

Datos con perspectiva de género en el área STEM

1th María Beatriz Sánchez Díaz

Facultad de Ingeniería, UNAM

Ciudad de México, México

bety.sanchez.diaz@gmail.com

Abstract—La brecha de género es una condición existente desde hace mucho tiempo y se ve reflejada en diferentes aspectos como oportunidades de acceso, recursos económicos, oportunidades laborales, situaciones políticas, sociales, culturales, etc.

De acuerdo con la Comisión Económica para América Latina (CEPAL) algunas condiciones que existen a causa del género son: sobrecarga de trabajo no remunerado, menor participación laboral, menores niveles de ingreso, brechas salariales, menor participación en empleos formales, entre otras.

Existen diferentes indicadores y fuentes para analizar esta situación en México, una de las fuentes es la proporcionada por el INEGI: la Encuesta Nacional de Ocupación y Empleo (ENOE) la cual es considerada la principal fuente de información sobre mercado laboral en México, ofrece datos mensuales y trimestrales sobre empleabilidad para las 32 entidades federativas.

Actualmente gracias a la existencia de los datos abiertos es posible analizar y obtener patrones y conocimiento de los datos que se proporcionan con el fin de entender de una manera más amplia la situación y encontrar algún valor en los datos y que esto pueda ayudar a solucionar el problema identificado. En este documento se visualizará como por el género existen grandes diferencias desde el salario hasta el número de personas que se dedican a una profesión en el área STEM y se verá reflejado que existe aproximadamente el doble de probabilidad de que alguien del género masculino entre en estas áreas a una mujer. La razón de hacer más visible esta situación es que para poder abordar y generar acciones lo primero es poder identificar y tener datos de lo que ocurre y en donde y como se manifiesta la desigualdad de género, este es uno de los objetivos principales en esta investigación.

Key words: Clasificación, Empleo, Brecha, género.

I. INTRODUCCIÓN

La tecnología y la ciencia está teniendo un gran impacto en la vida de todas las personas y se ha convertido en uno de los recursos con más consumo en la actualidad aparte de que tienen efectos positivos en diferentes áreas como el crecimiento económico y las condiciones de vida, a lo largo de los años estas áreas constituyen un pilar en el desarrollo cultural, social y económico.

Uno de los impactos importantes que tiene es el surgimiento de nuevas áreas laborales pues con su crecimiento

se necesitan a más especialistas sin embargo, ¿Qué tan grande es la brecha de género al hablar de oportunidades laborales en el área STEM?

De acuerdo con datos de la AMIIF en México el 38 por ciento de las mujeres estudian carreras STEM, otros estudios publicados por la UNAM demuestran que en carreras como Matemáticas e Ingenierías solo eran el 19 por ciento de mujeres. Estos son solo alguno de los datos que demuestran que a pesar de que la población en México es mayormente constituida por mujeres no tienen mucha presencia en estas áreas.

En cuanto al campo laboral, se sabe que los puestos de trabajo relacionados con la ciencia y tecnología son los mejores pagados y siguen en crecimiento, el hecho de que la participación de la mujer en esta área sea muy poca significa que existen grandes diferencias salariales, otra cuestión que también surge es que aunado a que la probabilidad de que la mujer pueda tener un rol en el sector STEM una vez que lo logran suelen tener salarios menores a los que tiene el género masculino teniendo el mismo cargo, esto lleva a una discriminación de género salarial.

A lo largo de este documento se buscará analizar datos sobre empleabilidad y Ocupaciones en México esto con el fin de visualizar la brecha de género que existe en empleo y salario en el área STEM (Ciencia, Tecnología, Ingeniería y Matemáticas).

II. ANTECEDENTES

México es uno de los primeros países que se encuentran en el Índice Global de la Brecha de Género ocupando el lugar 25 con 75.4 por ciento y a lo largo de la historia esta brecha de género ha sido visible en diferentes aspectos, algunos más visibles que otros pero siempre ha estado presente y a pesar de que en los últimos 50 años se ha logrado un gran avance aún hay muchas acciones que tomar para lograr una mayor igualdad de género en México y el mundo, tan solo en Latinoamérica de acuerdo con el Foro Económico Mundial, es una de las

regiones con mayor desigualdad de género, alcanzando el 70 por ciento de brecha para alcanzar la igualdad de género.

