**PROYECTO DE INVESTIGACIÓN FINAL**



**Tema:** Situación socio-económica de los migrantes venezolanos en el Perú

**Curso:** Machine Learning aplicado a las Ciencias Sociales

**Alumnos:**

**-** Sergio Romero Miraval (20167264)

**-** Santiago Nicandro Machaca Bernabé (20160464)

**Profesores:**

**-** Pedro Achanccaray

- Pavel Coronado

- José Incio

**Año:** 2021

**I.- Introducción**

Según cifras del Banco Mundial (2020), en el Perú se encuentran 1.2 millones de migrantes y refugiados venezolanos. Con dicha cifra, Perú se sitúa como el segundo país con más migrantes venezolanos en el mundo, solo por detrás de Colombia. Asimismo, se estima que esta es la migración más grande del siglo XXI en el Perú. Si bien existen mecanismo regionales y nacionales existentes para la recepción y regularización de migrantes venezolanos, ello no ha evitado que se lleven a cabo medidas *ad hoc.* Es decir, no ha sido suficiente la Declaración de Cartagena, ya que la generosidad inicial en toda la región se ha convertido en un contexto de mayor xenofobia, mayor políticas restrictivas y demás.

Actualmente Venezuela atraviesa una crisis socioeconómica y política, caracterizada por un aumento de la pobreza, la escasez de alimentos y el aumento de la inseguridad. Ello ha causado que muchos venezolanos decidan salir de su país mediante cualquier medio, muchas veces de manera irregular. Esta situación de vulnerabilidad se ha visto aumentada en los países de recepción. En el caso de Perú, Lima ha sido la ciudad con la mayor recepción de migrantes venezolanos. A pesar de las políticas implementadas por el Estado peruano, la población venezolana aún se encuentra en un estado de vulnerabilidad.

De un lado, la economía informal imperante en la capital lleva a que los migrantes venezolanos estén más expuestos a empleos vulnerables y se concentran en pequeñas empresas. Según cifras del Banco Mundial (2020), 67% de los venezolanos trabajan en pequeñas empresas, las cuales están siendo afectadas por la pandemia del Covid-19. Asimismo, dentro de este grupo, solo el 6% cuenta con algún contrato. De otro lado, los migrantes venezolanos ganan menos que los peruanos y tienen niveles de pobreza más altos. La brecha salarial se calcula en un 37%, el 18% de los venezolanos se encuentran en condiciones de pobreza y el 4% en extrema pobreza. A todo ello, se suma que los migrantes y refugiados venezolanos tienen un menor acceso a servicios de salud, y su acceso a programas sociales y de empleo es limitado.

En este contexto se sitúa la presente investigación. Teniendo en cuenta que el Estado peruano no puede ignorar en su agenda política la migración venezolana, y que ha mostrado señales de que es un punto relevante en política interna y exterior regularizar la situación de los migrantes venezolanos, se propone la siguiente pregunta de investigación: ¿Cuáles son las variables más importantes al analizar la situación socio-económica de la población venezolana en Lima Metropolitana? La presente investigación tiene dos objetivos principales. El primero es poder analizar los factores que llevan a una mejor o peor situación socioeconómica caracterizada en el sueldo percibido mensualmente. El segundo objetivo es analizar cuáles son las variables que más logran explicar la vulnerabilidad de los migrantes venezolanos, con el fin de ayudar a la priorización de recursos al momento de implementar políticas públicas dirigidas a dicha población.

La data que usaremos para nuestro análisis es la brindada por el Instituto Nacional de Estadística e Informática (INEI) en el marco del Plan de Acción Multianual 2017-2018. Dicha encuesta se llevó a cabo en los meses de noviembre-diciembre del año 2018. La Encuesta Dirigida a la Población Venezolana que reside en el país (ENPOVE) se llevó a cabo con la ayuda de organizaciones como el Banco Mundial, la Agencia de las Naciones Unidas para los Refugiados, la Organización Internacional para las Migraciones y el el Fondo de las Naciones Unidas para la Infancia. Dicha encuesta se realizó a 9487 personas, en las ciudades de Arequipa, Cusco, La Libertad y Lima Metropolitana. Asimismo, se consultaron 117 preguntas sobre temas varios, como vivienda, familia, salud, educación, empleo, discriminación, género, etc.

