

**Proyecto Final Mental Health**



***Buscando La salud mental en nosotros***

Marcos Capponi

**INDICE**

**Planteo de la búsqueda del proyecto………………………………………………………………………………..2**

**Objetivo……………………………………………………………………………………………………………………………2**

**Descripción del proyecto………………………………………………………………………………………………….2**

**Machine learning……………………………………………………………………………………………………………..6**

**Conclusión……………………………………………………………………………………………………………………….7**



**Planteo de la búsqueda del proyecto**

En la búsqueda de una mejora en la salud mental de los trabajadores y con las nuevas propuestas de la actualidad como el trabajo en formato remoto o hibrido, buscaremos en base a la participación de 5.000 personas con diferentes trabajos en diferentes rubros, si el trabajar desde tu casa o de manera hibrida es beneficioso a la salud mental o no.

Con la aparición de la pandemia se comenzaron a explotar estas formas de trabajar para evitar la propagación y el contagio del virus, pero también se empezó a estudiar si esta metodología no era mejor para la gente. Si al estar trabajando desde la comodidad de tu hogar rendís mejor o te sentís más a gusto, o si te hace ser menos eficiente que estar en tu natural lugar de trabajo.

**Objetivo**

El objetivo general es que los algoritmos nos puedan predecir como clasifica la maquina las elecciones de la gente ante su metodología de trabajo y de ahí poder saber si esta metodología remota o hibrida son beneficiosas o no a la salud mental.

Particularmente queremos aprender del dataset y poder crear a traves del machine learning y de algoritmos de clasificación, una herramienta para poder predecir en base a datos que nos ofrecen personas, si su forma de trabajar, remota, hibrida o presencial, son beneficiosas o no para la salud mental.

**Descripción del proyecto**

En este caso usaremos un dataset que sacamos de la página kaggle donde, donde 5.000 personas participaron. Las columnas que contiene este dataset son las siguientes:

1. Employee\_ID
2. Age
3. Gender
4. Job\_role
5. Industry
6. Years\_of\_Experience
7. Work\_Location
8. Hours\_Worked\_Per\_Week
9. Number\_of\_Virtual\_Meetings
10. Stress\_Level
11. Mental\_Health\_Condition
12. Access\_to\_Menetal\_Health\_Resources
13. Productivity\_Change
14. Social\_Isolation\_Rating
15. Satisfaction\_with\_Remote\_Work
16. Company\_Support\_for\_Remote\_Work
17. Physical\_Activity
18. Sleep\_Quality
19. Region

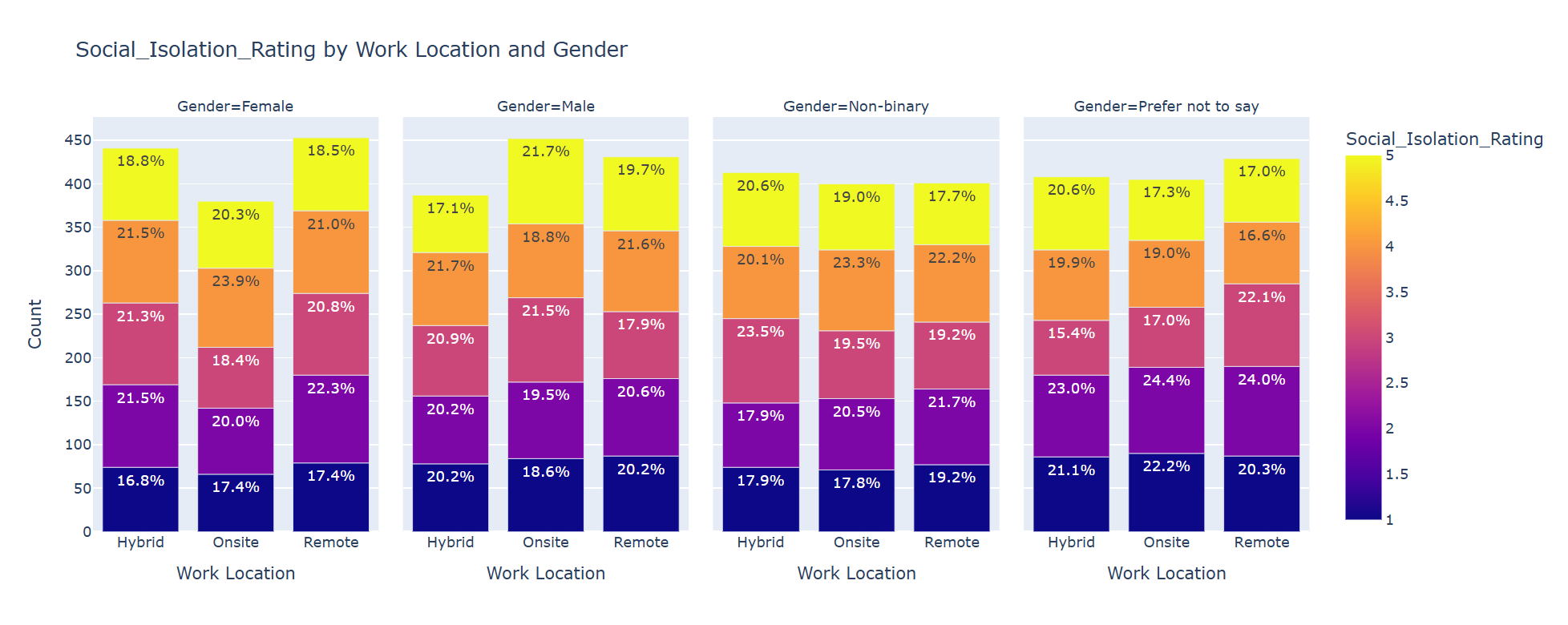
De todas estas columnas las mas relevantes para nuestro trabajo son: Work\_Location, Stress\_Level, Mental\_Health\_Condition, Access\_to\_Mental\_Health\_Resources, Productivity\_Change, Social\_Isolation\_Rating, Satisfaction\_with\_Remote\_Work, Company\_Support\_for\_Remote\_Work, Physical\_Activity, Work\_Life\_Balance\_Rating y Sleep\_Quality,.