De cada 10 trabajadores entre 25 y 54 años de edad en América Latina, solo cuatro son mujeres, y cabe mencionar que la diferencia no solo se ve en el número de trabajadores sino en el tipo de empleo, horas laborales y salarios promedio.

III. MATERIAL Y MÉTODOS

Esta investigación se desarrollo principalmente con el objetivo de hacer visible lo que para muchos es invisible: la brecha de género en empleos relacionados con el área STEM, de igual forma se busca encontrar que en efecto el género influye en esta situación y se planteron algunas preguntas como.

¿Qué tanto influye el género en la participación de la mujer en el área STEM? ¿Qué otras variables afectan esta situación? ¿En dónde se debe poner enfoque para acabar/mejorar la desigualdad de género?

Se siguió el siguiente pipeline para el desarrollo del proyecto:



Fig. 1. Pipeline

- ETAPA 1 - Extracción, limpieza y transformación de datos: para esta etapa se hizo uso de una herramienta dinámica que proporciona DataMéxico Tesseract UI este es un explorador de API's que ofrece acceso a diferentes Bases de Datos y permite hacer consultas dinámicas de los datos y extraerlos.

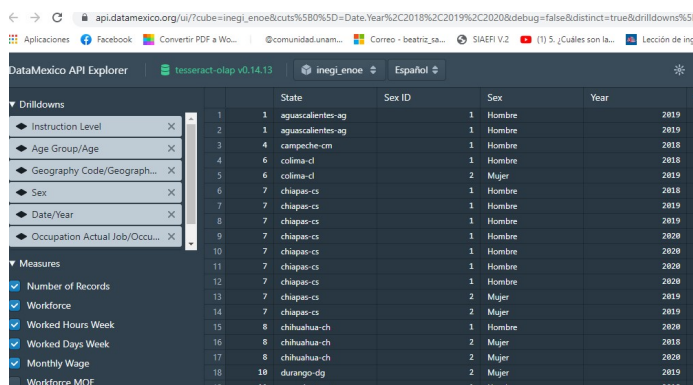


Fig. 2. Tesseract UI

Dentro de todas las bases disponibles la de interés fue la enoe, la cual es la principal fuente de información sobre

empleabilidad de las 32 entidades federativas en México y se puede filtrar por entidad, género, ocupación, situación laboral entre otros.

DATASET: El dataset extraído para el análisis del proyecto cuenta con las siguientes variables, el dataframe cuenta con 642601 rows × 15 columns:

Datos de los últimos tres años: 2018, 2019, 2020.

Estados: 32 entidades federativas.

Características:

- Instruction Level :Indica los niveles de escolaridad, posee once opciones.
- Age:Indica la edad
- State: Indica el estado
- Sex : Indica el género, posee dos opciones (femenino y masculino).
- Year: Indica la temporalidad de los datos
- Subgroup: Indica el sector de la ocupación
- Workforce: Indica la fuerza laboral.
- Worked Hours Week: Indica las horas trabajadas por semana.
- Worked Days Week: Indica los días trabajados por semana.
- Monthly Wage: Indica el salario promedio mensual en pesos mexicanos.

	Instruction Level ID	Instruction Level	Age	State ID	State	Sex ID	Sex	Year	Subgroup ID	
1	0	Ninguno	15	1	Aguascalientes	1	Hombre	2019	971	Trabajadores de P
2	0	Ninguno	15	4	Campeche	1	Hombre	2018	963	Lavi
3	0	Ninguno	15	6	Colima	1	Hombre	2018	963	Lavi
5	0	Ninguno	15	7	Chiapas	1	Hombre	2018	713	Yeseros, instala
6	0	Ninguno	15	7	Chiapas	1	Hombre	2019	611	Trabi
...
814867	99	No Sabe Nivel	98	28	Tamaulipas	1	Hombre	2020	511	Trabajadores e
814869	99	No Sabe Nivel	98	30	Veracruz de Ignacio de la Llave	1	Hombre	2018	262	Auxiliares y Té
814872	99	No Sabe Nivel	98	30	Veracruz de Ignacio de la Llave	1	Hombre	2019	112	P
814873	99	No Sabe Nivel	98	30	Veracruz de Ignacio de la Llave	1	Hombre	2019	152	Coordinadores
814875	99	No Sabe Nivel	98	30	Veracruz de Ignacio de la Llave	2	Mujer	2018	961	

Fig. 3. Dataframe

Limpieza y Transaformación: posterior a la extracción de los datos se hizo una transformación y eliminación de datos faltantes con python.

