

# Estudio de prevalencia y predictores de problemas de salud mental entre trabajadores tecnológicos

Máster Universitario en Ciencia de Datos  
Área de Ciencia de datos aplicada a Salud  
Trabajo de Fin de Master  
Autor: Byron Vinicio Lima Rojas  
Consultor: Jose Luis Iglesias Allones  
Febrero/2019

---

# Índice

1. Introducción
    - a. Contexto y justificación del trabajo
    - b. Objetivos generales y específicos
    - c. Enfoque y método seguido
    - d. Motivación Personal
  2. Salud Mental
  3. Minería de datos y Aprendizaje Automático
  4. Herramientas TIC
  5. Implementación
  6. Resultados
  7. Conclusiones
-

# Contexto y justificación del trabajo

- En la actualidad, son frecuentes diagnósticos como anorexia, ansiedad, bulimia, depresión y otros tipos de trastornos en personas de cualquier edad, estos problemas de salud mental se detectan diariamente a nivel mundial y representan el 14% de las patologías conocidas.
- En departamentos de TI, los desordenes relacionados a la salud mental se encuentran en constante aparición, entre las múltiples causas se encuentra el estrés, presión laboral, antecedentes familiares, horarios de trabajo ajustados, problemas familiares entre otros.
- El objetivo es realizar el análisis de datos recolectados mediante encuesta a más de 1000 personas, centrándonos únicamente en profesionales de TI que permitan identificar factores, patrones y síntomas que puedan convertirse en problemas para el colaborador con la finalidad de generar sensibilización y mejorar las condiciones de los trabajadores con trastornos de salud mental en su ambiente de trabajo.

# Objetivos

## Objetivo general

Analizar la prevalencia de problemas de salud mental entre trabajadores tecnológicos y definir predictores para detectar síntomas de forma temprana.

## Objetivos específicos

- Identificar síntomas de prevalencia entre trabajadores tecnológicos y la afectación por género del trabajador.
- Evaluar la apertura de los trabajadores para buscar ayuda y las prestaciones de las organizaciones para tratar síntomas en sus colaboradores en base a factores de apertura/ayuda.
- Evaluar si el tipo de empresa (grande, mediana o pequeña) afecta a la proporción de los beneficios y prestaciones para la atención mental de sus colaboradores.

# Enfoque y método seguido

## 1. Encuesta de salud mental en tecnología 2016 - OSMI

## 2. Procesos de Minería de datos

- Selección del conjunto de datos
- Análisis de las propiedades de los datos
- Transformación del conjunto de datos de entrada
- Selección y aplicación de técnicas de minería de datos
- Extracción de conocimiento
- Interpretación y evaluación de datos

## 3. Modelo en cascada/secuencial

# Motivación personal

- Experiencia Laboral
- Seguros públicos y privados
- Prestaciones laborales

# Salud Mental

- El estrés laboral es cada vez más frecuente y una de las principales causas de baja laboral. Cada año aumenta el número de trabajadores que sufre de estrés y ansiedad por motivos laborales, y se estima que en 2020 sea la primera causa de muerte a nivel mundial.
- Para estimar la prevalencia, la gravedad y el tratamiento del diagnóstico y estadístico de trastornos mentales, la OMS se encuentra constantemente analizando la gravedad de los trastornos mentales no tratados a nivel mundial.
- A nivel mundial, el costo global anual por salud mental es de \$ 2.5 billones y se espera que esta cifra aumente a \$ 6 mil millones en 2030.
- La constante aparición de tecnologías disruptivas y la evolución de la transformación digital están realizando un importante incremento de la calidad de vida a nivel social y profesional.

## Minería de datos y Aprendizaje Automático

- La minería de datos utiliza los métodos de la inteligencia artificial, aprendizaje automático, estadística y sistemas de bases de datos. El objetivo general del proceso de minería de datos consiste en extraer información de un conjunto de datos y transformarla en una estructura comprensible para su uso posterior.
- El aprendizaje automático incorpora bases de la minería de datos, pero también puede crear correlaciones automáticas y aprender de ellas para aplicarlas a nuevos algoritmos.
- Los algoritmos de aprendizaje automático se clasifican en dos grandes categorías: supervisados (predictivos) y no supervisados (descubrimiento del conocimiento).