A este dataset le agregaremos finalmente una columna llamada work\_benefit, que indica en base a los registros de las columnas, Satisfaction\_with\_Remote\_Work, Productivity\_Change, Stress\_Level, Work\_Life\_Balance\_Rating, Mental\_Health\_Condition, Social\_Isolation\_Rating, si su metodologia de trabajo es beneficiosa, neutral o no beneficiosa.

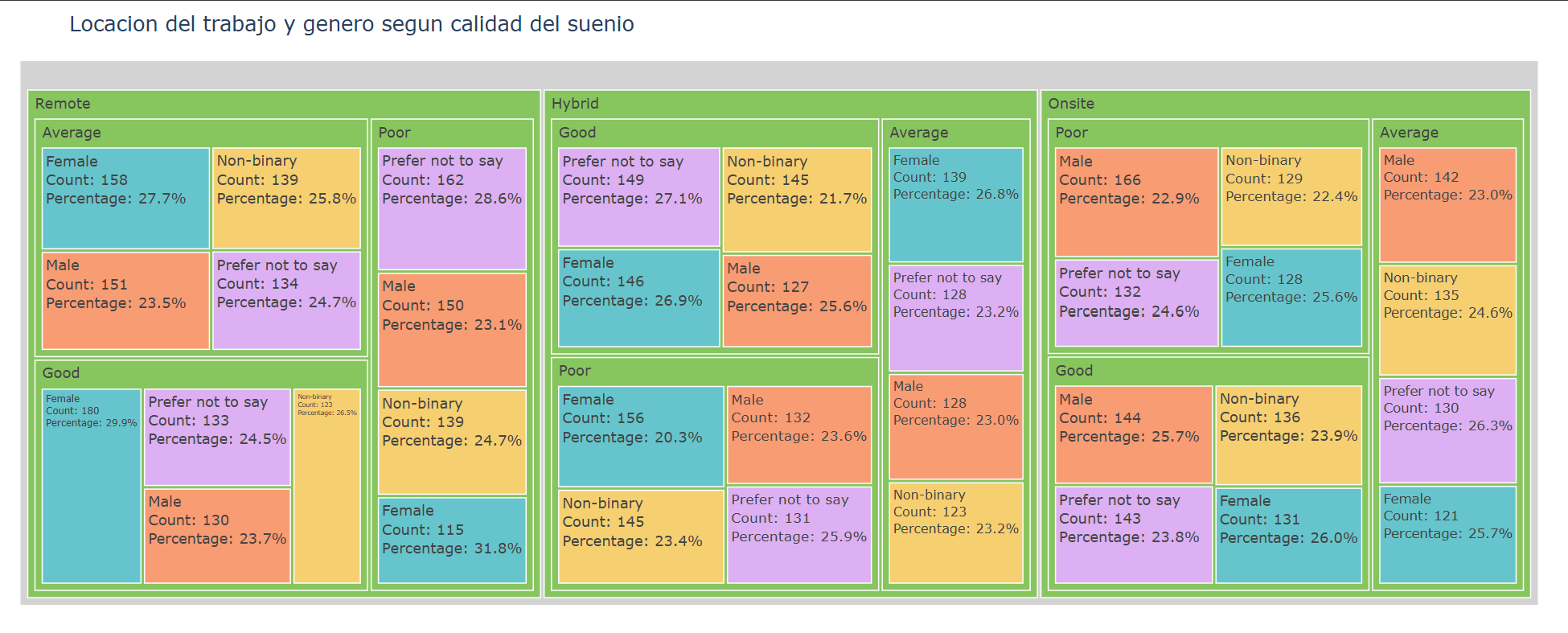
Haremos un análisis exploratorio de datos donde mostraremos a través de gráficos según el género y la locación del trabajo, un análisis de las columnas, Work\_Life\_Balance\_Rating, Job\_Role, Industry, Stress\_Level, Mental\_Health\_Condition, Access\_to\_Mental\_Health\_Resources, Productivity\_Change, Social\_Isolation\_Rating, Satisfaction\_with\_Remote\_Work, Company\_Support\_for\_Remote\_Work, Physical\_Activity, Sleep\_Quality.

