatoriedad En Computrabajo .....	38
Tabla 17. Predicción Escenario 3, Con Aleatoriedad En Linkedin .....	39
Tabla 18. Predicción Escenario 3, Con Aleatoriedad En Agencia De Empleo .....	40

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

## LISTA DE FIGURAS

Figura 1. Tasa De Desempleo Histórica.....	22
Figura 2. Tasa De Desempleo Vs Modelo De Regresión Lineal.....	33
Figura 3. Tasa De Desempleo Vs Predicción Escenario 1 .....	36
Figura 4. Tasa De Desempleo Vs Predicción Escenario 2 .....	37
Figura 5. Tasa De Desempleo Vs Predicciones Escenario 3 (Computrabajo Variable) ....	39
Figura 6. Tasa De Desempleo Vs Predicciones Escenario 3 (Linkedin Variable) .....	40
Figura 7. Tasa De Desempleo Vs Predicciones Escenario 3 (Agencia De Empleo Variable) .....	41

## LISTA DE ANEXOS

Figura 1. Empleo .....	46
Figura 2. Ofertas Laborales .....	46
Figura 3. Vacantes.....	46
Figura 4. Hoja De Vida.....	46
Figura 5. Teletrabajo.....	47
Figura 6. Trabajo Sin Experiencia .....	47
Figura 7. Trabajando Colombia.....	47
Figura 8. Agencia De Empleo .....	47
Figura 9. Empleo Sena .....	47
Figura 10. Magneto Empleo.....	48
Figura 11. Linkedin .....	48
Figura 12. Computrabajo .....	48
Figura 13. Elemplo.....	48
Figura 14. Trabaja Con Nosotros .....	48
Figura 15. Trabajos.Com .....	49
Figura 16. Servicio Público De Empleo .....	49
Figura 17. Indeed.....	49
Figura 18. Ministerio De Trabajo.....	49
Figura 19. Trabajo Ya .....	50
Figura 20. Buscador De Empleo .....	50
Figura 21. Código. Librerías Y Palabras Claves.....	50
Figura 22. Código. Librerías Y Palabras Claves.....	51

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

Figura 23. Código. Escenario 1.....	51
Figura 24. Código. Escenario 2.....	52
Figura 25. Código. Escenario 3.....	52
Figura 26. Código. Escenario 4.....	52
Figura 27. Código. Escenario 5.....	53
Figura 28. Código. Escenario 6.....	53
Figura 29. Código. Escenario 6.....	54

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

## RESUMEN

Debido a los niveles de incertidumbre que se crean en la economía, por la espera de los resultados de los indicadores de las variables macroeconómicas, se propone un modelo para pronosticar de manera semanal la tasa de desempleo en Colombia. La espera de los resultados de las variables que explican los movimientos del mercado y ciclos económicos puede perjudicar la toma de decisiones que se lleven a cabo en las entidades públicas; debido al retraso que tienen las entidades generadoras de estadísticas en publicar sus informes mensualmente. El desarrollo de un indicador que pronostique la tasa de desempleo en un “tiempo real” permite dar un panorama de cómo se puede estar comportando esta variable en el tiempo y así facilitar la toma de decisiones políticas, más que todo en tiempos donde el mercado no se encuentre en una estabilidad.

Para el desarrollo del indicador de Now-cast se dará uso al indicador de Google Trends, en el cual se buscarán palabras claves que tengan relación con personas que se encuentren buscando empleo en un lugar determinado y en cierto horizonte temporal. Con los datos suministrados por Google Trends se procederá a realizar un modelo de regresión lineal simple en R studio. Una vez se tenga el modelo, se estimarán y se realizarán los pronósticos respectivos, se mide la eficiencia del pronóstico y se comparan con otros escenarios. Con lo anterior se espera lograr una buena predicción de la tasa de desempleo, el cual sea de uso útil para tomar medidas de control en el caso en que se muestre un incremento en la variable.

**Palabras Clave:** *Nowcasting, palabras claves, hits, tasa de desempleo, pronóstico, regresión lineal*



## ABSTRACT

Due to the levels of uncertainty that the economy has because of the waiting of the indicator of the macroeconomic variables, a forecasting model was proposed to forecast the Unemployment rate in Colombia on a weekly basis. The delay of the results of the variables that explain the market movements and the economic cycles can harm the decision-making in government entities due to the delay that the Statistics Generating Entities have in publishing its monthly reports. The development of an indicator that forecast the unemployment rate in a “real time” shows an outlook of how this variable can behave by the time. which helps to political decision-making when the marketplace is not stable.

For the development of the Now-casting indicator, a google trends indicator is used to search for key words that has a link with people who are looking for a job in a certain place and in a certain time horizon. With the data that google trends provides, a simple linear regression model will be performed in R studio. Once the model is created, the respective forecasts are estimated and made, the efficiency of the forecast is measured and compared with other scenarios. It is expected to achieve a good prediction of the unemployment rate, which is useful for taking control measures if an increase in the variable is shown.

**Keywords:** *Nowcasting, key words, hits, unemployment rate, forecast, linear regression*

## INTRODUCCIÓN

Las variables económicas son de alta importancia en el desarrollo y crecimiento de un país, ya que estas nos indican como se van comportando este en términos económicos, crecimiento poblacional, nivel de pobreza, desempleo, entre otras. Debido a esto, es fundamental tener un seguimiento constante de los movimientos de dichas variables. En este caso, la variable desempleo, la cual es un factor clave para observar cómo se está comportado los niveles de pobreza, crecimiento industrial y la economía; es de suma importancia analizar los cambios que esta variable pueda tener de una manera más constante, con el fin de servir como ayuda a la toma de decisiones de los representantes públicos en cuanto al ciclo económico.

Dado lo anterior, el propósito del presente trabajo es presentar un modelo de pronóstico inmediato o, más comúnmente conocido como Now-casting, con el fin de analizar como se puede comportar la tasa de desempleo en un período de tiempo semanal. El modelo se procedió a realizar bajo un modelo de regresión lineal siempre, la cual es una de las metodologías más comunes a la hora de realizar pronósticos. Dicha metodología fue la escogida debido a que se planteo el supuesto de que la serie iba a tener un crecimiento lineal, ya que los datos suministrados para realizar el modelo de predicción son de manera mensual, por lo que se optó por hacer una interpolación con las series y así predecir la tasa para las próximas cuatro semanas de marzo del presente año.

En el presente trabajo, se llevó a cabo una búsqueda de palabras claves que tuvieran relación con la tasa de desempleo, por lo que se utilizaron palabras que las personas posiblemente indaguen en la internet o navegadores a la hora de buscar un empleo; como lo son plataformas de empleo como LinkedIn, computrabajo o palabras claves más relacionadas a la palabra empleo como trabajo, vacantes, empleo, entre otras. Luego se procedió a modelar diferentes escenarios por medio de una regresión lineal en el programa R studio, con el fin de ver que tan significativas eran las palabras para el modelo y cual de los escenarios se ajustaba mejor al comportamiento de la tasa de desempleo.

Finalmente, se dio paso a la predicción de la tasa de desempleo, la cual fue analizada bajo tres diferentes escenarios; interpolación, aleatoriedad y una combinación de los dos; con el fin de analizar cual de estos escenarios se acercaba más al resultado real de la tasa de desempleo para marzo, y también, para indagar sobre como varían los resultados de la predicción cuando las tres palabras claves más significativas del modelo tomaban valores bajos, medios y altos.

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

# 1. PRELIMINARES

## 1.1 Planteamiento del problema

La tasa de desempleo es uno de los principales indicadores para hacer seguimiento al desempeño de una economía, debido a que estas poseen información de gran relevancia para evaluar el comportamiento o situación del ciclo económico. Una de estas fuentes de información se da a través de la relación que posee esta variable con el producto interno bruto (Domenech & Gómez, 2005). Al igual que la correspondencia entre los consumidores sobre las utilidades, debido a que, acorde con la teoría de los nuevos keynesianos, la tasa de desempleo puede determinarse por dicha relación, y por ende puede modificar movimientos en los precios al aumentar la demanda agregada, ya que el salario es el encargado de igualar la oferta y demanda laboral (García & Cruz, 2017). Con lo anterior se quiere decir que, esta variable económica trae consigo varias fuerzas de mercado en el cual se puede evidenciar el comportamiento económico de un lugar en específico.

Una de las principales dificultades que se tiene a la hora de publicar los datos del desempleo, ya sea en el país o de los mercados laborales regionales, es el tiempo en que tarda la entidad generadora de estadísticas oficiales, en el caso de Colombia, el Departamento Administrativo Nacional de Estadística (DANE) en informar los resultados de los principales indicadores de desempleo, ya que esta entidad transmite dicho informe mensualmente. La causalidad de publicar estos informes en dicha periodicidad, y no de una manera más constante es que se crea una gran incertidumbre durante el periodo y retarda la toma de decisiones de la política pública, lo cual es especialmente problemático en tiempos de alta volatilidad macroeconómica. No obstante, es importante entender el fenómeno del desempleo o saber cómo se está moviendo esta variable constantemente, ya que ayuda a cuantificar la magnitud que produce un cambio en el ciclo económico (Trejo, Rivera, & Rios, 2017).

Por esta razón el diseño de metodologías para medir esta variable de manera más frecuente es muy útil para la toma de decisiones en cuanto a los factores que afecta el desempleo, y para tener una idea sobre las dinámicas del corto plazo del mercado de trabajo, ya sea diaria o semanalmente, lo que lleva a la siguiente pregunta: ¿Cómo se pueden apreciar los posibles resultados de la tasa de desempleo en un “tiempo real”?

## 1.2 OBJETIVOS DEL PROYECTO

### 1.2.1 Objetivo General

Construir un modelo de nowcasting para generar estadísticas líderes de desempleo a nivel nacional, para hacer seguimiento a las dinámicas de la tasa de desempleo durante periodos

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

semanales, con el fin de servir de ayuda para la toma de decisiones tanto a nivel nacional como a nivel regional.

### 1.2.2 Objetivos Específicos

- Operacionalizar la recolección de bases de datos para la predicción del desempleo en tiempo real.
- Definir las metodologías apropiadas para el pronóstico del desempleo en tiempo real.
- Validar un modelo de predicción de desempleo en tiempo real para el país.

### 1.3 MARCO DE REFERENCIA

Richardson, P. (2018) en su trabajo *Nowcasting and the Use of Big Data in Short – Term...* Menciona que, gracias a la evolución de los métodos econométricos, se ha permitido mejorar una previsión en las variables macroeconómicas, como el PIB. La manipulación de grandes bases de datos ha proporcionado una mejor calidad en la información para economistas y analistas financieros, los cuales utilizan variables para predecir el producto interno bruto como las búsquedas frecuentes en internet, información de redes electrónicas como Twitter y transacciones económicas. Con la búsqueda de palabras claves en internet, se pueden crear indicadores de series temporales semanales, ya que este recurso permite la búsqueda de información de manera rápida, a la cual se tiene acceso en tiempos frecuentes teniendo en cuenta palabras claves y la geografía donde se está ubicado; se ha encontrado en varios estudios que la utilización de indicadores de Google Trends (palabras claves) para pronosticar variables macroeconómicas ha sido estadísticamente significativo a comparación de otros elementos que utilizan las encuestas tanto empresariales como al consumidor. Encuentran resultados positivos para indicadores de consumo de EE. UU, para las ventas de electrodomésticos y ventas de hogares, para predecir la refinanciación hipotecaria en Estados Unidos, para precios de casas en el Reino Unido y también para el turismo en Hong Kong (*Richardson, 2018*).

En los mercados financieros también ha sido útil esta búsqueda, ya que investigadores como Andrade et al (2009) utiliza dichas métricas para pronosticar burbujas de volatilidad en las empresas como fuertes correlaciones entre datos de volumen por nombre de empresa y volumen de operaciones y exceso de devoluciones de acciones en el índice s&p500; sin embargo, es una métrica que no es muy sólida para utilizarla en los mercados financieros, ya que tienen bajo poder predictivo. Por otra parte, el uso de medios sociales como Twitter, permite la información continua de los tweets de los usuarios y un tamaño de muestra mayor, al igual que accede a una mayor variabilidad de los datos y permiten un enfoque más estratificado, ya que analiza la información de un gran grupo de usuarios. Se encontró que, en los mercados financieros, los indicadores que utilizan dicha herramienta son más factibles a un corto plazo. Un estudio que se realizó por medio de los comentarios

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

de Twitter, donde mencionaban acciones del S&P500, el precio de las acciones empresariales y el volumen de negociaciones; en el cual se utilizó un modelo de regresión lineal autorregresiva para predecir los indicadores bursátiles, encontró que ocho de cada diez sectores industriales y empresariales poseían correlaciones estadísticamente significativas para los volúmenes de negociación, pero no para el precio de las acciones; mientras que con Apple se observa una correlación estadísticamente significativa tanto en el precio de las acciones como en los volúmenes de negociación (*Richardson, 2018*).

Cuando se realiza una búsqueda en Google sobre un tema en específico, ya sean autos, desempleo, productos, etc.; se pueden observar la cantidad de personas que han estado buscando dichos productos o elementos en un cierto lapso de tiempo. Esta técnica ha ayudado a los pronosticadores para recolectar datos y poder estimar lo que se va a predecir, ya que, si en un lapso de tiempo un alto porcentaje de personas han estado buscando empleo en internet, muy probablemente sea porque la tasa de desempleo este aumentando. Paul Smith utiliza la técnica de MIDAS y recolección de datos de Google Trends para llevar a cabo su pronóstico en cuanto al nivel de desempleo en Inglaterra; para su trabajo se recolectaron datos de la siguiente manera: Se utilizó el término raíz en las búsquedas de Google, la palabra desempleo y sus términos relacionados y despido, de esta manera se llevó a cabo la formación un índice de Google para medir el desempleo; y se centraron en dos partes, beneficios disponibles para desempleados y vacantes en un campo en particular de las empresas de construcción (*Smith, 2016*).