- ETAPA 2 - Análisis exploratorio de los Datos y visualizaciones: el principal objetivo de este paso es visualizar de forma intectiva los datos que se ofrecen, obtener algunos insights y hacerlo de forma interactiva por lo que se decidió utilizar Power BI el cual es una herramienta de análisis de datos orientado justamente a proporcionar visualizaciones interactivas.
- ETAPA 3 - Implementación de algoritmos de ML: en esta etapa se busca generar algún modelo de ML que

clasifique en área STEM u OTRA de acuerdo a las características proporcionadas dentro de las cuales se incluye el género, para esta clasificación se usó la clasificación por Bosques Aleatorios, este es un algoritmo de aprendizaje automático muy utilizado que consiste en combinar múltiples árboles de decisión hasta poder obtener alguna predicción más precisa, su implementación suele ser muy sencilla por lo que es fácil generar buenos modelos.

Se caracteriza por ser uno de los modelos con buen desempeño en sus predicciones y clasificaciones, también es muy útil para evitar el sobreajuste de modelos como en caso de los árboles de decisión y a pesar de contar con algunas desventajas como el tiempo computacional o la dificultad de interpretación siguen siendo más efectivos.



Fig. 4. Representación de Random Forest

IV. RESULTADOS

A continuación se mostrarán los resultados en cada uno de las etapas, detallando un poco acerca de la situación en la que México se encuentra respecto a la brecha de género.

Recordando, los datos utilizados son de los años 2018-2020 en México obtenidos de la enoe sobre empleabilidad, se clasificaron los subgrupos de ocupación aquellos que se relacionaran con el área STEM, seleccionando los siguientes:

- 'Auxiliares y Técnicos en Ciencias Físicas, Matemáticas, Biológicas, Químicas, del Medio Ambiente y Agronómicas',
- 'Técnicos Eléctricos, en Electrónica y de Equipos en Telecomunicaciones y Electromecánicos',
- 'Directores y Gerentes en Informática, Telecomunicaciones, Transporte y en Investigación y Desarrollo Tecnológico',
- 'Investigadores y Especialistas en Sistemas Computacionales',
- 'Investigadores y Especialistas en Ciencias Sociales',

- 'Ingenieros Químicos, Mecánicos, Industriales, Mineros y Metalúrgicos',
- 'Ingenieros Eléctricos y en Electrónica',
- 'Médicos Generales y Especialistas',
- 'Investigadores y Especialistas en Ciencias Biológicas, Químicas y del Medio Ambiente',
- 'Ingenieros Civiles, Topógrafos y Arquitectos',
- 'Especialistas en Ciencias Agronómicas',
- 'Ingenieros en Comunicaciones y Telecomunicaciones',
- 'Investigadores y Especialistas en Física, Matemáticas, Estadística y Actuaría',
- 'Coordinadores y Jefes de Área en Informática, Telecomunicaciones, Transporte y en Investigación y Desarrollo Tecnológico'

VISUALIZACIÓN DE DATOS

La etapa de visualización y análisis exploratorio fue de las más relevantes pues sirvió para encontrar ciertos patrones y hallazgos.

Se generaron 3 dashboards en Power BI con diferentes perspectivas, el primero tiene un enfoque mayor en mostrar la presencia de la mujer en cada uno de los estados de la república.

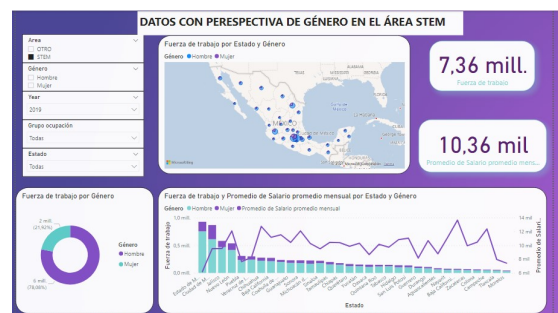


Fig. 5. Dashboard 1

En resumen lo que se pudo visualizar de este primer dashboard fue lo siguiente:

* Para los tres años analizados se encontró que los estados con mayor porcentaje de participación femenina en ciencia y tecnología son: Guadalajara, Estado de México, Ciudad de México y Nuevo León, sin embargo, el porcentaje de hombres sigue dominando el área.