Luego de este análisis haremos una primera conclusión y luego usaremos el maching learning para poder predecir si los empleados están satisfechos con el método de trabajo o no. Usaremos algoritmos de clasificación y de ahí sacaremos las conclusiones finales.

Aquí mostraremos algunos gráficos a modo de ejemplo.

****

****



**Machine Learning**

A este trabajo le aplicaremos algoritmos de clasificación con la variable dependiente a predecir, la variable work\_benefit.

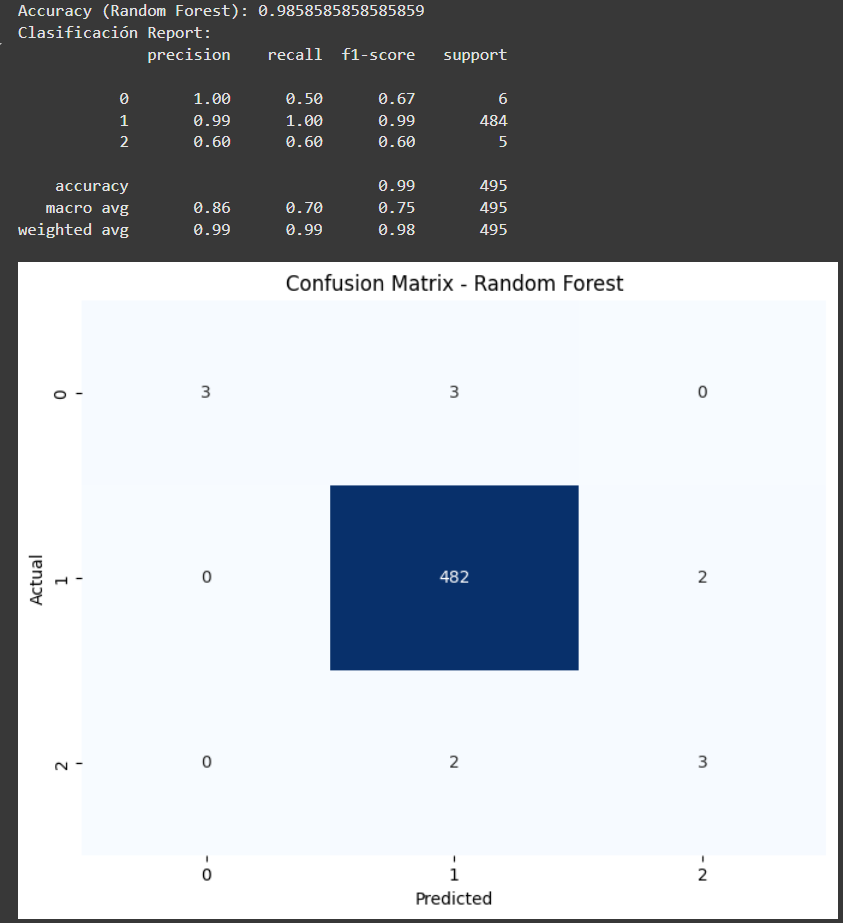
Usaremos los algoritmos:

1. Logistic Regression
2. Decision Tree
3. Random Forest
4. SVM
5. Gradient Boost Machine

Haremos una función para mostrar clasification report y un heatmap para graficar los resultados de cada algoritmo.