Se utilizaron también datos de encuestas de empresas y consumidores para el pronóstico, las cuales responden a preguntas como las vacantes en empresas, demanda de empleados y disponibilidad de mano de obra. Se creó una regresión lineal entre el desempleo como variable dependiente a tres meses, los datos de Google y las encuestas, de lo cual se encontró que, de todas las variables explicativas, una gran mayoría de estas eran significativas, es decir, servían para explicar el modelo. Una de estas variables fue el índice compuesto de redundancia de Google (GRI), la cual hace parte de los indicadores de este y se define como la suma de volúmenes de término de búsqueda individuales ponderados para cada semana durante el período de muestra. La variable del término raíz también funcionó bien, por el contrario, una de las pocas variables que resultaron no significativas fue el JSA Google. Por otra parte, el modelo MIDAS permite predecir el desempleo, teniendo en cuenta que los datos relativos de esa variable son publicados mensualmente, pero que pueden ser pronosticados por variables que sus datos son informados diaria o semanalmente. De este modelo se observó que los datos que se utilizan para los indicadores de Google son más eficientes para explicar cambios en la variable del desempleo que las encuestas que se realizan a las empresas y consumidores; adicionalmente el modelo a medida de que los datos se fueron actualizando y/o acumulando en el indicador de Google, se fue disminuyendo la incertidumbre; mientras que al inicio del período de previsión el modelo no fue tan óptimo, debido a que no se acumulaban datos frecuentemente, lo cual afectaba la predicción (*Smith, 2016*).

Generar y evaluar predicciones de variables financieras y macroeconómicas en el mercado de Estados Unidos también fue un gran estudio que se llevó a cabo para evaluar la metodología empleada por Edward S. Knotek II y Saeed Zaman, los cuales utilizaron

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

métodos predictivos como modelos financieros de predicción inmediata basadas en datos de alta frecuencia como una caminata aleatoria, predicciones condicionales con pronósticos financieros y macroeconómicos, previsiones condicionales con pronósticos financieros, modelo de muestreo de datos mixtos (MIDAS) y métodos de pronóstico de frecuencia mixta; con el fin de estimar el comportamiento de variables como el índice de acciones S&P500, el rendimiento de los bonos del tesoro a diez años, el rendimiento de los bonos del tesoro a 3 meses, rendimiento de los bonos corporativos, el valor ponderado del dólar frente a otras moneas y variables macroeconómicas como lo es el PIB, la tasa de desempleo, IPC y tasa efectiva de fondos federales (Zaman, 2019).

Durante el estudio se pudo observar que las variables pronosticadas mostraron mejores resultados con los pronósticos condicionales a comparación de los pronósticos incondicionales; no obstante, la predicción condicional con datos financieros mostró peores resultados en su pronóstico en un corto plazo, sin embargo, los datos mensuales que se utilizan para su estimación ayudan a mejorar el pronóstico del PIB en un mediano plazo. Adicionalmente los resultados obtenidos al realizar predicciones condicionales con pronósticos financieros y macroeconómicos fueron más eficientes a la hora de pronosticar la tasa de desempleo y la tasa de fondos federales. Finalmente se compararon pronósticos realizados con modelos BVAR aumentados con predicciones financieras y pronósticos realizados con modelos de frecuencia mixta, los cuales contienen datos de frecuencias altas y bajas; se demostró que los métodos son comparables entre sí, es decir, el método VAR con modelos bayesianos es tan eficiente para el pronóstico de dichas variables económicas como el método de frecuencia mixta (Zaman, 2019).

En el desarrollo del “Labor Market Indicator for Colombia (LMI)”, se dio paso a construir un indicador del mercado laboral en Colombia que permite a las políticas de mercado evaluar la situación actual de este. Posicionando como variables principales los elementos que influyen en el mercado laboral, por medio de un proceso de modelo autorregresivo, con el fin de eliminar posibles shocks idiosincráticos del indicador, es decir, con el fin de eliminar elementos que causan ruido en el LMI. Por medio de un Dynamic Factor Model, el cual es un proceso de estimación de variables, analizaron cada uno de los componentes que hacen parte del ciclo del mercado laboral, como los hogares, tasa de desempleo, tasa de participación laboral, tasa de ocupación, salario promedio, entre otras y estimaron un vector el cual seguía un proceso autorregresivo (Cristiano Botia, Hernandez Bejarano, & Ramos Veloza, 2019).

Para el tratamiento de la serie, extrajeron los componentes cíclicos de cada serie con el fin de eliminar falsas señales y obtener estimaciones robustas para ajustar el modelo. Dichas señales fueron componentes estacionales y permanentes; luego dieron paso a testear el modelo por medio de una prueba de estacionalidad, ajustaron estacionalmente quince series faltantes y pronosticaron cuatro años de periodos no observados, para lo cual utilizaron un modelo ARIMA, para así identificar la cronología de los puntos de inflexión; lo cual ayuda a identificar las fases de altos y bajos crecimientos. Finalmente compararon el LMI con el PIB y la tasa de desempleo (Cristiano Botia, Hernandez Bejarano, & Ramos Veloza, 2019).

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

Los hallazgos del estudio anterior mostraron que su indicador retrata el estado actual del mercado laboral con más información que el único análisis de la tasa de desempleo. Descubrieron que las expectativas, las vacantes, el empleo industrial y la presencia de cuellos de botella también juegan un papel importante durante las fases de crecimiento bajo y alto. Además, su estudio obtuvo una evolución similar a la del PIB y las brechas de la tasa de desempleo. Finalmente, encontraron que la introducción de nueva información no afecta el diagnóstico realizado. Dado que la mayoría de las series son publicadas y no revisadas, y las estimaciones de los parámetros del modelo estadístico no cambian entre estimaciones (Cristiano Botia, Hernandez Bejarano, & Ramos Veloza, 2019).

El desarrollo del modelo de nowcasting para pronosticar la tasa de desempleo de Estados Unidos, elaborado por Shintaro Nagao, Fumiko Takeda y Riku Tanaka, examinó si los datos de intensidad de búsqueda obtenidos de Google Trends contribuyen al pronóstico inmediato de la tasa de desempleo de Estados Unidos, comparándolo así con un modelo ARIMA tradicional. En donde se analizó dos elementos principales, el cambio en la especificación de tendencias de Google y el posible cambio de valor de punto final de las series ajustadas estacionalmente (Nagao, Takeda, & Tanaka, Nowcasting of the U.S. unemployment rate using Google Trends, 2019).

Nagao, Takeda y Tanaka para obtener los datos de la tasa de desempleo, descargaron las series de ALFRED, el cual es una página web del Banco de la Reserva Federal de St. Louis, el cual permite el acceso a los indicadores económicos publicados en el momento de cada anuncio oficial incluyendo datos antes de la revisión. Por otra parte, los datos de intensidad de búsqueda los obtuvieron de Google Trends, donde definieron como palabras claves los términos “trabajos” y “ofertas de trabajo”. Sus palabras claves las definieron con base a estudios previos de Askitas y Zimmermann, 2009. Luego realizaron el respectivo ajuste estacional utilizando un método llamado X13-ARIMA-SEATS, el cual es un programa desarrollado por la oficina del Censo de EE. UU y banco de España para realizar análisis estadísticos (Nagao, Takeda, & Tanaka, Nowcasting of the U.S. unemployment rate using Google Trends, 2019).

Para su proceso de estacionalidad, se utilizaron dos diferentes casos, el primero en el que toda la serie de intensidad de búsqueda para el periodo de la muestra se ajusta estacionalmente y el segundo caso cuando la intensidad de la búsqueda hasta un mes determinado se usa para pronosticar después de que se ajusta estacionalmente cada vez que se realiza una previsión mensual. Para el pronóstico utilizaron un modelo AR, el cual fue aumentado con la intensidad de búsqueda de Google Trends para una ventana de tiempo de 5 años (Nagao, Takeda, & Tanaka, Nowcasting of the U.S. unemployment rate using Google Trends, 2019).

Encontraron que el pronóstico de la tasa de desempleo utilizando Google Trends es más eficiente que el modelo AR en un corto plazo, sin embargo, a un largo plazo no es tan adecuado. Además, concluyeron que el indicador de Google podría ser más preciso en sus resultados al utilizar más palabras claves para su pronóstico (Nagao, Takeda, & Tanaka, Nowcasting of the U.S. unemployment rate using Google Trends, 2019).

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

### 1.3.1 El Nowcasting como herramienta de pronóstico

El Now-casting se define como la predicción del presente y del futuro muy cercano utilizando el pasado reciente, el cual puede ser entendido como pronóstico inmediato. El Now casting anteriormente se utilizaba en el área de la meteorología, pero hoy en día su uso más frecuente es para realizar pronósticos con variables económicas, ya que estas variables presentan un retraso al ser publicadas por las entidades encargadas de mostrar sus resultados mensual o trimestralmente, como por ejemplo el PIB, el cual es estimado oficialmente un mes después del final del trimestre de referencia. El principio básico de los modelos del Now-cast es explotar la información que se publica con antelación y, posiblemente, a frecuencias más altas que la variable objetivo de interés, con el fin de obtener una estimación temprana de esta mucho antes de que la información sea publicada por los entes encargados (Bańbura, 2011).

Cuando se quiere pronosticar una variable económica haciendo uso del Now-casting, se debe de estudiar los componentes que causan movimientos en dicha variable, por ejemplo, en el caso del PIB se puede mirar como es el comportamiento del gasto, consumo de personas, producción, entre otras variables que lo afecten. Un componente clave en esta metodología es usar un modelo con una representación del espacio estado, el cual se puede escribir con dos tipos de ecuaciones: Ecuaciones de medición, las cuales vinculan las series observadas a un proceso en estado latente, es decir, el estado que no es observado de la economía y las ecuaciones de transición, las cuales describen la dinámica del proceso en estado. Cabe resaltar que la representación del espacio del estado permite hacer frente fácilmente a las características de un conjunto de información que se está elaborando actualmente, es decir, cuando al final de la muestra faltan datos por recopilar debido a la falta de sincronización de liberación de datos (Bańbura, 2011).

Para el uso de herramientas predictivas como el Now-casting, se debe de tener acceso a grandes fuentes de información, los cuales permiten realizar el pronóstico de una manera más efectiva, como por ejemplo el Big data. Una posibilidad para obtener una clasificación en general de la información es dividirla en cuatro partes: Volumen o escala de datos, velocidad o análisis de la transmisión de datos, variedad (diferentes formas de datos) y la veracidad la cual se refiere a la incertidumbre de los datos; sin embargo, dicha clasificación resulta ser muy amplia para adoptar modelos de Now-casting. Otra métrica que se utiliza para clasificar la información es identificando tres principales tipos de datos: Altos, los cuales son pocas variables y muchas observaciones, como por ejemplo las consultas de búsqueda y datos sobre operaciones financieras. Gordo, muchas variables y pocas observaciones, como por ejemplo bases de datos transversales. Y enorme, muchas variables y muchas observaciones, como por ejemplo Google Trends. Es el tipo de datos más utilizado en la actualidad (Kapetanios & Papailias, 2018).

Una tercera posibilidad para clasificar información de gran tamaño es identificando el contenido de los datos, como, por ejemplo: Datos en redes sociales, los cuales son informaciones registradas por experiencias humanas, como Facebook, Twitter, Tumblr, Blogs y sus respectivos comentarios, fotos, videos en YouTube, búsquedas en internet, mensajes de texto, mapas generados por usuarios y correo electrónico. Sistemas

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.



tradicionales de negocios, los cuales registran y supervisan eventos de interés, como por ejemplo los registros de un cliente, fabricación de un producto, toma de pedidos, entre otros. Un último elemento que ayuda a filtrar información requerida para el uso de pronósticos es el internet de los objetos o datos generados por maquinas, los cuales son derivados de sensores y maquinas utilizadas para medir los eventos y situaciones que se presenten en el entorno físico; un ejemplo muy claro de esto podría ser los sensores fijos como de climatización, contaminación tráfico o los sensores móviles como la localización de teléfonos móviles, imágenes de satélite y automóviles; no obstante desde el punto de vista económico, son potencialmente relevantes los datos de sistemas informáticos como transacciones de tarjeta de crédito, registros web como Tweets y búsquedas de palabras claves en internet (Kapetanios & Papailias, 2018).

Los grandes datos ofrecen muchos beneficios para la modelización estadística, sin embargo, en varios casos se podrían presentar situaciones donde los resultados presenten sesgos debido a un exceso de falsos positivos, por ejemplo, a confundir correlación con causalidad, ignorar sesgos de muestreo, datos de mala calidad y selección por métodos inapropiados. Debido a lo anterior existen métodos de Now-cast basados en regresiones, ya sea Bridge, modelos de regresión de muestreo de datos mixtos (MIDAS), UMIDAS o en el filtro de Kalman, el cual se aplica para el manejo de frecuencias mixtas. Los modelos de regresión de muestreo con datos mixtos (MIDAS) proporcionan una herramienta muy útil para la manipulación de datos de series de tiempo, de los cuales se realizan muestreo en diferentes frecuencias y son estimados a través del método de mínimos cuadrados no lineales. Las regresiones MIDAS involucran especificaciones parsimoniosas basadas en polinomios de retardo distribuidos, los cuales proporcionan un pronóstico directo de una variable condicional de baja frecuencia de datos y alta frecuencia de rezagos (Ghysels & Qian, 2018).

Una excepción importante en los modelos MIDAS es el modelo de regresión U-MIDAS, el cual significa modelo de regresión de muestreo con datos mixtos sin restricciones. Dicho modelo no utiliza polinomios de retardo distribuidos y es más atractivo en situaciones donde la diferencia en las frecuencias de muestreo es más pequeña. Estos modelos implican una mezcla de datos trimestrales y mensuales con tres observaciones de alta frecuencia que estén asociados con cada punto de datos de baja frecuencia, es decir, el modelo de regresión U-MIDAS a comparación del modelo MIDAS, es que en el modelo sin restricciones solo se involucran los métodos de regresión lineal y no lineal (OLS) (Ghysels & Qian, 2018). Por otra parte, el Now-casting utiliza el filtro de Kaplan para manejar regresiones como el MIDAS, este filtro hace referencia a un campo aleatorio espaciotemporal, el cual se utiliza para representar fenómenos que cambian con el tiempo, pues este modelo incorpora observaciones de un fenómeno temporal, por lo tanto, cuando se va a realizar el campo aleatorio, dichas informaciones deben de ser compatibles con las observaciones en diferentes lapsos de tiempo. El filtro de Kalman se ha utilizado más que todo en las ciencias físicas para representar variedades de fenómenos y en la ingeniería geotécnica en procesos transitorios de flujo de agua (Felipe & Gabriel, 2015).