Para realizar el entrenamiento y las validaciones de los modelos predictivos se utilizarán los métodos *hold-out* y *k-fold-cross-validation* para el resampleo de la data. Así, se obtendrá un primer set que se entrenará con los métodos Ridge, LASSO, Random forest y Boosted trees y se validará con 5-fold-cross-validation. Posteriormente se evaluarán los modelos predictivos con un set de validación y se guardará un último set para una validación final previa a la presentación de resultados. Asimismo, se aprovecharán los resultados de la regresión LASSO para identificar los predictores más determinantes en el sueldo monetario de la población venezolana en Perú

**II.- Revisión de Literatura**

Respecto a la revisión de literatura, está dividida en dos secciones. Según la literatura revisada, existen ciertas condiciones en las que las personas de un determinado país se ven forzadas a migrar, por lo que es importante notar sus condiciones. Asimismo, la literatura presenta las condiciones que expone a los migrantes a situaciones vulnerables, destacando aspectos económicos y sociales.

En primer lugar, según información del Departamento de Asuntos Económicos y Sociales de la Organización de las Naciones Unidas (UNDESA), solo durante el año 2010 hubo 214 millones de migrantes internacionales en todo el mundo. Ello convierte a la migración internacional en algo muy relevante en términos económicos y políticos para los países. Las mismas cifras revelan que los migrantes se concentran fuertemente en los países más desarrollados y muchos menos en los países de desarrollo. Para el año 2010 se estimaba que el 60% de los migrantes vivían en países desarrollados. Dentro de los principales motivos para migrar se encuentran los altos niveles de desigualdad en un país, la seguridad del Estado y seguridad humana, los avances técnicos que alientan la movilidad, la demanda de una mano de obra usualmente más barata y cambios demográficos (Castles, 2014, 236-238). Dicha situación está caracterizada en el caso venezolano.

La procedencia de las personas migrantes es también un aspecto importante a considerar, mucho más en el contexto de América Latina y el Caribe, por lo que se distingue entre migración intra y extrarregional. Si bien se han establecido acuerdos en distintos niveles (regional, subregional y nacional) de movilidad y residencia, tales como los de integración económica de la Comunidad Andina de Naciones (CAN) y el MERCOSUR, en comparación con otras regiones existe un menor avance en términos de libre circulación y residencia (Carrasco y Suárez, 2018, 13-15). Pese a esta situación, los migrantes deciden salir de sus países para poder tener una mayor protección, ser parte de los sistemas de protección social y mejorar su calidad de vida, ya que no lo pueden hacer en sus países de origen. No obstante, una vez llegado al país de destino, la situación de vulnerabilidad puede ser igual o incluso peor.

En segundo lugar, la literatura revisada considera algunos aspectos que llevan a una situación de vulnerabilidad de los migrantes. Entre el 2016 y 2018, Ecuador y Perú fueron relativamente generosos con los migrantes venezolanos. En el caso peruano, Perú fue el primer país de la región en crear un permiso específico que regulariza la situación migratoria de los venezolanos. A esto se le llamó el Permiso de Permanencia Temporal (PTP), que les autorizaba vivir y trabajar en el país durante el periodo de un año. Sin embargo, el anuncio de la exigencia de pasaporte para el ingreso de inmigrantes venezolanos ocurrió a mediados del 2018. Esta medida fue objeto de demanda por parte de la Coordinadora Nacional de Derechos Humanos y, si bien en un inicio se revocó, entró en vigencia a inicios del 2019, y luego se recortó el plazo del PTP (Freire y Castillo, 2020a, 2). Dichas medidas restrictivas se han sumado al contexto del Covid-19, que ha complejizado mucho más la dinámica migratoria.

