

Universidad del Rosario

Maestría en matemática Aplicada y Ciencias de la Computación

Machine Learning

1 semestre

Estudiante: Hector Leandro Rojas Serrano

Índice Global de Felicidad

Objetivo

Determinar los principales factores que contribuyen a la felicidad global.

Para determinar el país más feliz del mundo, los investigadores analizaron datos completos de encuestas de Gallup de 149 países durante los últimos tres años, monitoreando específicamente el desempeño en seis categorías particulares: producto interno bruto per cápita, apoyo social, esperanza de vida saludable, libertad para tomar sus propias decisiones de vida., generosidad de la población en general y percepciones de los niveles de corrupción interna y externa. (World Happiness Report 2022).

Para este proyecto como propuesta vamos a incluir 3 nuevas variables de desarrollo mundial y elementos del índice de Desarrollo Humano y evaluaremos su efecto en el índice Global de Felicidad, esta nueva propuesta tratara de aumentar y explicar de una mejor forma el comportamiento de dicho índice.

Triandis(2010) define factores que influyen en el bienestar social o felicidad ya que están altamente relacionados, entre ellos tenemos la salud, las libertades políticas, eficiencia de las instituciones públicas, el GNI, entre otros, de forma negativa podemos mencionar, los conflictos internacionales, gobiernos no democráticos, desastres naturales, desempleo, deficiencias en salud.

Variables- Generalidades:

Datasets

Para construir este modelo utilizamos 3 diferentes fuentes de información, dentro de ellas tenemos:

1.**Happiness score:** es la variable que pretendemos explicar por medio de modelos de regresión.extraido de : https://happiness-report.s3.amazonaws.com/2022/Appendix_2_Data_for_Figure_2.1.xls

2.**Unemployment, total (% of total labor force):** se obtuvo de <https://databank.worldbank.org/source/world-development-indicators#> Datos para 2021

3.**Political stability and absence of violence:** se obtuvo de: <https://databank.worldbank.org/source/worldwide-governance-indicators#> datos para 2021, este indicador mide las percepciones de la probabilidad de que el gobierno sea desestabilizado o derrocado por medios inconstitucionales o violentos, incluida la violencia y el terrorismo por motivos políticos.

4.**Mean years of schooling:** data extraída del índice de desarrollo humano Human Development Report 2021-22 https://hdr.undp.org/sites/default/files/2021-22_HDR/HDR21-

[22 Statistical Annex HDI Table.xlsx](#) número promedio de años de educación recibidos por personas de 25 años o más, convertidos a partir de los niveles de educación alcanzados utilizando las duraciones oficiales de cada nivel.

5. **Gross national income (GNI) per cápita:** data extraída del índice de desarrollo humano Human Development Report 2021-22 https://hdr.undp.org/sites/default/files/2021-22_HDR/HDR21-22_Statistical_Annex_HDI_Table.xlsx

6. **Control of corruption:** se obtuvo de <https://databank.worldbank.org/source/worldwide-governance-indicators#> datos para 2021.

Construcción del dataset

Se realizó una combinación de los datos mencionados en la sección anterior para crear un dataset para los países analizados y poder explicar y entender el comportamiento del índice global de felicidad. Las secciones son:

1. Datos del World Happiness Report 2022. (variable de interés)
2. Indicadores del banco Mundial (Tasa de desempleo, control de corrupción, expectativa de vida al nacer, percepción de estabilidad política y ausencia de violencia).
3. Años promedio de estudio e Ingreso nacional bruto (INB) per cápita del índice de desarrollo humano 2021-2022.

Variables descripción

Gross national income (GNI) per capita: Ingreso agregado de una economía generada por su producción y su propiedad de factores de producción, menos los ingresos pagados por el uso de factores de producción propiedad del resto del mundo, convertidos a dólares internacionales utilizando tasas de PPA, dividido por mitad de año población.

Life expectancy at birth: Número de años que tiene un recién nacido podría esperar vivir si los patrones prevalecientes de las tasas de mortalidad específicas por edad en el momento del nacimiento se mantuvieran iguales durante toda la vida del infante.

Control of Corruption: El Control de la Corrupción capta las percepciones de la medida en que el poder público se ejerce para beneficio privado, incluidas formas de corrupción menores y mayores, así como la "captura" del Estado por parte de élites e intereses privados. La estimación da la puntuación del país en el indicador agregado, en unidades de una distribución normal estándar, es decir, que van desde aproximadamente -2,5 a 2,5.