Otra métrica muy reconocida que es implementada por los analistas a la hora de generar pronósticos inmediatos (Now-cast), es el modelo de vector autorregresivo (VAR), el cual se

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

utiliza cuando se quiere caracterizar las interacciones simultaneas entre un grupo de variables por medio de ecuaciones de forma reducida sin restringir, es decir, que los valores contemporáneos de las variables de un modelo no aparecen como variables explicativas de cada ecuación. El hecho de que sean ecuaciones sin restringir significa que aparece en cada una de ellas el mismo grupo de variables explicativas. En las ecuaciones se pueden incluir variables explicativas de naturaleza determinística, como por ejemplo variables ficticias estacionales, posible tendencia temporal, entre otras. El modelo VAR es muy útil cuando existe evidencia de simultaneidad entre un grupo de variables, es decir, cuando las variables presentan cierta relación entre sí actuando del mismo modo en un cierto marco de referencia temporal, y cuando sus relaciones se transmiten a lo largo de un determinado número de períodos. Más, sin embargo, cabe resaltar que la principal motivación de utilizar los modelos de vector autorregresivo es la dificultad en identificar variables exógenas al modelo, como es preciso hacer la identificación de un modelo de ecuaciones simultáneas (Novales, 2017).

### **1.3.2 El desempleo**

El desempleo como variable económica posee varios factores que hacen que esta tasa sea vea afectada a través del tiempo. Uno de estos factores claves es el nivel de pobreza de un país o región, ya que cuando los niveles de personas vulnerables económicamente aumentan, hay una gran posibilidad de que el nivel de desempleo se incremente (Muñoz, 2019). Lo anterior debido a que la escasez de empleo trae consigo la imposibilidad de sustentar los costos y gastos que se generan en la vida diaria. Por ende, cuando la tasa de desempleo es alta, los niveles de pobreza suelen incrementar y viceversa (Muñoz, 2019).

Acorde con el DPN, Colombia posee dos índices los cuales miden el nivel de pobreza de la población, estos dos indicadores son conocidos como índice de pobreza monetaria e índice de pobreza multidimensional. Donde el índice de pobreza monetaria se refiere al porcentaje de población que devenga ingresos menores que el mínimo establecido por la ley para sustentar sus gastos. Y el índice multidimensional, el cual se encarga de medir las carencias básicas de bienestar en los hogares (Planeación, 2017).

Así mismo como lo ilustra Nicolas Alfredo Farfán en su investigación acerca de la relación entre el desempleo y la pobreza en Colombia, los niveles de pobreza incrementan en un país debido a factores como: Territorios geográficos afectados por los grupos al margen de la ley, desplazamiento forzoso, escasa ayuda por parte del gobierno nacional, poca experiencia laboral, analfabetismo y escases de recursos económicos para acceder a la educación media o superior.

Un segundo factor por el cual el nivel de desempleo se ve afectado, es la discriminación que se genera en contra del género femenino a la hora de ofertar ciertos empleos en áreas como la industria, construcción, ventas y gerencia (Muñoz, 2019). Lo anterior es causado por la preferencia que puede desarrollar ciertas compañías por el género masculino para desempeñarse laboralmente en estos campos. Esto causa un aumento en el desempleo en mujeres y, por ende, un incremento en la tasa de desempleo total (Muñoz, 2019).

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

Otro factor que puede incidir en el aumento del desempleo es la lenta velocidad a la que los salarios se ajustan a los precios económicos, lo cual no permite tener el mercado en equilibrio a un corto plazo. Esto se da cuando las empresas mantienen salarios a la baja (Pachón, 2012).

Las recesiones económicas son factores de alta importancia y que poseen un gran impacto en la tasa de desempleo. Así como lo menciona Fernando Rocha, la gran Recesión, la cual ha sido uno de los eventos que más ha afectado a las economías mundiales, tuvo un fuerte impacto sobre el nivel de desempleo en los países de la Unión Europea, lo cual trajo consigo fuertes consecuencias en los mercados de trabajo y condiciones de vida (Rocha, 2012). Pues esta gran recesión económica destruyó más de 7.2 millones de puestos de trabajo entre los años 2008 y 2012, donde España representaba el 23% de las personas desempleadas en la Unión Europea. Todo este suceso tuvo fuertes impactos en la vida de los españoles y en su economía como tal (Rocha, 2012).

Igualmente ocurrido con la crisis del 2001 en Argentina, la cual fue una crisis económica, social y política, que dejó consigo una alta tasa de desempleo debido a las decisiones políticas tomadas en ese entonces por el presidente Juan Fernando de la Rúa (Merino, 2012). Así mismo ilustrado en el trabajo *“Las crisis económicas y sus efectos en el mercado de trabajo, en la desigualdad y en la pobreza de México”*, por Mario Camberos y Joaquín Bracamontes, donde se exhibe como las crisis económicas ocurridas en México han aumentado de una manera masiva el nivel de los desempleados del país. Esto debido a las decisiones político-monetarias que han tomado los diferentes gobernantes de México, los cuales han afectado el mercado laboral, los salarios y la tasa de desempleo (Castro & Nevárez, 2014).

Es importante entender como la tasa de desempleo se ve afectada por varios factores que inciden en la economía en la vida cotidiana; sin embargo, para el pronóstico inmediato del Nowcasting se hace uso de ciertas palabras claves con el fin de analizar cuantas personas han buscado palabras relacionadas con desempleo o vacantes laborales, para así dar paso al pronóstico y analizar en un tiempo real el número de personas que se encuentran sin empleo o buscando oportunidades laborales. Lo anterior descrito se explicará en la metodología, donde se expondrá el proceso que se tuvo de selección de palabras claves y el modelo econométrico que más se ajusta.

## 2. METODOLOGÍA

En el presente trabajo se llevará a cabo un método de modelación, en el cual se desarrolla un modelo para investigar la realidad económica de Colombia, el cual posee una analogía estructural y funcional con el objeto de investigación. Valle (2007) informa que para iniciar la modelación de un estudio que se esté llevando a cabo, es importante la contemplación de la realidad, ya que una vez este se materialice, se tiene como resultado los factores los cuales sirven para explicar el modelo teniendo en cuenta la realidad que los conforma. Con lo anterior, el autor propone que las acciones a implementar para desarrollar este método son los análisis respectivos al problema planteado, creación del modelo, desarrollo del modelo, obtención de resultados y la comparación entre los resultados obtenidos y el problema planteado inicialmente (Valle, 2007).

Para la recolección de datos que se necesitan al predecir el desempleo a nivel nacional, se hará uso de la herramienta Google Trends, la cual permite buscar datos de manera geográfica y temporal por medio de palabras claves. Dichas palabras ayudarán a cuantificar las personas que se encuentran desempleadas y que están en búsqueda de un nuevo trabajo; estas serán palabras usuales que la gente desempleada suele buscar en internet a la hora de encontrar una vacante. Un claro ejemplo de las palabras claves que se van a utilizar para adquirir información de datos relevantes a predecir la tasa de desempleo, hacen referencia a palabras como: Compu trabajo, servicio público de empleo, empleo.com y otras, la cuales se van a ir incluyendo a medida de la exploración de datos. Se dará uso a una herramienta como Google Trends, ya que la mayoría de los individuos utilizan Google tanto como un motor de búsqueda, como directorio de sitios web de interés, ya que las personas prefieren buscar el nombre de cierto sitio web en Google que dirigirse directamente a la barra de direcciones (Askitas & Zimmerman, 2009).

Para el desarrollo del trabajo se hará uso de una metodología, la cual se va a implementar para la creación del pronóstico de Now-cast para la variable económica. Esta metodología corresponde a una regresión lineal simple, la cual se refiere a un método estadístico de predicción el cual se basa en describir la relación que hay entre una variable dependiente y una o más variables independientes, con el fin de predecir el valor de la variable dependiente, es decir, el análisis de una regresión lineal es una metodología que permite estudiar cómo puede cambiar una variable con la información que le suministran las otras variables explicativas. En una regresión lineal simple se busca que la variable dependiente pueda ser expresada en función de las independientes, con la finalidad de predecir los valores futuros que se quiere con la variable dependiente (Camacho, 2006).

Se debe de tener en cuenta que la relación que hay entre las variables en esta regresión es lineal; de ahí su nombre; por ende, se deben de descartar otro tipo de relaciones (Camacho, 2006). Dicho lo anterior, un modelo de regresión lineal simple se escribe como:

$$Y = \alpha + \beta X + \varepsilon$$

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

Donde  $X$  representa a la variable independiente o explicativa observada la cual quiere ser estudiada para saber qué tanta influencia tiene sobre la variable dependiente, en este caso, las palabras claves del índice de Google Trends. Y hace referencia a la tasa de desempleo, la cual es la variable que se quiere estudiar, y se conoce como la variable dependiente o variable explicada. La variable  $\varepsilon$  representa el componente de los residuales o errores en el modelo, los cuales resultan de mediciones incorrectas de la variable dependiente, variabilidad inherente u omisión de factores externos que influyen en el pronóstico de la variable, y  $\alpha$  hace referencia a la constante del modelo (Cortés, Bielsa, & Cobo, 2015).

La construcción del modelo de regresión lineal se llevará a cabo por cuatro etapas principales:

- Identificación: es donde se analizan las palabras claves que se van a utilizar para modelar la variable desempleo. Es donde se analiza el comportamiento de dichas variables, que tan buscadas han sido y que tan influyentes pueden llegar a ser para el modelo.
- Estimación: se realiza la inferencia del modelo con las palabras claves elegidas.
- Validación: se hacen diagnósticos para comprobar si la serie es válida para pronosticar cierta variable, es decir, para ver si el modelo escogido se ajusta a los datos. En caso de que el modelo no sea válido, en la validación también se encuentra la fase de buscar dichos errores que no dejan que el modelo se ajuste a los datos, con el fin de mejorar dichas discrepancias.
- Predicción: es la etapa donde se realizan los pronósticos de las variables y también donde se puede validar el modelo para saber si predijo adecuadamente o no.

Google Trends permite la recolección de datos en tiempo real, lo que facilita la predicción del desempleo. El proceso ideal que se realiza con el indicador de Google Trends es estimar un buen modelo que se ajuste a la base de datos, incluir el indicador de Google Trends como una variable explicativa al modelo y finalmente evaluar la mejora en las predicciones de la variable macroeconómica a través de métodos comparativos de predicción. (Artola & Galán, 2012). Después de obtener el modelo de regresión lineal adecuado, el cual se obtiene con ayuda del programa R studio, se crearán las estimaciones y predicciones necesarias con el fin de modelar diferentes escenarios y escoger el mejor de estos.

A continuación, se presentan los datos extraídos del Banco de La República, los cuales hacen referencia a las tasas de desempleo que ha tenido Colombia desde el 2001 hasta febrero del presente año.

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

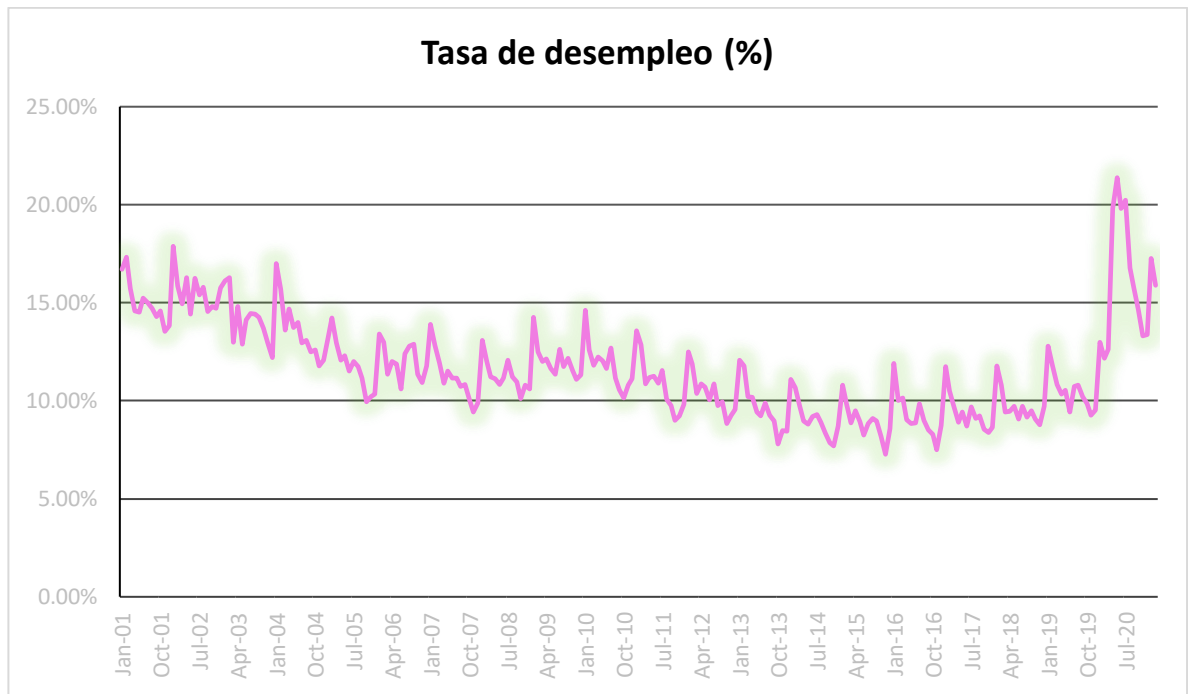


Figura 1. Tasa de desempleo histórica

### Recolección de información

Para la elaboración del presente trabajo se dio inicio a identificar las palabras claves que se utilizarán para el pronóstico de la tasa de desempleo colombiana para las próximas dos semanas. Para esto se recurrió a investigaciones pasadas donde predecían la tasa de desempleo en países como Estados Unidos y otra investigación llevada a cabo por Cristiano-Botia, Hernandez-Bejarano y Ramos-Velozza en 2019, donde se desarrolló un indicador para el mercado laboral en Colombia.

Para predecir la tasa de desempleo de los Estados Unidos, Hyungyoung Choi y Hal Varían, en su trabajo "Predicting the Present with Google Trends", utilizaron una ventana de tiempo de tres años en donde buscaban en todas las subregiones de los Estados Unidos ciertas palabras claves relacionadas con el desempleo, las cuales estaban divididas por tres secciones: "Trabajos", "Bienestar y desempleo" y "Reclutamiento y dotación de personal". Donde cada subdivisión contaba con nueve palabras claves, así como se ilustra en la siguiente tabla.

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

<b>Trabajos</b>	<b>Bienestar y Desempleo</b>	<b>Reclutamiento y Dotación de personal</b>
Monster	Social Security	CareerBuilder
Indeed	Social Security Office Locations	Kelly Service
Jobs	Social Security Administration	Manpower
Job Search	Unemployment Benefits	Temp Agencies
Resume	Social Security Disability	Robert Half
Job Search Engines	Social Security Gov	Spherion
Linkedin	Unemployment Office	Aerotek
Hotjobs	Food Stamps	Walmart Jobs
Cover Letter	Department of Labor	Appleone

Tabla 1. Palabras claves para el pronóstico de la tasa de desempleo de EE. UU

Por otra parte, para el desarrollo del indicador del mercado laboral en Colombia, Cristiano-Botia, Hernandez-Bejarano, & Ramos-Veloza, 2019, en la variable desempleo, dieron uso a palabras claves como “Vacantes de trabajo” y “Sitios de trabajo”. Finalmente, Shintaro Nagao, Fumiko Takeda y Riku Tanaka, en la elaboración de su modelo de regresión lineal para pronosticar la tasa de desempleo de EE. UU utilizando Google Trends, utilizaron dos palabras claves para la búsqueda del desempleo: “Trabajo” y “Ofertas laborales” (Nagao, Takeda, & Tanaka, Nowcasting of the U.S. unemployment rate using Google Trends, 2019).