Freire y Castillo consideran que en los últimos años, crisis como la migración venezolana ha representado una prueba para la aplicación de normas migratorias generosas en situaciones reales. No obstante, se evidenciaron problemas como la débil coordinación regional y la dificultad para garantizar derechos, sobre todo los sociales. Al mismo tiempo, la politización de la migración venezolana ha dificultado un abordaje técnico y sensato sobre el fenómeno migratorio, lo cual llevó a una disparidad en las respuestas estatales, dejando más vulnerables a los migrantes venezolanos (2020b, 53-55). Sumado a ello, Hernández-Vasquéz *et al*. encuentran que la no utilización de servicios de salud afecta directamente la calidad de vida de los venezolanos. El no contar con suficiente dinero, el carecer de seguro de salud, o simplemente no disponer de tiempo eran motivos por los que no utilizaban los servicios de salud, lo cual empeoraba su situación en el Perú (2019, 583). En esa misma línea, se encuentra el estudio de Mendoza y Miranda, que analizan los desafíos y las situaciones a las que se enfrentan los migrantes venezolanos en Perú. Así, los autores encuentran que existen principalmente dos factores que explican la situación de vulnerabilidad de los migrantes venezolanos. Estos son las experiencias de discriminación y violencia, y la imposibilidad de homologar los estudios que tuvieron en Venezuela (Mendoza y Miranda, 2019, 503-505).

Con todo lo revisado, podemos concluir que la literatura presenta dos escenarios. El primero de ellos es la situación de vulnerabilidad de la que parten en sus países de origen. Es decir, situaciones de pobreza, hambre o conflictos violentos. Ello resulta importante en tanto nos permite dilucidar en qué áreas los migrantes venezolanos pueden llegar a verse muchos más expuestos a situaciones de vulnerabilidad, empleos informales y demás. El segundo escenario que nos presenta la literatura es la situación interna de los países receptores. Dicho de otro modo, la situación legal en la que se encuentran los migrantes, si han estado expuestos a situaciones de discriminación o no logran hacer uso de los serivicos médicos por diversoso motivos. Estos factores, que por razones analíticas materializamos en variables independientes para explicar la situación socioeconómica, serán operacionalizados y luego analizados con nuestra variable dependiente: el sueldo total mensual percibido.

**III.- Metodología y data**

El trabajo se realizará utilizando como insumo principal los resultados de la “Encuesta Dirigida a la Población Venezolana que Reside en el País” - ENPOVE 2018 realizada en los meses de noviembre - diciembre del año 2018. La encuesta mencionada recopila los datos de 9487 ciudadanos venezolanos residentes en Perú y contiene respuestas a 117 preguntas en temas de vivienda y hogar, miembros del hogar, migración, salud, educación, empleo, discriminación, y género y victimización.

De dichas respuestas, tendremos como variable dependiente y objetivo de predicción los ingresos monetarios que recibieron en el mes anterior a la encuesta. Y como variables independientes y posibles predictores consideramos la condición de ingreso (turista, estudiante, permiso temporal de permanencia, refugiado), sexo, lugar de residencia, nivel de estudios, autoidentificación étnica, pertenencia a la comunidad LBGT+, padecimiento de alguna enfermedad, si se ha sentido discriminado, uso de tecnologías, etc.

Con el propósito de poder realizar tanto el entrenamiento como las validaciones de los modelos predictivos se utilizarán como métodos de resampleo el método *hold-out* y el *k-fold-cross-validation*, de forma similar a lo realizado por Kleinbert et al (2017). Así, primero se dividirá la data aleatoriamente en tres conjuntos. Un primer conjunto contendrá el 65% de los datos y será utilizado para el entrenamiento de los modelos predictivos con un 5-fold-cross-validation, por lo que se dividirá a su vez en 5 grupos aleatorios, cada uno de los cuales se utilizará para validar el modelo predictivo generado por los otros 4. El segundo grupo contendrá el 15% de los datos y será utilizado como set de validación para evaluar la capacidad de predicción de los modelos para datos distintos a los utilizados en el entrenamiento. El tercer conjunto contendrá el 20% restante y será reservado hasta el momento de publicar los resultados y se utilizará como set de validación final.

Dado que uno de los objetivos del trabajo es hallar un modelo con el que pueda realizarse una predicción, se entrenará los datos con cuatro estrategias de estimación distintas y se comparará los resultados. Las estrategias a utilizar serán la regresión Ridge, la regresión LASSO, Random forest y Boosted trees.