Fig. 6. Mapa fuerza de trabajo por estado

De manera más visual y observando de mejor manera la brecha que existe entre el género femenino y masculino se puede observar en la siguiente visualización donde claramente vemos que en ninguno de los casos la mujer domina o está en una misma situación que el hombre, en la gráfica también se pueden ver los sueldos promedios mensuales que si los filtramos por género se obtienen pequeñas diferencias en ellos.

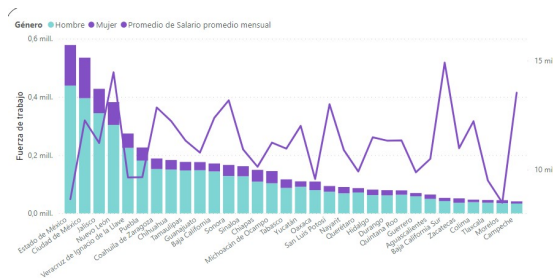


Fig. 7. Fuerza de trabajo vs Estado

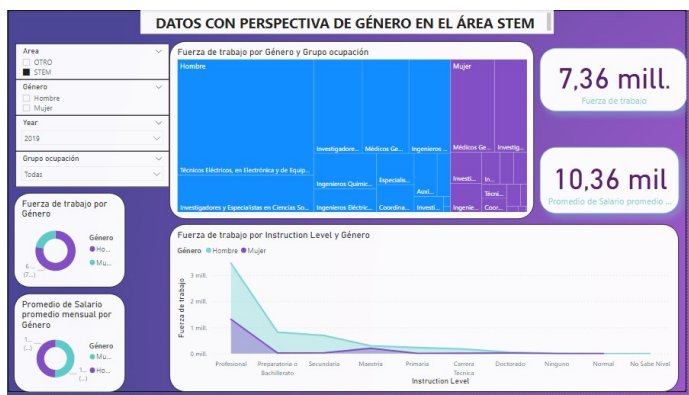


Fig. 8. Dashboard 2

El segundo dashboard tiene un enfoque hacia las ocupaciones y nivel de estudios, de este se obtuvieron los siguientes hallazgos: La mayor fuerza de trabajo en el área STEM se encuentra ocupada por el género masculino, mientras que las ocupaciones más ocupadas por las mujeres son los empleos informales siendo así que el empleo más ocupado por la mujer

pertenece al grupo de "Trabajadoras domésticas" y cualquiera de las áreas STEM la mujer tiene una participación mínima dando como resultado que tan solo el 2.16 por ciento de mujeres tiene participación en estas áreas.

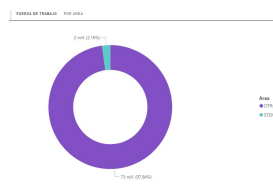


Fig. 9. Gráfico de anillos

En cuanto al nivel académico que estudiaron las personas para poder dedicarse a una profesión dentro del área, de parte del género femenino la mayor parte estudiaron una carrera profesional o maestría mientras que el género masculino estudió una carrera profesional, preparatoria o secundaria.

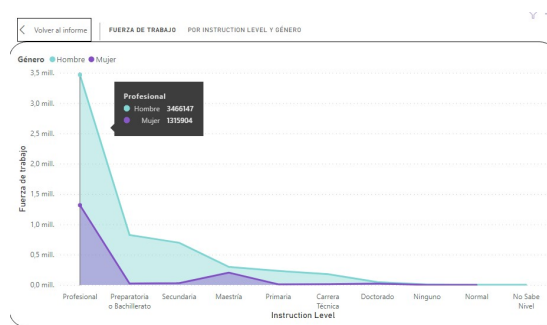


Fig. 10. Fuerza de trabajo por nivel de educación y género

A manera de resumen en la siguiente visualización se puede ver que la mayor parte de la fuerza laboral se encuentra dominada por los hombres mientras que las mujeres ocupan un porcentaje menor.



Fig. 11. Fuerza de trabajo por género y ocupación

Finalmente con el uso de una herramienta que proporciona Power BI se buscaron los principales elementos que influyen en que el área a la que se dediquen se encuentre en STEM.

De esta visualización se obtuvo lo siguiente:

* Cuando el nivel de educación es Profesional hay 5.82 más veces de probabilidad de que el área sea STEM a comparación de los otros valores.