Teniendo en cuenta las palabras claves utilizadas en las investigaciones anteriormente mencionadas, y considerando pertinente el uso de más palabras claves como lo son las plataformas web más comúnmente usadas por la población colombiana; se procedió a la elección de veinte palabras claves. Estas fueron agrupadas en dos categorías: la categoría de palabras relacionadas con el término empleo y la categoría de plataformas de empleo, como se puede ilustrar en el siguiente cuadro.

<b>Palabras relacionadas con el término empleo</b>	<b>Plataformas de empleo</b>
Empleos	Magneto empleos
Ofertas laborales	LinkedIn
Vacantes	Computrabajo
Hoja de vida	Eempleo
Teletrabajo	Trabaja con nosotros
Trabajo sin experiencia	Trabajos.com
Trabajando Colombia	Servicio público de empleo
Agencia de empleo	Indeed
Empleo SENA	Mintrabajo
	Trabajoya
	Buscadordeempleo

Tabla 2. Lista de palabras claves

Luego se procedió a la búsqueda de estas palabras claves en Google Trends por medio del programa R studio. Dicha búsqueda se filtró por geografía, ya que nos interesa saber cuántas veces han sido buscadas estas palabras en Colombia. Es importante resaltar que Google Trends lanza unos “hits”, los cuales son un rango de números del 0 al 100 e indican que tan buscada ha sido la palabra cada mes en el tiempo; siendo 100 lo más buscado y 0 lo menos buscado.

En las figuras 1-20 se pueden observar las variaciones que han tenido las veinte palabras claves durante los últimos 16 años, Es importante resaltar que no todas las palabras poseen información desde el 2004, por ende, se procedió a separar las palabras por grupos acorde a su período de tiempo, donde ocho palabras poseen información desde el 2004, dos palabras a partir del 2006, una desde el 2009, cuatro desde el 2011, dos desde el 2012 y 2014 y una palabra con información a partir del 2019. Lo anterior se puede entender mejor observando la información en la siguiente tabla.

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.



<b>Palabra</b>	<b>Horizonte de tiempo</b>
Empleo	2004
ofertas laborales	2004
Vacantes	2004
Hoja de vida	2004
teletrabajo	2004
trabajando Colombia	2004
empleo Sena	2004
eempleo	2004
computrabajo	2006
servicio público de empleo	2006
trabajos.com	2009
Ministerio de trabajo	2011
Linkedin	2011
Indeed	2011
Trabajo ya	2011
trabajo sin experiencia	2012
trabaja con nosotros	2012
agencia de empleo	2014
buscador de empleo	2014
magneto empleo	2019

Tabla 3. Palabras claves y su horizonte temporal

Así mismo se procedió a analizar siete posibles escenarios con base a un período de tiempo en específico y que se podrían realizar para el pronóstico de la tasa de desempleo. Así como se ilustra en la tabla.

<b>Posibles escenarios</b>		
<b>Escenarios</b>	<b>Número de palabras</b>	<b>Horizonte de tiempo en años</b>
1	9	16
2	11	14
3	12	11
4	15	9
5	17	8
6	19	6
7	20	1

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

Tabla 4. Posibles escenarios

Con base a lo anterior se procedió a descartar las palabras que tuvieran un horizonte temporal menor a cinco años, ya que este será el tiempo mínimo establecido para proyectar la tasa de desempleo. Por dicha razón, la única palabra con un tiempo menor a cinco años fue magneto empleo, la cual posee información a partir del 2019 y por ende fue la palabra clave descartada. Por ende, la nueva lista de palabras claves con su respectivo escenario y venta de tiempo a estimar quedo así:

<b>Escenarios</b>		
<b>Escenarios</b>	<b>Número de palabras</b>	<b>Horizonte de tiempo en años</b>
1	9	16
2	11	14
3	12	11
4	15	9
5	17	8
6	19	6

Tabla 5. Escenarios para el modelo de nowcasting

### 3. PRESENTACIÓN Y DISCUSIÓN DE RESULTADOS

Con la información suministrada por el Banco de la República (tasa de desempleo) y el buscador en Google Trends, se procedió a realizar el análisis estadístico con cada uno de los seis diferentes escenarios que se plantearon para el pronóstico de la tasa de desempleo. Teniendo en cuenta que cada escenario posee número de datos y ventanas de tiempo distintas, pre procedió a ajustar la tasa de desempleo y las palabras claves a dichos criterios con el fin de obtener un análisis más robusto.

El modelo de regresión lineal que se llevó a cabo para la estimación del modelo de nowcasting sigue la siguiente ecuación general:

$$y = \alpha + \beta x + \varepsilon i$$

Donde “y” hace referencia a la variable que se desea modelar, en este caso la tasa de desempleo. Alfa es el intercepto o constante,  $\beta$  son los coeficientes angulares de la regresión, “x” son cada una de las variables independientes, es decir, cada una de las palabras claves y  $\varepsilon i$  hace referencia a los errores del modelo.

Con la ecuación anterior, se dio inicio a la inferencia de los datos, los cuales lanzan ciertos análisis estadísticos con el fin de analizar la significancia del modelo y que tan significativas pueden ser las palabras claves. Para lo anterior se observan algunos criterios estadísticos como lo son el valor P, R cuadrado ajustado, R cuadrado múltiple y el error residual estándar. Dichos criterios reflejan que tan significativas son las variables en el modelo de regresión, para lo cual se tiene que el valor p y el error residual entre más pequeño sea de 0.05, más significativo es y para los otros elementos como el R cuadrado ajustado y el R cuadrado múltiple, entre más alto sea su valor, mayor nivel de significancia tendrá la variable (Pontaque, 2005).

#### Primer escenario:

MODELO ESCENARIO 1					
Residuals:					
	Min	1Q	Median	3Q	Max
	-0.043077	-0.015249	-0.000813	0.00949	0.083522
Coefficients:					
		Estimate	Std. Error	t value	Pr(> t )
(Intercept)		9.93E-02	8.50E-03	11.689	2.00E-16 ***
empleo1		-2.97E-04	2.08E-04	-1.432	0.15386
ofertaslaborales1		2.87E-04	1.29E-04	2.223	0.02739 *
vacantes1		2.06E-05	1.32E-04	0.157	0.87567
hojadevida1		-9.89E-05	1.87E-04	-0.528	0.59779
teletrabajo1		3.19E-04	1.09E-04	2.919	0.00393 **
trabajandocolombia1		4.32E-04	1.07E-04	4.058	7.15E-05 ***
empleosena1		2.61E-04	1.39E-04	1.88	0.0616 .
eempleo1		-1.98E-04	1.02E-04	-1.952	0.05238 .
ministeriodetrabajo1		6.17E-04	1.39E-04	4.424	1.60E-05 ***
Residual standard error:		0.02047			
Multiple R-squared:		0.2199			
Adjusted R-squared:		0.1841			
F-statistic:		6.138			
p-value:		1.33E-07			

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

Tabla 6. Escenario 1

Acorde a los resultados del escenario número uno, se puede ilustrar que cuatro de las nueve palabras claves escogidas para este escenario fueron altamente significativas, ya que su valor p es menor al 0.05, siendo estas ofertas laborales, teletrabajo, trabajando Colombia y Ministerio de Trabajo. Por otra parte, se puede ver que el modelo es estadísticamente significativo ya que su valor P es del 1.33E-07 y su error estándar es de 0.020.

**Segundo escenario:**

MODELO ESCENARIO 2					
Residuals:					
	Min	1Q	Median	3Q	Max
	-0.044223	-0.009664	-0.001887	0.00833	0.077021
Coefficients:					
	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	9.85E-02	8.87E-03	11.104	2.00E-16	***
es2_empleo	-6.90E-04	2.77E-04	-2.491	0.0137	*
es2_ofertaslaborales	6.32E-05	2.04E-04	0.309	0.7574	
es2_vacantes	4.24E-04	1.67E-04	2.542	0.0119	*
es2_hojadevida	3.76E-04	2.50E-04	1.504	0.1345	
es2_teletrabajo	4.94E-04	1.22E-04	4.063	7.40E-05	***
es2_trabajandocolombia	6.37E-04	1.47E-04	4.32	2.65E-05	***
es2_empleosena	3.55E-04	2.17E-04	1.635	0.1039	
es2_elemplo	1.23E-06	1.17E-04	0.011	0.9916	
es2_ministeriodetrabajo	-2.39E-04	2.84E-04	-0.84	0.402	
es2_computrabajo	-6.94E-04	1.48E-04	-4.677	5.90E-06	***
es2_serviciopublicoempleo	2.56E-04	1.09E-04	2.343	0.0203	*
---					
Residual standard error:	0.01905				
Multiple R-squared:	0.326				
Adjusted R-squared:	0.2824				
F-statistic:	7.4775				
p-value:	2.085E-10				

Tabla 7. Escenario 2

En el escenario número dos, se observa que el modelo de regresión muestra una alta significancia acorde a su valor p y su error, 2.085E-10 y 0.019 respectivamente. Además, seis de sus once variables mostraron ser altamente significativas en el modelo. Se resalta que algunas de las variables que no fueron estadísticamente significativas en el primero escenario, como “empleo” y “vacantes”, se volvieron significativas en el segundo escenario. Y variables que fueron estadísticamente significativas en el primer escenario, dejaron de serlo en el segundo escenario, en este caso “Ministerio de Trabajo”. Lo anterior posiblemente se deba al número de buscas que tuvieron dichas palabras en esta ventana de tiempo.

**Tercer escenario:**

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

MODELO ESCENARIO 3					
Residuals:					
	Min	1Q	Median	3Q	Max
	-0.058032	-0.010343	-0.002208	0.007898	0.07435
Coefficients:					
		Estimate	Std. Error	t value	Pr(> t )
(Intercept)		9.13E-02	1.11E-02	8.24	1.43E-13 ***
es3_empleo		-9.64E-04	3.68E-04	-2.619	0.00983 **
es3_ofertaslaborales		2.88E-04	4.06E-04	0.71	0.47879
es3_vacantes		7.51E-04	2.35E-04	3.193	0.00176 **
es3_hojadevida		2.45E-04	3.34E-04	0.735	0.4635
es3_teletrabajo		7.23E-04	1.39E-04	5.207	7.12E-07 ***
es3_trabajandocolombia		3.23E-04	2.32E-04	1.392	0.16629
es3_empleosena		3.08E-04	2.87E-04	1.073	0.28502
es3_eempleo		-4.17E-04	1.73E-04	-2.415	0.01709 *
es3_ministeriodetrabajo		-5.72E-04	4.13E-04	-1.383	0.16901
es3_computrabajo		1.99E-05	2.63E-04	0.075	0.94004
es3_serviciopublicoempleo		2.23E-04	1.39E-04	1.603	0.11124
es3_trabajos_com		5.39E-05	1.06E-04	0.506	0.61341
---					
Residual standard error:		0.01924			
Multiple R-squared:		0.4273			
Adjusted R-squared:		0.3757			
F-statistic:		8.271			
p-value:		1.69E-11			

Tabla 8. Escenario 3

En el presente escenario se obtuvieron cuatro variables altamente significativas para el modelo de regresión lineal. Al igual que el escenario dos, se observa que algunas palabras claves que fueron estadísticamente significativas en el segundo escenario, dejaron de serlo en el tercero. Sin embargo, el modelo lanzó un valor p de 1.69E-11, lo cual prueba su significancia estadísticamente.

#### Cuarto escenario:

MODELO ESCENARIO 4					
Residuals:					
	Min	1Q	Median	3Q	Max
	-0.047245	-0.0097	0.000133	0.010077	0.0503
Coefficients:					
		Estimate	Std. Error	t value	Pr(> t )
(Intercept)		4.46E-02	1.57E-02	2.836	0.00548 **
es4_empleo		-1.31E-03	4.15E-04	-3.155	0.00209 **
es4_ofertaslaborales		5.59E-04	4.52E-04	1.237	0.21868
es4_vacantes		9.93E-04	3.15E-04	3.149	0.00213 **
es4_hojadevida		9.82E-05	3.50E-04	0.281	0.77944
es4_teletrabajo		6.12E-04	1.32E-04	4.652	9.55E-06 ***
es4_trabajandocolombia		5.59E-04	2.76E-04	2.024	0.0455 *
es4_empleosena		2.07E-04	3.81E-04	0.544	0.58762
es4_eempleo		-2.90E-04	2.72E-04	-1.065	0.28932
es4_ministeriodetrabajo		1.84E-04	3.99E-04	0.462	0.64519
es4_computrabajo		3.76E-04	3.08E-04	1.219	0.22544
es4_serviciopublicoempleo		-8.01E-06	1.59E-04	-0.051	0.95979
es4_trabajos_com		4.02E-05	1.44E-04	0.278	0.78126
es4_linkedin		6.89E-04	2.38E-04	2.894	0.00462 **
es4_indeed		-4.10E-04	1.64E-04	-2.508	0.01367 *
es4_trabajoya		2.29E-04	1.14E-04	2.012	0.04672 *
---					
Residual standard error:		0.01661			
Multiple R-squared:		0.6323			
Adjusted R-squared:		0.5803			
F-statistic:		12.15			
p-value:		2.20E-16			

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

Tabla 9. Escenario 4

Se observa que siete de sus quince palabras claves fueron altamente significativas, al igual que el modelo en general ya que este arrojó un valor p de 2.20E-16 y un error residual de 0.016. Y, al igual que en los escenarios pasados, ciertas variables bajaron su nivel de significancia cuando lo fueron en los escenarios pasados.