Tambien usaremos la técnica GridSearchCV para ajustar hiperparametros y poder usar la mejor combinación de datos para obtener un mayor rendimiento del algoritmo y los datos.

Aquí debajo dejaremos un ejemplo.



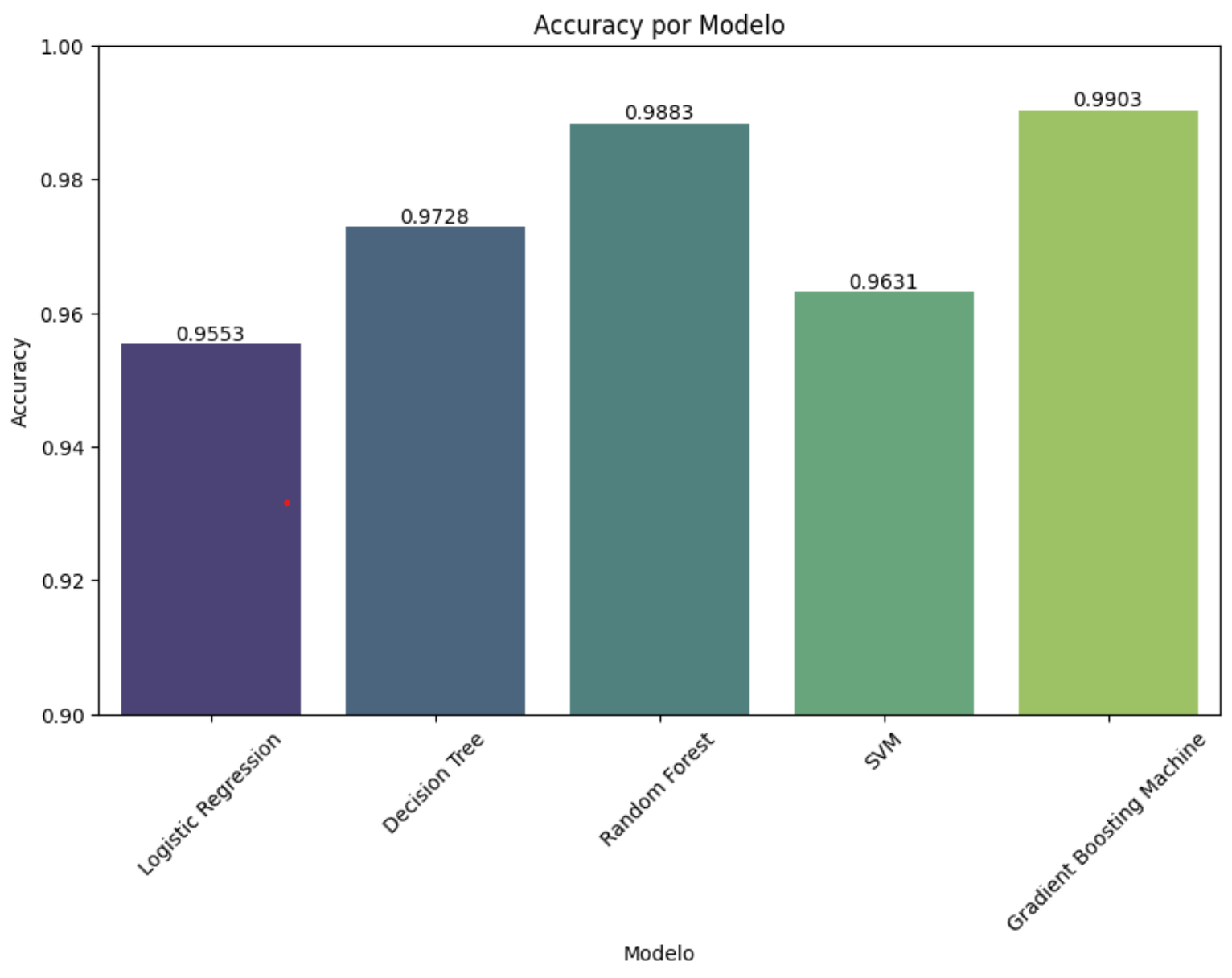
**Conclusión**

Como conclusión podemos expresar que la mayoría de las personas se encuentra de forma neutral a la hora de representar si el trabajo remoto o hibrido son buenos para la salud mental.

Hemos obtenidos grandes accuaracy con nuestros algoritmos, pero el mas eficaz es el Gradient Boosting Machine con un valor de 098/0.99.

Aquí veremos dos graficos de barras con los valores que dejaron los algoritmos luego del test.

Este es el caso del trabajo remoto.



Este es el caso del trabajo hibrido.