* Cuando el género es hombre hay 2.16 más de probabilidad de que el área sea STEM a comparación de los otros valores.

* Cuando el estado es Nuevo León hay 1.23 más veces de probabilidad de que el área sea STEM a comparación de los otros valores y 1.19 más en la Ciudad de México.

Con esto se puede ver que efectivamente el género tiene importancia sin embargo, existen otros factores a tomar en cuenta.

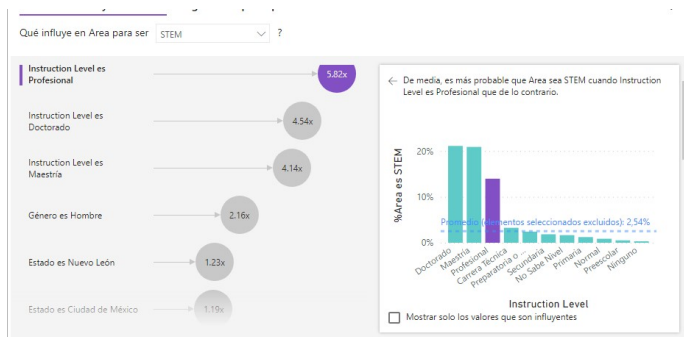


Fig. 12. Elementos influyentes

MODELO DE CLASIFICACIÓN

Con la aplicación de árboles aleatorios para la clasificación se busca que a partir de las características que se tienen como: Nivel de Educación, Edad, Estado y Género se pueda clasificar a que área es más probable que se dedique.

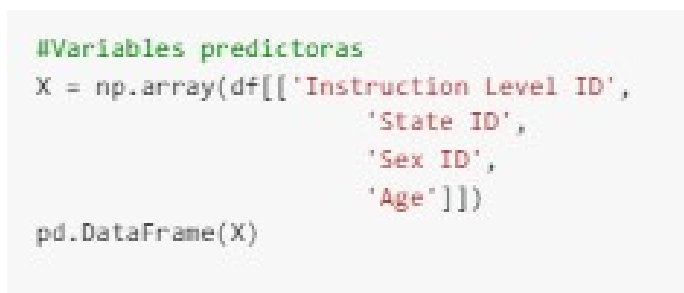


Fig. 13. Resultados del modelo

Se implementó el modelo donde las variables predictorias son las anteriormente mencionadas y la variable a predecir/variable clase es el área (STEM u OTRO).

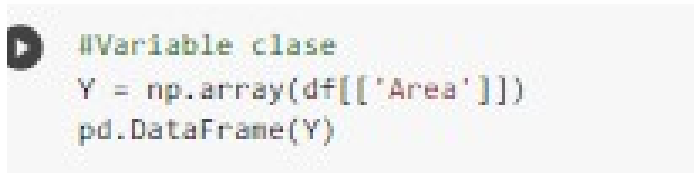


Fig. 14. Resultados del modelo

La exactitud promedio de la validación fue del 94.81 por ciento donde se obtuvo la siguiente matriz de clasificación. También se generó la importancia de cada una de las variables en el resultado teniendo como más relevancia el nivel de educación.

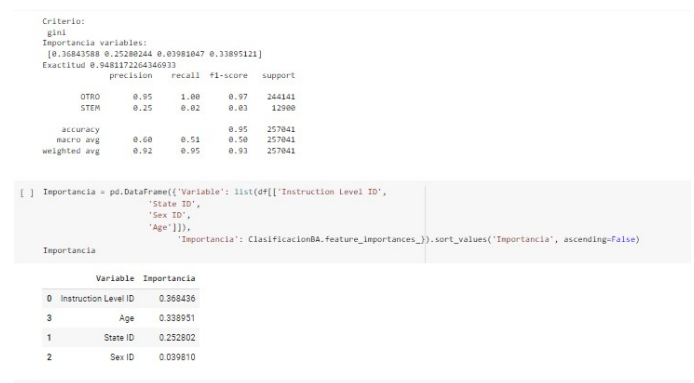


Fig. 15. Resultados del modelo

Este modelo ayuda principalmente a poder predecir por características la probabilidad de estar laboralmente en área STEM. Puede ayudar a identificar factores de riesgo y de esta manera proponer soluciones.