**Quinto escenario:**

MODELO ESCENARIO 5					
Residuals:					
	Min	1Q	Median	3Q	Max
	-0.045485	-0.00776	-0.00079	0.00824	0.048866
Coefficients:					
	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	4.37E-02	1.83E-02	2.393	0.01873	*
es5_empleo	-1.47E-03	5.22E-04	-2.821	0.00586	**
es5_ofertaslaborales	7.14E-04	5.01E-04	1.425	0.15762	
es5_vacantes	9.19E-04	4.24E-04	2.169	0.0327	*
es5_hojadevida	7.50E-05	4.05E-04	0.185	0.85371	
es5_teletrabajo	5.85E-04	1.43E-04	4.108	8.63E-05	***
es5_trabajandocolombia	9.45E-04	4.09E-04	2.31	0.02311	*
es5_empleosena	3.73E-04	4.43E-04	0.843	0.4016	
es5_empleo	-1.79E-04	3.29E-04	-0.544	0.58781	
es5_ministeriodetrabajo	2.44E-04	4.23E-04	0.577	0.56536	
es5_computrabajo	1.72E-04	3.99E-04	0.431	0.66714	
es5_serviciopublicoempleo	-4.60E-05	1.76E-04	-0.261	0.79481	
es5_trabajos_com	8.09E-05	2.54E-04	0.318	0.75122	
es5_linkedin	6.98E-04	2.57E-04	2.719	0.00783	**
es5_indeed	-4.20E-04	1.89E-04	-2.214	0.02928	*
es5_trabajoya	2.48E-04	1.33E-04	1.871	0.06452	
es5_trabajosinexperiencia	-6.83E-06	3.28E-04	-0.021	0.98346	
es5_trabajaconnosotros	1.23E-04	2.90E-04	0.424	0.67275	
----					
Residual standard error:	0.01699				
Multiple R-squared:	0.656				
Adjusted R-squared:	0.5924				
F-statistic:	10.32				
p-value:	9.05E-15				

Tabla 10. Escenario 5

En el escenario número cinco, al igual que en los otros anteriores, se muestra una gran significancia del modelo (valor p = 9.05E-15 y error residual = 0.017). Por otra parte, se observa que solo cinco de 17 palabras tuvieron una alta significancia estadística para el modelo de predicción de la tasa de desempleo.

**Sexto escenario:**

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

MODELO ESCENARIO 6					
Residuals:					
	Min	1Q	Median	3Q	Max
	-0.033411	-0.009321	0.000259	0.009493	0.037497
Coefficients:					
		Estimate	Std. Error	t value	Pr(> t )
(Intercept)		3.40E-02	2.34E-02	1.455	0.15054
es6_empleo		-1.56E-03	7.37E-04	-2.116	0.03816 *
es6_ofertaslaborales		5.35E-04	6.42E-04	0.832	0.40816
es6_vacantes		8.90E-04	5.28E-04	1.685	0.09663
es6_hojadevida		2.21E-05	5.74E-04	0.039	0.96937
es6_teletrabajo		5.08E-04	1.90E-04	2.678	0.00933 **
es6_trabajandocolombia		9.57E-04	9.42E-04	1.017	0.31298
es6_empleosena		-1.48E-03	7.06E-04	-2.099	0.03966 *
es6_empleo		-6.83E-04	4.39E-04	-1.555	0.12484
es6_ministeriodetrabajo		4.32E-05	4.57E-04	0.094	0.92507
es6_computrabajo		1.71E-03	7.36E-04	2.324	0.02323 *
es6_serviciopublicoempleo		-9.21E-05	2.07E-04	-0.446	0.65708
es6_trabajos_com		-1.25E-04	4.30E-04	-0.291	0.77218
es6_linkedin		1.02E-03	3.79E-04	2.681	0.00926 **
es6_indeed		-2.38E-04	2.40E-04	-0.993	0.32432
es6_trabajoya		2.60E-04	1.55E-04	1.681	0.09746
es6_trabajosinexperiencia		2.71E-04	3.85E-04	0.703	0.48429
es6_trabajaconnosotros		-4.26E-05	3.26E-04	-0.131	0.89641
es6_agenciadempleo		9.98E-04	3.29E-04	3.028	0.00351 **
es6_buscadoreempleo		-7.16E-04	2.21E-04	-3.245	0.00185 **
---					
Residual standard error:		0.01664			
Multiple R-squared:		0.7506			
Adjusted R-squared:		0.6788			
F-statistic:		10.46			
p-value:		2.47E-13			

Tabla 11. Escenario 6

El sexto escenario, con dieciséis palabras claves, siete de ellas altamente significativas, obtuvo un valor p de 2.47E-13 y un error residual de 0.016; lo que significa que el modelo es altamente significativo. Además, se repite la misma situación que con los escenarios anteriores, donde ciertas palabras pierden su significancia y otras las recuperan.

Teniendo en cuenta que todos los modelos fueron estadísticamente significativos para la predicción de la tasa de desempleo, se debe de escoger el mejor entre ellos, por lo que se procede a analizar los criterios estadísticos de cada modelo, con el fin de saber cuál de los seis modelos es el más adecuado para pronosticar la tasa de desempleo. Así como se ilustra en la siguiente tabla.

	Residual standard error:	Multiple R-squared:	Adjusted R-squared:	F-statistic:	p-value:
escenario 1	0.02047	0.2199	0.1841	6.138	1.33E-07
escenario 2	0.01905	0.326	0.2824	7.4775	2.085E-10
escenario 3	0.01924	0.4273	0.3757	8.271	1.69E-11
escenario 4	0.01661	0.6323	0.5803	12.15	2.20E-16
escenario 5	0.01699	0.656	0.5924	10.32	9.05E-15
escenario 6	0.01664	0.7506	0.6788	10.46	2.47E-13

Tabla 12. Criterios estadísticos por escenarios

Teniendo en cuenta los criterios estadísticos arrojados por el modelo, se observa que el escenario número 6 es el que posee mayor R cuadrado múltiple y mejor R cuadrado ajustado. Sin embargo, el escenario 4 fue el escenario con menor valor p y error residual. Asimismo, se procedió a analizar el número de variables estadísticamente significativa que tuvieron estos dos escenarios, con el fin de escoger al que tuviera más número de palabras claves significativas. Y como resultado se obtuvo que los dos escenarios poseen el mismo número de variables significativas, las cuales son siete.

Dada la situación anterior y teniendo en cuenta el informe publicado por Colombia TIC, el cual informa la penetración de internet que ha tenido Colombia, su banda ancha, internet móvil y demás, para el año 2011 la penetración de internet fue de 11%, 12%, 12.8% y 13.4% para los cuatros trimestres de dicho año, respectivamente. Por otra parte, la penetración de internet que tuvo Colombia en el primer, segundo, tercer y cuarto trimestre del 2016, fue de 28.1%, 29.5%, 31.01% y 32.5% respectivamente. Lo cual quiere decir que las personas tuvieron un mayor acceso a internet, páginas web o exploradores de Google durante el 2016 (TIC, 2017). Por lo cual posiblemente las personas que se encontraban buscando empleo en este año tuvieron más acceso a plataformas digitales o buscaban empleo por la computadora en vez de medios convencionales como periódicos, radio o televisora. Razón por la cual el escenario número 6 podría ajustarse más al pronóstico de la tasa de desempleo, por lo tanto, fue el modelo escogido para la realización del nowcasting.

En la siguiente figura se puede ver como se ajusta el comportamiento del modelo de regresión lineal con el escenario 6 y la tasa de desempleo, desde el año 2016 hasta febrero del 2021



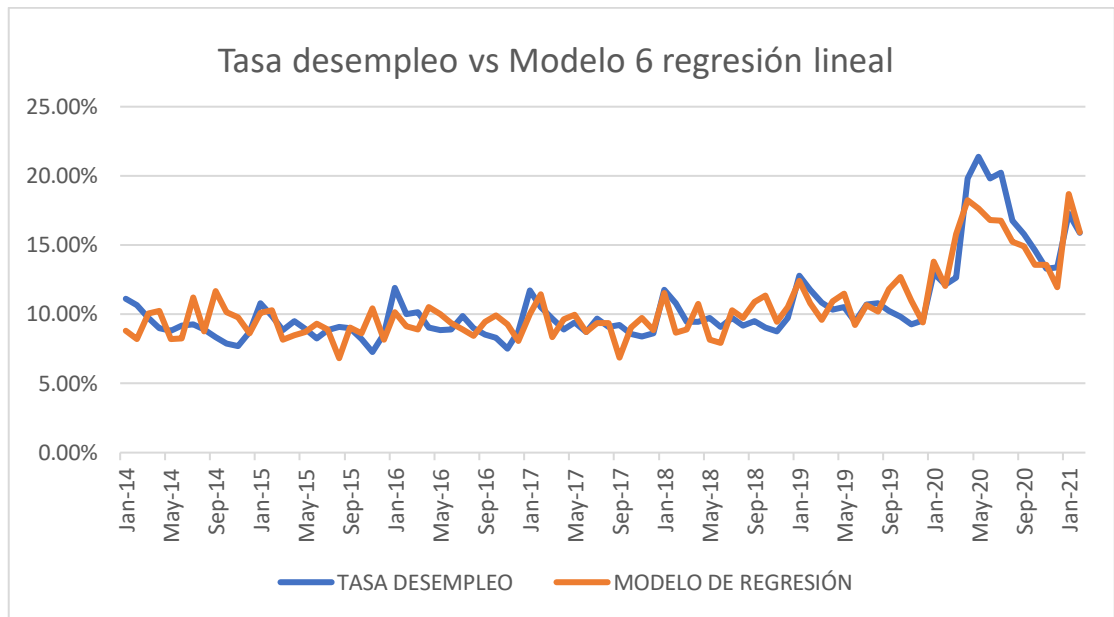


Figura 2. Tasa de desempleo vs modelo de regresión lineal

Acorde con la figura número dos, se puede ver como se asemeja el comportamiento que han tenido las búsquedas de las palabras claves a través del tiempo con la variación de la tasa de desempleo. Lo cual nos confirma que el modelo de regresión lineal elegido es adecuado para dar paso al proceso de predicción por medio de las búsquedas del índice de Google Trends.

### **Predicción de la tasa de desempleo**

Para el pronóstico de la tasa de desempleo, se llevaron a cabo tres diferentes escenarios, en donde se utilizaron distintos métodos para llevar a cabo la predicción de esta tasa de manera semanal.

Teniendo en cuenta que los datos suministrados por el índice de Google Trends son de manera mensual, y que el informe de la tasa de desempleo suministrado por el DANE o por el Banco de la República, al igual que el índice de Google Trends, es de manera mensual, para el primer escenario, se hizo el supuesto de que las variables iban a tener un crecimiento lineal, para lo cual se llevó a cabo una interpolación de los datos que lanza cada una de las palabras de Google Trends, con el fin de obtener los hits de cada palabra clave en las próximas cuatro semanas del mes de marzo y así poder pronosticar la tasa de desempleo de manera semanal.

Para la obtención de los hits de las próximas cuatro semanas, se halló la pendiente para cada variable así:

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

$$M = \frac{Y_2 - Y_1}{X_2 - X_1}$$

Donde:

$Y_2$ : Valor de la variable en el mes de febrero

$Y_1$ : Valor de la variable en el mes de enero

$X_2$ : Mes 2 (febrero)

$X_1$ : Mes 1 (enero)

Para lo cual se obtuvieron las siguientes pendientes:

Pendientes (M)	
Empleo	-10
Ofertas laborales	-11
Vacantes	-19
Hoja de vida	-10
Teletrabajo	-10
Trabajando colombia	3
Empleo sena	-7
Empleo	-2
Ministerio de trabajo	-1
Computrabajo	-6
Servicio publico de empleo	-10
Trabajos.com	-9
Linkedin	-6
Indeed	-2
Trabajo ya	2
Trabajos sin experiencia	-21
Trabaja con nosotros	-10
Agencia de empleo	-14
Buscardor de empleo	-4

Tabla 13. Pendientes

Como se quiere predecir la tasa de desempleo de manera semanal, se procedió a dividir cada una de las pendientes en cuatro partes, (ya que se hizo el supuesto de que esta variable iba a crecer o decrecer de manera lineal), para luego sumársela al dato anterior y así poder obtener los hits de cada una de las variables para las próximas cuatro semanas de marzo. Es importante resaltar que, en los casos donde la interpolación fuese menores a cero o mayores a cien, su resultado en las próximas cuatro semanas iba a ser igual a cero.

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

Lo anterior debido a que el índice de Google Trends lanza los hits de un rango de cero a cien. Lo anterior ocurrió solamente para una variable, la cual fue trabajos.com

Finalmente, cada uno de los resultados obtenidos en la interpolación, se multiplicaron por cada uno de los coeficientes de las variables, dando como resultado el pronóstico para las próximas cuatro semanas de marzo, así como se ilustra en la tabla 14.

<b>Semana 1</b>	<b>15.21%</b>
<b>Semana 2</b>	<b>14.49%</b>
<b>Semana 3</b>	<b>13.77%</b>
<b>Semana 4</b>	<b>13.05%</b>

Tabla 14. Predicción escenario 1

Acorde a los resultados obtenidos de la predicción, se puede observar que, para el primer escenario, la tasa de desempleo va disminuyendo semanalmente hasta llegar a 13.05%, la cual es la tasa que se espera obtener para el mes de marzo. Acorde al informe mensual de la tasa de desempleo, publicado por el Banco de La República, la tasa de desempleo que se obtuvo para el mes de marzo del 2021 fue de 14.20%, lo cual nos indica que el método de predicción que se utilizó para el primer escenario nos lanza un buen resultado para la tasa de desempleo.

Con el fin de obtener una mejor visualización del modelo de regresión junto con el comportamiento de la tasa de desempleo, se procedió a graficar la tasa de desempleo desde enero del 2014 hasta marzo del 2021 junto con el modelo de regresión y su pronóstico para el mes de marzo, el cual fue de 13.05%. Así como se ilustra en la siguiente figura.