Las regresiones Ridge y LASSO son métodos *shrinkage*, que reducen el valor de la varianza acercando el valor de los coeficientes de la regresión a cero ayudando así también a prevenir el *overfitting* (sobreajuste) del modelo predictivo a los datos de entrenamiento. La regresión Ridge reduce algunos coeficientes a valores cercanos a cero, mientras que la regresión LASSO iguala su valor a cero. Debido a ello, la regresión Ridge da mejores resultados cuando la variable a predecir sea el resultado de muchos predictores con pesos similares, y la regresión LASSO resulta más efectiva para variables que se predicen con un número reducido de predictores muy significativos (James et al, 2013, 224). Sin embargo, dado que la naturaleza de la relación entre los predictores y la variable a predecir nos es desconocida, se utilizarán ambos métodos.

Los métodos Random forest y Boosted trees derivan son el resultado de mejorar la capacidad predictiva del método de *decision trees* (árboles de decisión) con el bootstrapping y el boosting respectivamente. Los árboles de decisión implican la segmentación de los resultados en regiones generadas en base a los predictores; sin embargo, son deficientes en su capacidad predictiva porque tienden al *overfitting*. Para solucionar este impase, el método de Random Forest utiliza el bootstrapping para crear simultáneamente una cantidad significativa de árboles de decisión distintos en base a los datos de entrenamiento, y promediando los resultados, obteniendo así un modelo predictivo más consistente. Por otro lado, el método de boosted trees es el resultado de crear diversos árboles de decisión consecutivos, tal que cada árbol “aprende” del anterior, mejorando así gradualmente la capacidad de predicción del modelo. Por lo tanto, la diferencia entre random forest y boosted trees radica en que el primero es el resultado de promediar árboles simultáneos y el segundo de una mejora gradual (James et al, 2013, 321).

Si bien se hará uso de todos los métodos antes mencionados, tenemos especial interés en el método LASSO. Dado que este iguala diversos coeficientes a cero, permite hacer la selección de los predictores más relevantes, ayudando así a cumplir otro de los objetivos de este trabajo, que es el de reconocer las variables más importantes en la predicción.

**IV.- Bibliografía**

Banco Mundial. (2020). *Migrantes y Refugiados venezolanos en el Perú: el impacto de la crisis del Covid-19*. <http://documents1.worldbank.org/curated/en/647431591197541136/pdf/Migrantes-y-Refugiados-Venezolanos-en-El-Peru-El-Impacto-de-la-Crisis-del-Covid-19.pdf>

Carrasco, I., & Suárez, J. (2018). *Migración internacional e inclusión en América Latina*. CEPAL, Cooperación Alemana. <https://repositorio.cepal.org/bitstream/handle/11362/43947/S1800526_es.pdf?sequence=1&isAllowed=y>

Castles, S. (2014). Las fuerzas tras las migración global. *Revista Mexicana de Ciencias Políticas y Sociales*, *40*(220). <http://www.scielo.org.mx/pdf/rmcps/v59n220/v59n220a8.pdf>

Freier, F. (2020). Movilidad y políticas migratorias en América Latina en tiempos de COVID-19. En S. Castillo (Ed.), *Inmigración en tiempos de Covid-19* (pp. 50-68). CIBOD. <https://www.cidob.org/es/publicaciones/serie_de_publicacion/anuario_cidob_de_la_inmigracion/inmigracion_en_tiempos_de_covid_19_anuario_cidob_de_la_inmigracion_2020?fbclid=IwAR3gLnoaVeqimI75zZhTsEIYGYMMnDUT35ISByGe2vSniMNHaL8UBzBeuRY>

Freier, F., & Castillo, S. (2020). *Regional responses to Venezuela’s mass population displacement*. <https://www.e-ir.info/pdf/87681>

Gareth et al., J. (2017). *An Introduction to Statistical Learning with applications in R*. Springer. <https://static1.squarespace.com/static/5ff2adbe3fe4fe33db902812/t/601cc86d7f828c4792e0bcae/1612499080032/ISLR+Seventh+Printing.pdf>

Hernández-Vásquez et al., A. (2019). Factores asociados a la no utilización de servicios de Salud en inmigrantes venezolanos en Perú. *Rev Peru Med Exp Salud Publica*, *36*(4). <https://www.scielosp.org/pdf/rpmesp/2019.v36n4/583-591/es>

INEI. (2019). *Condiciones de vida de la población venezolana que reside en Perú*. <https://r4v.info/en/documents/details/70521>