V. CONCLUSIONES

Sin duda alguna es de suma importancia comenzar desde un análisis profundo sobre los datos con los que se cuentan para entender cualquier problema, en este artículo específicamente sobre la brecha de género, esto con el fin de tener un panorama más amplio partir de ahí poder atacar el problema. Esta investigación muestra de una forma visual las diferencias entre los dos géneros en cuanto a la fuerza de trabajo, rango salarial y nivel de educación dando porcentajes que deberían ser tomados en cuenta por cada uno de los estados de la República e implementar nuevos programas para la integración de la mujer en el área STEM. También se presentó la integración de un modelo de Machine Learning que nos permitió ver que

factores toman más importancia a la hora de la clasificación y se puede concluir que a pesar de que en efecto el género influye hay otros factores a tomar en cuenta aunado a que se necesitan aún más factores para poder tener un mejor resultado en el modelo pero sin duda esto puede ayudar a saber que acciones tomar de acuerdo a la importancia de cada característica y saber que mejorar.

VI. TRABAJO FUTURO

El objetivo de esta investigación es poder crear una fuente de información sobre la situación de la desigualdad de género en áreas como la tecnología, la ciencia, ingeniería y matemáticas por lo que una mejora a futuro es crear una aplicación web donde cada uno de los estados pueda ver la brecha de género que existe y tomar acciones para erradicar el problema e inclusive después también enfocarlo en empresas pues la brecha de género en el área laboral es algo en lo que cada persona, estado y empresa puede aportar para mejorar.

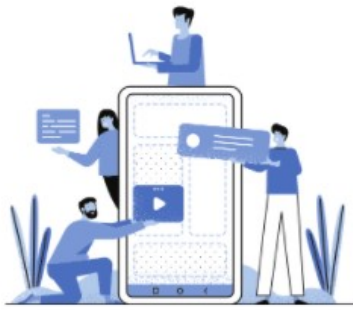


Fig. 16. Resultados del modelo

Otro punto a mejorar es la exactitud del modelo aplicado pues el incluir este tipo de tecnologías en investigaciones con temas sociales suma bastante y poder crear y mejorar el modelo es relevante para que pueda ser tomado como punto de referencia en diferentes aspectos. Otro punto que se quiere tomar en cuenta para un trabajo futuro es la integración de modelos de predicción a lo largo del tiempo, esto con el fin de saber como puede ir mejorando o empeorando la situación si las variables cambian.

REFERENCIAS

- [1] Marchionni, M., *BRECHAS DE GÉNERO EN AMÉRICA LATINA*, Banco de Desarrollo de América Latina, Accedido el 09 de Diciembre, 2021. [En línea]. Disponible en: <https://scioteca.caf.com/bitstream/handle/123456789/1401/Brechas%20de%20genero%20en%20America%20Latina.%20Un%20estado%20de%20situacion.pdf>
- [2] Eternod M., *“Brechas de género, INMUJERES, Agosto 2018*. Accedido el 09 de Diciembre, 2021. . [En línea]. Disponible en: https://crpd.cepal.org/3/sites/crpd3/files/presentations/panel2_marcelaeternod.pdf
- [3] Gonzalese L., *Bosques Aleatorios Clasificación*, APREEN-DEIA, 2019. Accedido el 06 de Diciembre, 2021. . [En línea]. Disponible en: <https://aprendeia.com/bosques-aleatorios-clasificacion-teoria-machine-learning/>
- [4] Gonzalese L., *Bosques Aleatorios Clasificación*, APREEN-DEIA, 2019. Accedido el 06 de Diciembre, 2021. . [En línea]. Disponible en: <https://aprendeia.com/bosques-aleatorios-clasificacion-teoria-machine-learning/>
- [5] México - *Índice Global de la Brecha de Género*, datosmacro, 2020. Accedido el 06 de Diciembre, 2021. . [En línea]. Disponible en: <https://datosmacro.expansion.com/demografia/indice-brecha-genero-global/mexico/>
- [6] DATAMEXICO, DATAMEXICO. Accedido el 03 de Diciembre, 2021. . [En línea]. Disponible en: <https://datamexico.org/es/profile/geo/mexico>
- [7] *Random Forest (Bosque Aleatorio): combinando árboles*, IArtificial.net. Accedido el 04 de Diciembre, 2021. . [En línea]. Disponible en: <https://www.iartificial.net/random-forest-bosque-aleatorio/>
- [8] Molero G. (2021) *Bosques Aleatorios*, Minería de Datos, UNAM. Accedido el 02 de Diciembre, 2021.