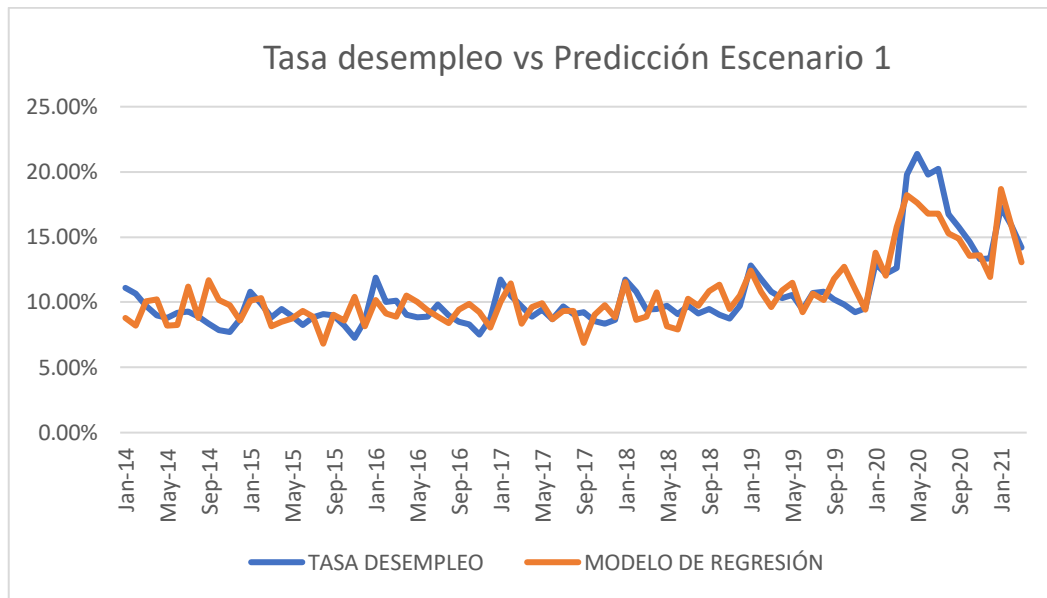


Figura 3. Tasa de desempleo vs Predicción escenario 1

Para la predicción del segundo escenario, se utilizó la metodología de generar números aleatorios para obtener los hits de las próximas cuatro semanas del mes de marzo. Estos números aleatorios se generaron dentro de un rango de dos números, donde a cada una de las variables del modelo se le extrajo su mínimo y máximo valor que obtuvieron desde enero del 2014 a febrero del 2021, y con estos dos valores se creó el rango de aleatorios para la obtención de los hits de las semanas de marzo.

Luego con estos datos, se procedió a calcular el pronóstico semanal de la tasa de desempleo, el cual obtuvo los siguientes resultados:

<b>Semana 1</b>	<b>14.95%</b>
<b>Semana 2</b>	<b>4.00%</b>
<b>Semana 3</b>	<b>13.76%</b>
<b>Semana 4</b>	<b>23.17%</b>

Tabla 15. Predicción escenario 2

Con el escenario dos se puede notar que el valor que toma la tasa de desempleo de manera semanal es mucho más alto y volátil que con el primer escenario, esto debido a que los aleatorios que genera poseen, en algunas variables, un rango más amplio de números, lo

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

cual hace que el resultado final tenga variaciones altas en el pronóstico de la tasa de desempleo entre semanas. Además, con el escenario dos no se determina un alza o decline en el valor de la tasa de desempleo, sino que se observan picos de altos y bajos.

Por otra parte, en su pronóstico mensual utilizando el nowcasting, se obtuvo como resultado un pronóstico de la tasa de desempleo del 10.20%, un valor mucho menor al resultado real del mes de marzo. Lo anterior se ilustra en la siguiente gráfica.

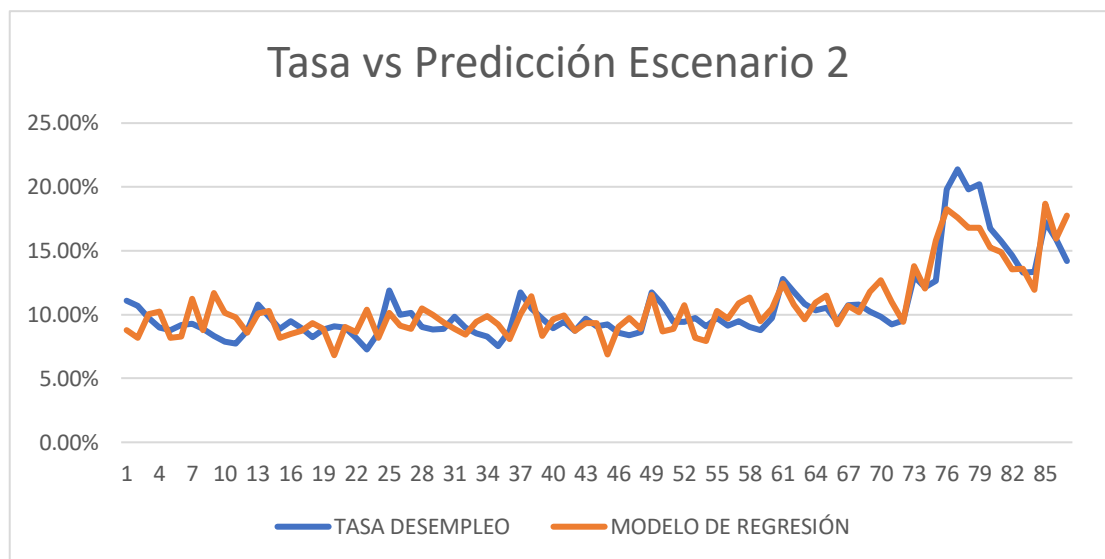


Figura 4. Tasa de desempleo vs Predicción escenario 2

Finalmente, para el tercer y último escenario, se hizo una combinación de las dos metodologías anteriores, en donde tres variables del modelo fueron seleccionadas para que su valor variara de manera aleatoria, mientras que las otras variables del modelo se les aplicó la metodología de interpolación utilizada en el escenario 1.

Las variables que se seleccionaron para la aleatoriedad fueron variables las cuales obtuvieron una de las mayores significancias en el modelo de regresión lineal, además, son variables que corresponden a la categoría de plataformas de empleo, las cuales obtuvieron grandes hits en el buscador de Google y poseen una tendencia alcista. Dichas variables fueron Computrabajo, LinkedIn y Agencia de Empleo.

La aleatoriedad que se realizó para cada una de las variables escogidas consta de nueve combinaciones, donde a cada variable se le aplican tres diferentes escenarios descritos de la siguiente manera:

Primer escenario: Aleatoriedad con hits bajos. Entre 0 y 39

Segundo escenario: Aleatoriedad con hits medios. Entre 40 y 69

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

Tercer escenario: Aleatoriedad con hits altos. Entre 70 y 100

Es importante tener en cuenta que mientras se realizó la aleatoriedad con hits bajos, medios y altos para una variable, las otras dos variables elegidas acogieron la modalidad de interpolación, al igual que las otras variables del modelo. Lo anterior con la intención de analizar qué tan significativo pueden llegar a ser, para la tasa de desempleo, estas variables cuando toman valores pequeños, medianos o grandes.

Para la variable computrabajo, cuando esta toma valores bajos, medios y altos, los pronósticos obtenidos de manera semanal para el mes de marzo fueron los siguientes:

<b>COMPUTRABAJO</b>	<b>Valores Bajos</b>	<b>Valores Medios</b>	<b>Valores Altos</b>
<b>Semana 1</b>	<b>12.15%</b>	<b>19.40%</b>	<b>28.28%</b>
<b>Semana 2</b>	<b>13.69%</b>	<b>21.33%</b>	<b>25.26%</b>
<b>Semana 3</b>	<b>15.40%</b>	<b>20.87%</b>	<b>25.14%</b>
<b>Semana 4</b>	<b>10.79%</b>	<b>21.77%</b>	<b>26.39%</b>

Tabla 16. Predicción escenario 3, con aleatoriedad en computrabajo

Se puede observar que cuando la variable toma valores más bajos, la tasa de desempleo tiende a ser más bajita, mientras que cuando la variable toma valores más altos, el resultado del pronóstico es mucho más alto. Además, se puede notar que los pronósticos semanales en las tres modalidades diferentes muestran altos y bajos durante las cuatro semanas de marzo.

En los resultados de sus respectivos pronósticos mensuales, la tasa de desempleo estimada que se obtuvo para el mes de marzo, teniendo en cuenta valores bajos, medios y altos, fue de 11.86%, 18.18% y 26.90% respectivamente. Se puede notar que a medida que la variable obtiene mayores valores, el valor de la tasa de desempleo tiende a aumentar, lo cual muestra que entre más se busque la palabra computrabajo, es porque posiblemente la tasa de desempleo vaya a aumentar para dicho periodo de tiempo, ya que puede haber más personas en búsqueda de un empleo o trabajo.

En cuanto al resultado real del mes de marzo, se observa que la predicción más cercana al valor fue cuando la variable computrabajo tomo valores bajos, ya que cuando esta obtiene valores medios o altos, el resultado de la predicción dio como resultado valores mayores a la tasa real. Así como se evidencia en la siguiente gráfica.

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

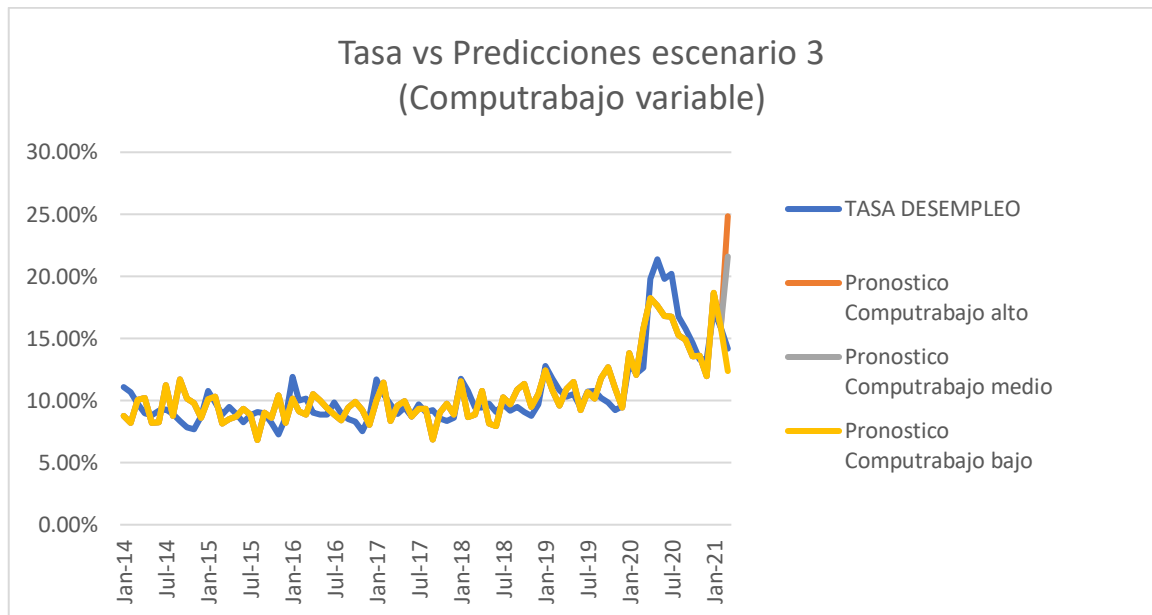


Figura 5. Tasa de desempleo vs Predicciones escenario 3 (computrabajo variable)

En el caso de la variable LinkedIn, cuando esta toma valores bajos, medios y altos, los resultados que se obtuvieron para el pronóstico del mes de marzo de la tasa de desempleo de manera semanal fueron:

<b>LINKEDIN</b>	<b>Valores Bajos</b>	<b>Valores Medios</b>	<b>Valores Altos</b>
<b>Semana 1</b>	<b>10.14%</b>	<b>12.92%</b>	<b>16.28%</b>
<b>Semana 2</b>	<b>6.54%</b>	<b>13.37%</b>	<b>17.14%</b>
<b>Semana 3</b>	<b>9.14%</b>	<b>11.48%</b>	<b>16.67%</b>
<b>Semana 4</b>	<b>7.78%</b>	<b>12.04%</b>	<b>15.70%</b>

Tabla 17. Predicción escenario 3, con aleatoriedad en LinkedIn

Con la variable LinkedIn se observa que cuando la variable toma valores más altos, el resultado del pronóstico de la tasa de desempleo es mucho más alto que cuando la variable toma valores medios o bajos. Al igual que ocurrió con la aleatoriedad en la variable computrabajo, se observa que, en las 3 modalidades, la tasa tiene altos y bajos durante las cuatro semanas del mes.

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

En cuanto a su pronóstico mensual, el resultado más cercano que se obtuvo al valor real de la tasa de desempleo para el mes de marzo fue cuando la variable LinkedIn tomo los valores más altos en los hits, ya que su pronóstico mensual fue de 14.89%, mientras que cuando la variable toma valores medios y bajos, su resultado fue de 11.53% y 9.80% respectivamente. Lo anterior se puede ver ilustrado en la siguiente gráfica:

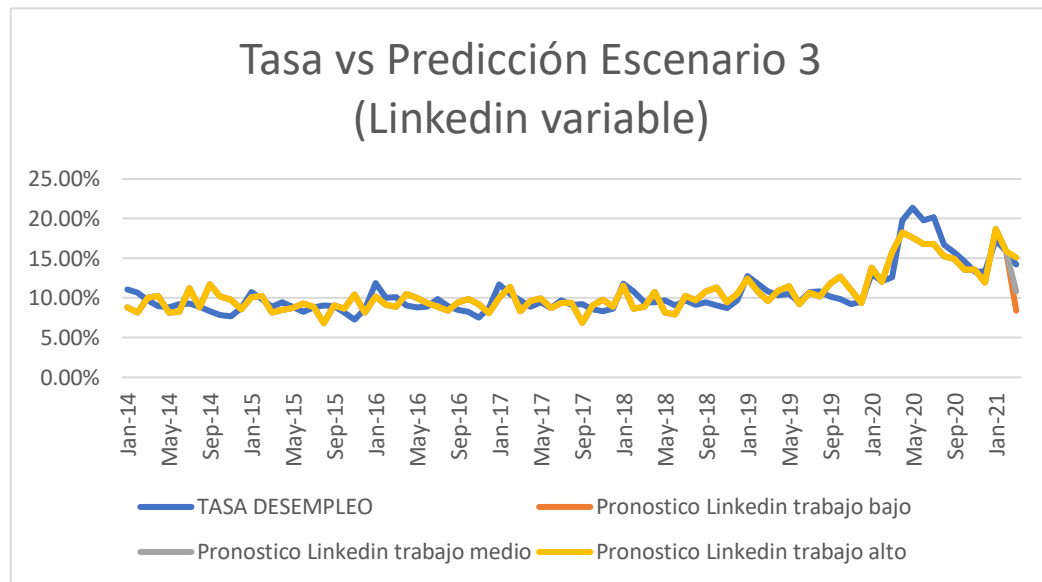


Figura 6. Tasa de desempleo vs Predicciones escenario 3 (LinkedIn variable)

Finalmente, para la variable Agencia de Empleo, las predicciones semanales obtenidas para marzo, con valores altos, medios y bajos, fueron los siguientes:

<b>AGENCIA DE EMPLEO</b>	<b>Valores Bajos</b>	<b>Valores Medios</b>	<b>Valores Altos</b>
<b>Semana 1</b>	<b>13.32%</b>	<b>15.96%</b>	<b>19.55%</b>
<b>Semana 2</b>	<b>13.36%</b>	<b>16.19%</b>	<b>20.28%</b>
<b>Semana 3</b>	<b>14.82%</b>	<b>17.91%</b>	<b>20.11%</b>
<b>Semana 4</b>	<b>13.18%</b>	<b>15.55%</b>	<b>18.34%</b>

Tabla 18. Predicción escenario 3, con aleatoriedad en Agencia de empleo

Al igual que con las otras dos variables, se puede observar que el pronóstico para la tasa de desempleo de cada semana va aumentando a medida que la variable toma mayores

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.



valores, de lo cual se puede decir que a medida que las personas busque más la palabra Agencia Empleo, la tasa de desempleo tiende a aumentar. Además, se observa que, para cada modalidad, las tasas entre semanas poseen altos y bajos, debido a la aleatoriedad que se maneja en cada una de las modalidades.