Political Stability and Absence of Violence/Terrorism: mide las percepciones de la probabilidad de inestabilidad política y/o violencia políticamente motivada, incluido el terrorismo. La estimación da la puntuación del país en el indicador agregado, en unidades de una distribución normal estándar, es decir, que van desde aproximadamente -2,5 a 2,5.

Unemployment, total (% of total labor force): Es la tasa de desempleo de la proporción de la fuerza laboral que no tiene trabajo pero que está disponible y buscando empleo.

Mean years of schooling: Promedio de años de educación recibidos por personas de 25 años o más, convertidos a partir de los niveles de educación alcanzados utilizando las duraciones oficiales de cada nivel.

Metodología:

1. Se realiza un análisis exploratorio de los datos
2. Se realizará modelos de regresión para predecir el comportamiento de la felicidad en los países con las variables mencionadas. Los modelos a utilizar serán (linnear Regression, Ridge CV y LassoCV. (De manera adicional se ejecuto el modelo OLS para contrastar lainformación y los estimadores)

Análisis exploratorio de datos

Matriz de correlación:

En la matriz de correlación se puede observar la relación negativa que existe entre la tasa de desempleo y el índice de felicidad, para las otras variables la relación es positiva o directa.

Boxplot:

De los diagramas de caja anteriores podemos concluir que hay valores atípicos en la variable de la tasa de desempleo y el ingreso bruto per cápita.

Histogramas:

De los histogramas podemos comentar que en su mayoría están sesgados hacia la derecha o hacia la izquierda, en cuanto a la tasa de desempleo su sesgo es más pronunciado a la izquierda y presenta pocos datos por encima del 20%. esta misma tendencia la tiene el ingreso bruto per cápita, respecto a la variable de interés o índice global de necesidad podemos afirmar que sigue una distribución normal con un pequeño sesgo a la derecha

Pairplot:

Siguiendo la teoría y varios autores podemos evidenciar en este gráfico bivariado de dispersión que para la variable desempleo se observa una relación contraria con el GNI donde cuando este último aumenta la felicidad aumenta, pero el desempleo tiende a la baja (5 grafica de la primera columna de arriba hacia abajo), existe por tanto una relación con los niveles de felicidad y el nivel de empleo con los ingresos per cápita.

Modelos de Regresión y Resultados

Linnear Regression

	predictor	coef Opt	coef Opt Train
0	GNI	0.000017	0.000015
1	Lifexpectancy	0.038346	0.040758
2	ControlofCorruption	-0.022178	-0.035837
3	PoliticalStability	0.167778	0.208002
4	Unemployment	-0.022996	-0.027458
5	Meanyearsofschooling	0.060292	0.071585

MCO Score o R2: 0.6837315084596044

MCO Train Score o R2: 0.670690951806701

MCO Test Score o R2: 0.7314263129519094

El error (rmse) para MCO es: 0.6123037409240174

El error (rmse) en Training para MCO es: 0.6385951224981687

El error (rmse) en Testing para MCO es: 0.5081308534410823

Ridge CV

	coef Opt Train
0	0.247249
1	0.263066
2	0.077834
3	0.118592
4	-0.113527
5	0.183961

Ridge Training Score 0.6626343581659409

Ridge Test Score 0.7517042539923952

El error (rmse) en Training para Ridge es: 0.5847220581089185

El error (rmse) en Testing para Ridge es: 0.48270084240838995

Lass

coef Opt train	
0	0.000039
1	0.000000
2	0.000000
3	0.000000
4	-0.000000
5	0.000000

Lasso Training Score -21.90362037908093

Lasso Test Score -22.613826779719748

El error (rmse) en Training para Lasso es: 4.817826104759959

El error (rmse) en Testing para Lasso es: 4.707351232925361

Conclusiones

1. Por sus resultados, el modelo de regresión que brinda mejores resultados es el modelo de Ridge CV ya que tiene el menor RMSE entre los demás modelos.
2. En cuanto a la felicidad de los países se observa que en gran parte los factores económicos y de riesgo país influyen positivamente en dicho índice.
3. Podemos expandir este estudio incluyendo más variables socioeconómicas que nos permitirán mejorar la efectividad de políticas públicas, inversiones, programas sociales, entre otros.