Por otra parte, el pronóstico mensual para el mes de marzo dio como resultados 18.04%, 15.35% y 13.65% cuando la variable toma valores altos, medios y bajos respectivamente. Lo anterior se ilustra en la siguiente gráfica:

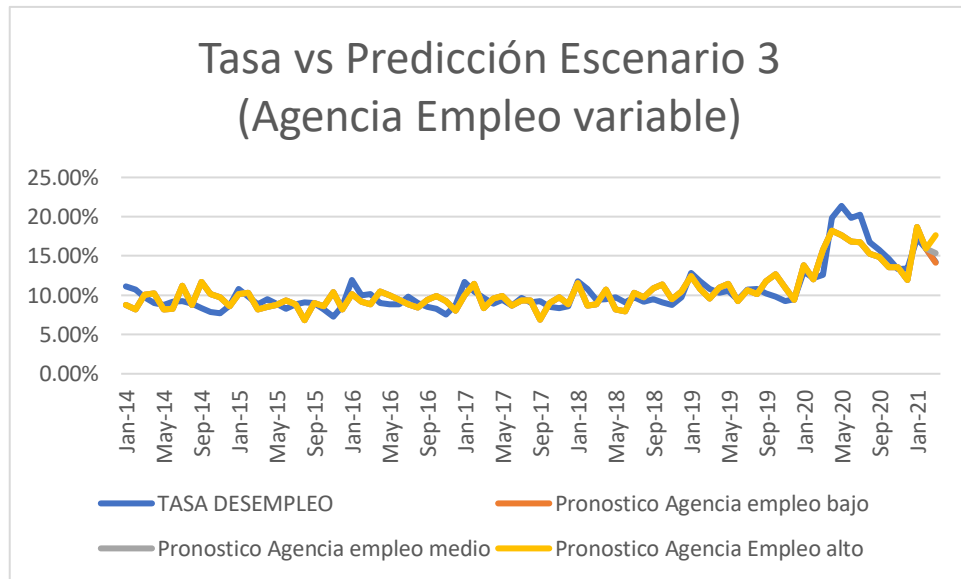


Figura 7. Tasa de desempleo vs Predicciones escenario 3 (Agencia de Empleo variable)

## 4. CONCLUSIONES Y CONSIDERACIONES FINALES

En el presente trabajo se presentó el modelo de Now-casting como una herramienta útil para llevar a cabo pronósticos en variables macroeconómicas, específicamente en la tasa de desempleo. El modelo de Now-casting muestra en un “tiempo real” que tanto puede llegar a cambiar una variable en específico, o que posibles picos pueda tener. Lo anterior por medio del uso de palabras claves que las personas suelen digitar en el navegador y que son medidas a través de un indicador de Google Trends.

El modelo que se utilizó con la finalidad de obtener una predicción o un acercamiento de como posiblemente se pueda mover la tasa de desempleo semanalmente, fue un modelo creado por 19 palabras claves, las cuales nos sirvieron de ayuda para analizar el comportamiento de dicha variable, ya que sus comportamientos se ajustaron al movimiento que ha tenido la tasa de desempleo durante los últimos seis años. Además, el modelo de regresión lineal que se utilizó fue un modelo estadísticamente significativo para analizar los movimientos futuros de la tasa de desempleo.

Por otra parte, se observa que entre más palabras claves posee el modelo, su significancia será mucho mayor, ya que al tener en cuenta más posibles palabras que las personas suelen buscar a la hora de encontrar un trabajo o empleo, el resultado final se ajustará más al resultado real, ya que se están teniendo en cuenta más posibilidades. Asimismo, se deben de tener en cuenta ciertos aspectos si se quiere obtener un buen pronóstico de Now-casting; en este caso fue el acceso a internet que tenían las personas en todo el período de tiempo que se analizó; esto debido a que son factores externos que pueden llegar a tener un alto impacto en el resultado final, ya que como se demostró durante el desarrollo del trabajo, la penetración del internet antes del 2016 era muy baja, lo que podía llevar a resultados más lejanos de lo real, ya que posiblemente no era tan común que las personas antes del 2016 buscaban un empleo por medio de una plataforma web, si no por medios más tradicionales como lo son el periódico y la radio.

Finalmente, se concluye que el modelo de predicción fue un modelo efectivo para analizar los posibles comportamientos de la tasa de desempleo, no obstante, se observa poca efectividad cuando se procede a generar aleatorios en el modelo, pero también se observa que el modelo adquiere una mayor efectividad cuando se da uso de la interpolación para realizar el pronóstico de manera semanal. Sin embargo, se hace la recomendación para investigaciones o proyectos posteriores de que los datos que vayan a utilizar para pronosticar cierta variable tengan la misma periodicidad a la que quieren predecir, ya que esto le dará una mayor efectividad y ajuste al modelo y por consecuencia al resultado final. También se recomienda el uso y comparación de otros tipos de modelos de predicción, como regresión lineal Multiple, red neuronal, modelos ARIMA, entre otros, con el fin de analizar cuál de estos se ajusta más al resultado final.

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

## 5. REFERENCIAS

- Anvik, C., & Gjelstad, K. (2010). *“Just Google It!”: Forecasting Norwegian unemployment figures with web queries*. Oslo: BI Norwegian School of Management.
- Artola, C., & Galán, E. (2012). TRACKING THE FUTURE ON THE WEB: CONSTRUCTION OF LEADING INDICATORS USING INTERNET SEARCHES. (1203), 29.
- Askatas, N., & Zimmerman, K. (06 de 2009). Google Econometrics and. *Applied Economics Quarterly*, 55(2), 107-120.
- Bańbura, M. (2011). *Nowcasting with daily data*. European Central Bank.
- Camacho, C. (2006). *Google Scholar*. Obtenido de <https://personal.us.es/vararey/regresion-simple.pdf>
- Castro, M. C., & Nevárez, J. B. (27 de 06 de 2014). Las crisis económicas y sus efectos en el mercadode trabajo, en la desigualdad y en la pobreza de México. México. Obtenido de <https://reader.elsevier.com/reader/sd/pii/S0186104215000042?token=5DDD92F8502A3E97AE86F3A7A14E7E7C433E57444DCF94CC42507699F2CDE18F0CEFB1318BC91263764830F82CF15E6E>
- Choi, H., & Varian, H. (18 de 03 de 2009). Predicting the Present with Google Trends. Estados Unidos. Obtenido de [https://www.frbsf.org/economic-research/files/Varian-part\\_1.pdf](https://www.frbsf.org/economic-research/files/Varian-part_1.pdf)
- Cortés, J., Bielsa, N., & Cobo, E. (2015). *Bioestadística para no estadísticos*. Cataluña: Universidad Politécnica de Cataluña.
- Cristiano Botia, D. J., Hernandez Bejarano, M. D., & Ramos Veloza, M. A. (25 de 06 de 2019). *Repositorio Banco de La República*. Obtenido de Banco de La República: <https://repositorio.banrep.gov.co/handle/20.500.12134/9965>
- Cristiano-Botia, D. J., Hernandez-Bejarano, M. D., & Ramos-Veloza, M. A. (25 de 06 de 2019). Labor Market Indicator for Colombia (LMI). Bogotá, Colombia.
- Domenech, R., & Gómez, V. (2005). *CICLO ECONÓMICO Y DESEMPLEO* (Vol. XXIX). Madris, España: Investigaciones Económicas.
- Felipe, V.-G., & Gabriel, A.-G. (2015). Simulación de campos aleatorios espacio-temporales utilizando un. *Ingeniería Investigación y Tecnología*, 1-12.
- García, L. F., & Cruz, M. (2017). *DESEMPLEO EN AMÉRICA LATINA: ¿FLEXIBILIDAD LABORAL O ACUMULACIÓN DE CAPITAL?*

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

- Ghysels, E., & Qian, H. (2018). Estimating MIDAS regressions via OLS with polynomial. *Econometrics and Statistics*, 2452-3062.
- González - Casimiro, M. P. (2009). *Análisis de Series Temporales: Modelos ARIMA*. País Vasco: Facultad de Ciencias Económicas y Empresariales. Universidad del País Vasco.
- Google. (s.f.). *Google Trends*. Obtenido de <https://trends.google.com/trends/?geo=CO>
- Kapetanios, G., & Papailias, F. (2018). Big Data & Macroeconomic Nowcasting: *ECONOMIC STATISTICS CENTRE OF EXCELLENCE*, 2-77.
- Merino, G. (2012). El Movimiento Obrero Organizado, la crisis de 2001 y el gobierno de Duhalde: El caso de la CGT disidente. En U. N. Plata, *Memoria Académica* (págs. 87-119). Facultad de Humanidades y Ciencias de la Educación (FaHCE) . Obtenido de [http://www.memoria.fahce.unlp.edu.ar/art\\_revistas/pr.5677/pr.5677.pdf](http://www.memoria.fahce.unlp.edu.ar/art_revistas/pr.5677/pr.5677.pdf)
- Muñoz, N. A. (2019). *ANÁLISIS DE LA RELACIÓN ENTRE EL DESEMPLEO Y LA POBREZA EN COLOMBIA EN LOS AÑOS 2010 AL 2018*. Ibagué: Universidad Cooperativa de Colombia. Obtenido de Google Scholar: [https://repository.ucc.edu.co/bitstream/20.500.12494/13219/1/2019\\_An%C3%A1lisis\\_Relaci%C3%B3n\\_Desempleo.pdf](https://repository.ucc.edu.co/bitstream/20.500.12494/13219/1/2019_An%C3%A1lisis_Relaci%C3%B3n_Desempleo.pdf)
- Nagao, S., Takeda, F., & Tanaka, R. (08 de 04 de 2019). Nowcasting of the U.S. unemployment rate using Google Trends. *Finance Research Letters*, 103-109. Obtenido de <https://reader.elsevier.com/reader/sd/pii/S1544612319301072?token=753CDF58C4EF82C6EEB594F0E5DA8D6D51542FC7071FE7589381F37B0753B3D4167EA3A311F4EC961BB677F27009EEBA>
- Nagao, S., Takeda, F., & Tanaka, R. (14 de 03 de 2019). Nowcasting of the U.S. unemployment rate using Google Trends. *Finance Research Letters*, 103-109.
- Novales, A. (2017). *Modelos vectoriales autoregresivos (VAR)*.
- Pachón, G. A. (12 de 2012). TENDENCIAS EN LA INVESTIGACIÓN PSICOLÓGICA SOBRE DESEMPLEO Y SALUD. *IBEROAMERICANA DE PSICOLOGÍA: CIENCIA Y TECNOLOGÍA*, 5, 17-30. Obtenido de <https://reviberopsicologia.iberu.edu.co/article/view/rip.5202/210>
- Planeación, D. N. (2017). *Pobreza monetaria y multidimensional departamental: necesidad de políticas*. DPN. Departamento Nacional de Planeación. Obtenido de <https://colaboracion.dnp.gov.co/CDT/Desarrollo%20Territorial/Portal%20Territorial/KitSeguimiento/Pobreza/Publicaci%C3%B3n%20lpm%20deptal.pdf>
- Pontaque, F. C. (2005). *Modelos Lineales*. Barcelona: Universidad de Barcelona.

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

- República, B. d. (s.f.). *Tasas de ocupación y desempleo*. Bogotá: DANE.
- Richardson, P. (2018). Nowcasting and the Use of Big Data in Short-Term. *Economie et Statistique / Economics and Statistics*, 505-506, 65–87. Obtenido de <https://www.insee.fr/en/statistiques/3706173?sommaire=3706269>
- Rocha, F. (01 de 05 de 2012). La crisis económica y sus efectos sobre el empleo en España. España. Obtenido de <http://docpublicos.ccoo.es/cendoc/035344CrisisEconomicaEfectos.pdf>
- Smith, P. (14 de 02 de 2016). Google's MIDAS Touch: Predicting UK Unemployment with. *Journal of Forecasting, J. Forecast*, 263–284.
- TIC, C. (2017). *Colombia TIC*. Obtenido de Colombia TIC: <https://colombiatic.mintic.gov.co/679/w3-propertyvalue-36359.html>
- Trejo, J., Rivera, E., & Rios, H. (2017). Análisis de la histéresis del desempleo en México ante shocks macroeconómicos. *Contaduría y Administración*, 62(4), 1228-1248. Obtenido de <https://www-sciencedirect-com.recursosbiblioteca.eia.edu.co/science/article/pii/S0186104217300670>
- Valle, A. (2007). *Algunos modelos importantes en la investigación pedagógica*. La Habana, Cuba: Instituto Central de Ciencias Pedagógicas.
- Zaman, E. K. (12 de 02 de 2019). Financial nowcasts and their usefulness in macroeconomic forecasting. *International Journal of Forecasting*, 1-17.

# ANEXO 1

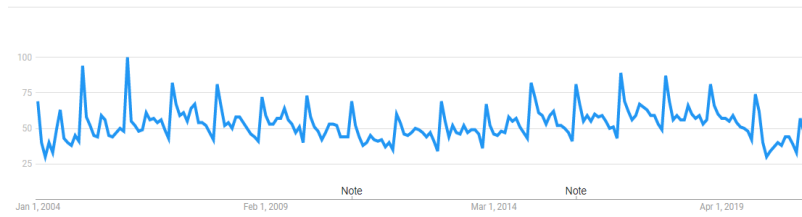


Figura 1. Empleo



Figura 2. Ofertas laborales

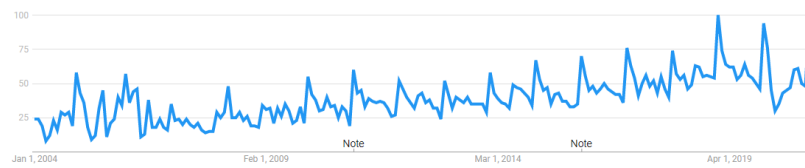


Figura 3. Vacantes



Figura 4. Hoja de vida

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.



Figura 5. Teletrabajo



Figura 6. Trabajo sin experiencia

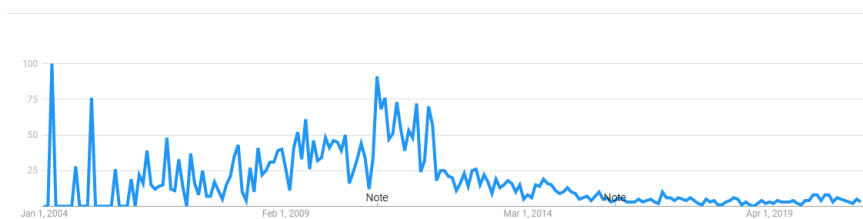


Figura 7. Trabajando Colombia

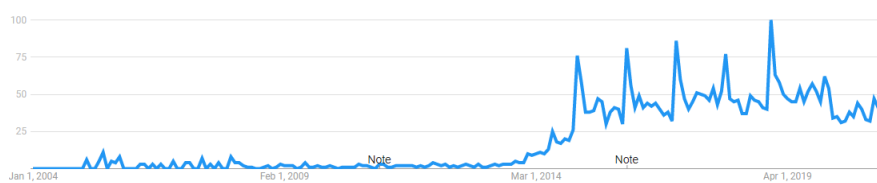


Figura 8. Agencia de empleo



Figura 9. Empleo SENA

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

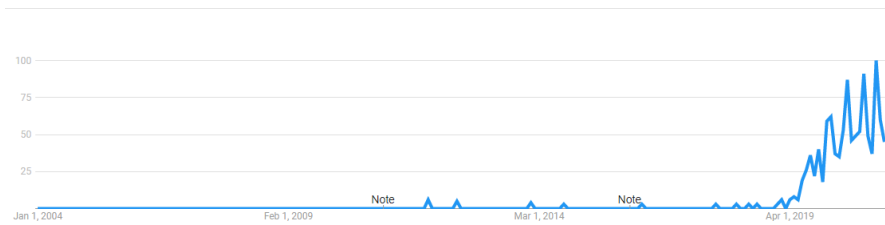


Figura 10. Magneto empleo



Figura 11. LinkedIn



Figura 12. Computrabajo

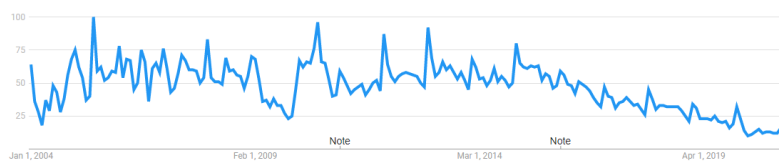


Figura 13. Eempleo



Figura 14. Trabaja con nosotros

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.



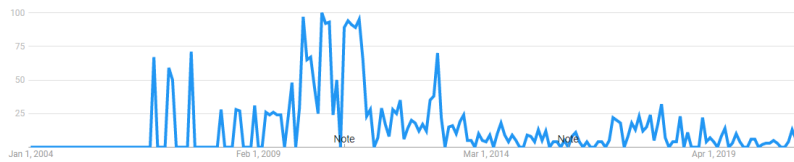


Figura 15. Trabajos.com



Figura 16. Servicio público de empleo

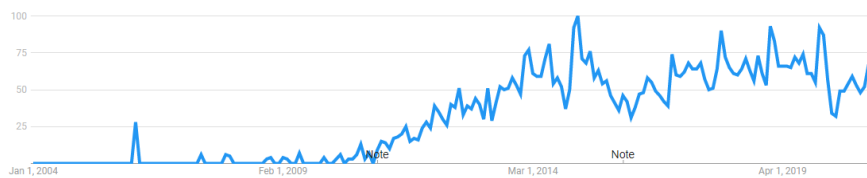
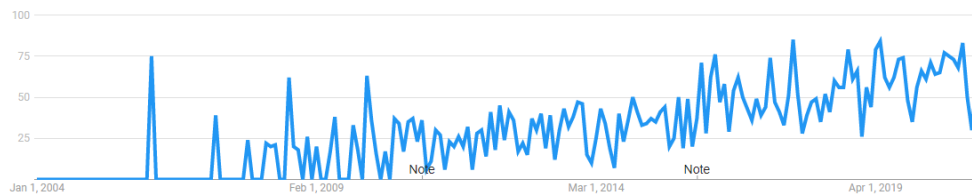


Figura 17. Indeed



Figura 18. Ministerio de trabajo



La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

Figura 19. Trabajo ya



Figura 20. Buscador de empleo

```
library(reshape2)
library(maps)
library(rworldmap)
library(tidyverse)
library(boot)
library(ggplot2)

attach(es1_desempleo)
attach(es1_empleo)
attach(es1_eempleo)
attach(es1_empleosena)
attach(es1_hojadevida)
attach(es1_ministeriodetrabajo)
attach(es1_ofertasborales)
attach(es1_teletrabajo)
attach(es1_trabajandocolombia)
attach(es1_vacantes)

attach(es6_desempleo)
attach(es6_empleo)
attach(es6_ofertaslaborales)
attach(es6_vacantes)
attach(es6_hojadevida)
```

Figura 21. Código. Librerías y palabras claves

```

attach(es6_teletrabajo)
attach(es6_trabajandocolombia)
attach(es6_empleosena)
attach(es6_eempleo)
attach(es6_ministeriodetrabajo)
attach(es6_computrabajo)
attach(es6_serviciopublicoempleo)
attach(es6_trabajos_com)
attach(es6_linkedin)
attach(es6_indeed)
attach(es6_trabajoya)
attach(es6_trabajosinexperiencia)
attach(es6_trabajaconnosotros)
attach(es6_agenciadempleo)
attach(es6_buscadorempleo)

```

Figura 22. Código. Librerías y palabras claves

```

##### ESCENARIO 1 #####
es1_desempleo<-unlist(es1_desempleo)
es1_empleo<-unlist(es1_empleo)
es1_ofertaslaborales<-unlist(es1_ofertaslaborales)
es1_vacantes<-unlist(es1_vacantes)
es1_hojadevida<-unlist(es1_hojadevida)
es1_teletrabajo<-unlist(es1_teletrabajo)
es1_trabajandocolombia<-unlist(es1_trabajandocolombia)
es1_empleosena<-unlist(es1_empleosena)
es1_eempleo<-unlist(es1_eempleo)
es1_ministeriodetrabajo<-unlist(es1_ministeriodetrabajo)

es1_modelo<- lm(es1_desempleo+es1_empleo+es1_ofertaslaborales+es1_vacantes+es1_hojadevida+
               es1_teletrabajo+es1_trabajandocolombia+es1_empleosena+es1_eempleo+es1_ministeriodetrabajo)

summary(es1_modelo)

```

Figura 23. Código. Escenario 1

```

##### ESCENARIO 2 #####
es2_desempleo<-unlist(es2_desempleo)
es2_empleo<-unlist(es2_empleo)
es2_ofertaslaborales<-unlist(es2_ofertaslaborales)
es2_vacantes<-unlist(es2_vacantes)
es2_hojadevida<-unlist(es2_hojadevida)
es2_teletrabajo<-unlist(es2_teletrabajo)
es2_trabajandocolombia<-unlist(es2_trabajandocolombia)
es2_empleosena<-unlist(es2_empleosena)
es2_eempleo<-unlist(es2_eempleo)
es2_ministeriodetrabajo<-unlist(es2_ministeriodetrabajo)
es2_computrabajo<- unlist(es2_computrabajo)
es2_serviciopublicoempleo<-unlist(es2_serviciopublicoempleo)

es2_modelo<- lm(es2_desempleo+es2_empleo+es2_ofertaslaborales+es2_vacantes
               +es2_hojadevida+es2_teletrabajo+es2_trabajandocolombia
               +es2_empleosena+es2_eempleo+es2_ministeriodetrabajo
               +es2_computrabajo+es2_serviciopublicoempleo)

summary(es2_modelo)

```

La información presentada en este documento es de exclusiva responsabilidad de los autores y no compromete a la EIA.

Figura 24. Código. Escenario 2

```
##### ESCENARIO 3 #####
es3_desempleo<-unlist(es3_desempleo)
es3_empleo<-unlist(es3_empleo)
es3_ofertaslaborales<-unlist(es3_ofertaslaborales)
es3_vacantes<-unlist(es3_vacantes)
es3_hojadevida<-unlist(es3_hojadevida)
es3_teletrabajo<-unlist(es3_teletrabajo)
es3_trabajandocolombia<-unlist(es3_trabajandocolombia)
es3_empleosena<-unlist(es3_empleosena)
es3_eempleo<-unlist(es3_eempleo)
es3_ministeriodetrabajo<-unlist(es3_ministeriodetrabajo)
es3_computrabajo<- unlist(es3_computrabajo)
es3_serviciopublicoempleo<-unlist(es3_serviciopublicoempleo)
es3_trabajos_com<-unlist(es3_trabajos_com)

es3_modelo<- lm(es3_desempleo~es3_empleo+es3_ofertaslaborales+es3_vacantes
+es3_hojadevida+es3_teletrabajo+es3_trabajandocolombia
+es3_empleosena+es3_eempleo+es3_ministeriodetrabajo+es3_computrabajo
+es3_serviciopublicoempleo+es3_trabajos_com)
summary(es3_modelo)
```

Figura 25. Código. Escenario 3

```
##### ESCENARIO 4 #####
es4_desempleo<-unlist(es4_desempleo)
es4_empleo<-unlist(es4_empleo)
es4_ofertaslaborales<-unlist(es4_ofertaslaborales)
es4_vacantes<-unlist(es4_vacantes)
es4_hojadevida<-unlist(es4_hojadevida)
es4_teletrabajo<-unlist(es4_teletrabajo)
es4_trabajandocolombia<-unlist(es4_trabajandocolombia)
es4_empleosena<-unlist(es4_empleosena)
es4_eempleo<-unlist(es4_eempleo)
es4_ministeriodetrabajo<-unlist(es4_ministeriodetrabajo)
es4_computrabajo<- unlist(es4_computrabajo)
es4_serviciopublicoempleo<-unlist(es4_serviciopublicoempleo)
es4_trabajos_com<-unlist(es4_trabajos_com)
es4_linkedin<- unlist(es4_linkedin)
es4_indeed<-unlist(es4_indeed)
es4_trabajoya<-unlist(es4_trabajoya)

es4_modelo<- lm(es4_desempleo~es4_empleo+es4_ofertaslaborales+es4_vacantes
+es4_hojadevida+es4_teletrabajo+es4_trabajandocolombia
+es4_empleosena+es4_eempleo+es4_ministeriodetrabajo+es4_computrabajo
+es4_serviciopublicoempleo+es4_trabajos_com+es4_linkedin+es4_indeed
+es4_trabajoya)
summary(es4_modelo)
```

Figura 26. Código. Escenario 4

```
##### ESCENARIO 5 #####
es5_desempleo<-unlist(es5_desempleo)
es5_empleo<-unlist(es5_empleo)
es5_ofertaslaborales<-unlist(es5_ofertaslaborales)
es5_vacantes<-unlist(es5_vacantes)
es5_hojadevida<-unlist(es5_hojadevida)
es5_teletrabajo<-unlist(es5_teletrabajo)
es5_trabajandocolombia<-unlist(es5_trabajandocolombia)
es5_empleosena<-unlist(es5_empleosena)
es5_eempleo<-unlist(es5_eempleo)
es5_ministeriodetrabajo<-unlist(es5_ministeriodetrabajo)
es5_computrabajo<- unlist(es5_computrabajo)
es5_serviciopublicoempleo<-unlist(es5_serviciopublicoempleo)
es5_trabajos_com<-unlist(es5_trabajos_com)
es5_linkedin<- unlist(es5_linkedin)
es5_indeed<-unlist(es5_indeed)
es5_trabajoya<-unlist(es5_trabajoya)
es5_trabajosinexperiencia<-unlist(es5_trabajosinexperiencia)
es5_trabajaconnosotros<-unlist(es5_trabajaconnosotros)

es5_modelo<- lm(es5_desempleo~es5_empleo+es5_ofertaslaborales+es5_vacantes
+es5_hojadevida+es5_teletrabajo+es5_trabajandocolombia
+es5_empleosena+es5_eempleo+es5_ministeriodetrabajo+es5_computrabajo
+es5_serviciopublicoempleo+es5_trabajos_com+es5_linkedin+es5_indeed
+es5_trabajoya+es5_trabajosinexperiencia+es5_trabajaconnosotros)
summary(es5_modelo)
```

Figura 27. Código. Escenario 5

```
##### ESCENARIO 6 #####
es6_desempleo<-unlist(es6_desempleo)
es6_empleo<-unlist(es6_empleo)
es6_ofertaslaborales<-unlist(es6_ofertaslaborales)
es6_vacantes<-unlist(es6_vacantes)
es6_hojadevida<-unlist(es6_hojadevida)
es6_teletrabajo<-unlist(es6_teletrabajo)
es6_trabajandocolombia<-unlist(es6_trabajandocolombia)
es6_empleosena<-unlist(es6_empleosena)
es6_eempleo<-unlist(es6_eempleo)
es6_ministeriodetrabajo<-unlist(es6_ministeriodetrabajo)
es6_computrabajo<- unlist(es6_computrabajo)
es6_serviciopublicoempleo<-unlist(es6_serviciopublicoempleo)
es6_trabajos_com<-unlist(es6_trabajos_com)
es6_linkedin<- unlist(es6_linkedin)
es6_indeed<-unlist(es6_indeed)
es6_trabajoya<-unlist(es6_trabajoya)
es6_trabajosinexperiencia<-unlist(es6_trabajosinexperiencia)
es6_trabajaconnosotros<-unlist(es6_trabajaconnosotros)
es6_agenciadempleo<-unlist(es6_agenciadempleo)
es6_buscadorempleo<-unlist(es6_buscadorempleo)
```

Figura 28. Código. Escenario 6

```
es6_modelo<- lm(es6_desempleo~es6_empleo+es6_ofertas laborales+es6_vacantes
+es6_hojadevida+es6_teletrabajo+es6_trabajandocolombia
+es6_empleosena+es6_elempelo+es6_ministeriodetrabajo+es6_computrabajo
+es6_serviciopublicoempleo+es6_trabajos_com+es6_linkedin+es6_indeed
+es6_trabajoya+es6_trabajosinexperiencia+es6_trabajaconnosotros
+es6_agenciadempleo+es6_buscadoremplo)
summary(es6_modelo)
plot(es6_modelo)
```

Figura 29. Código. Escenario 6