## Herramientas TIC

- Jupyter notebook para la edición y ejecución de código.
- Microsoft Excel para la revisión y preparación de datos.
- Plataforma de Kaggle para la obtención de información.
- Python 3.7 para el uso de librerías y comandos.
- Weka para Aprendizaje automático.
- Librerías de Pandas, Matplotlib, Scikit-learn, graphviz

# Implementación: Tratamiento de los datos



Open Sourcing Mental Illness, LTD

West Lafayette, Indiana, United States  
 Joined 3 years ago  
<https://osmihelp.org>

Se dispone en el conjunto de datos de 1433 encuestas válidas para el estudio a realizar y 63 preguntas o atributos

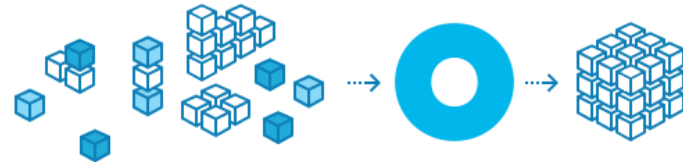
Descarga de información y carga en Python



Análisis de las propiedades de datos

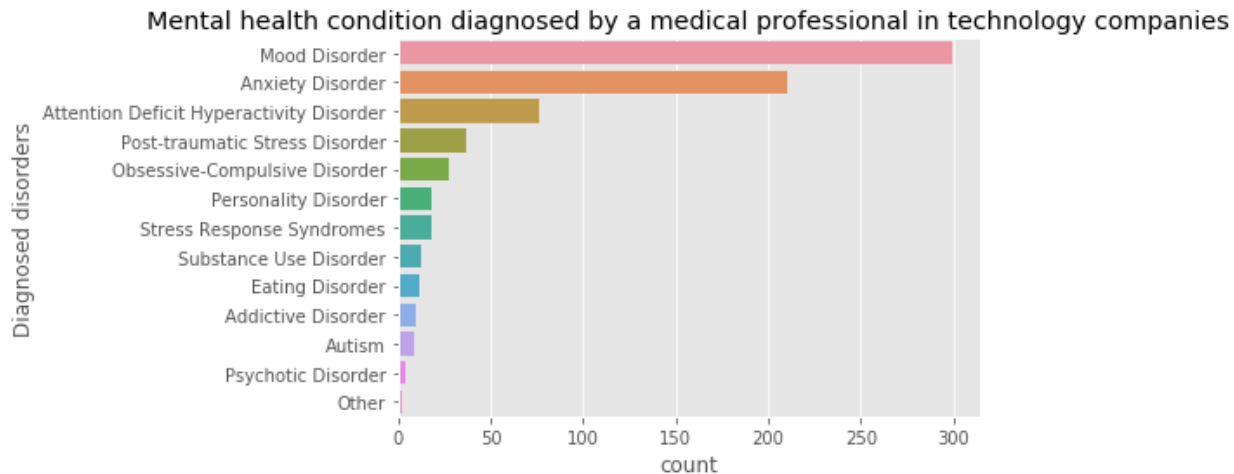
	Are you self-employed? company/organization?	In your primary job, do you work for a company?	Do you have medical coverage (group insurance or self-provided health insurance)?	Do you have a spouse/generic partner?	How do you rate your mental health from a neutral health perspective?	What is your age?
count	1433.00000	1146.00000	263.00000	237.00000	1433.00000	1433.00000
mean	0.202179	0.770000	0.362000	0.644000	0.802000	0.500000
std	0.401048	0.420000	0.482000	0.479671	0.322000	0.400000
min	0.000000	0.000000	0.000000	0.000000	0.000000	19.200000
25%	0.000000	1.000000	1.000000	0.000000	1.000000	0.000000
50%	0.000000	1.000000	1.000000	1.000000	1.000000	33.000000
75%	0.000000	1.000000	1.000000	1.000000	1.000000	38.000000
max	1.000000	1.000000	1.000000	1.000000	1.000000	53.000000

Transformación del conjunto de datos de entrada



Tomaremos en cuenta la información del 77,6% de los encuestados, seleccionando a los encuestados que trabajan en empresas de tecnología o relacionadas a la parte tecnológica, quedando únicamente con 788 registros y 43 variables.

## Implementación: Análisis estadístico de datos



**Grafica 1.** Condición de salud mental diagnosticada por un profesional en empresas tecnológicas

## Implementación: Extracción de conocimiento

Se utilizó el 80% de datos para entrenamiento y 20% de datos para prueba, el detalle de la cantidad de los datos se muestra a continuación:

- Entrenamiento X: (630, 16)
- Entrenamiento Y: (630,)
- Validación X: (158, 16)
- Validación Y: (158,)

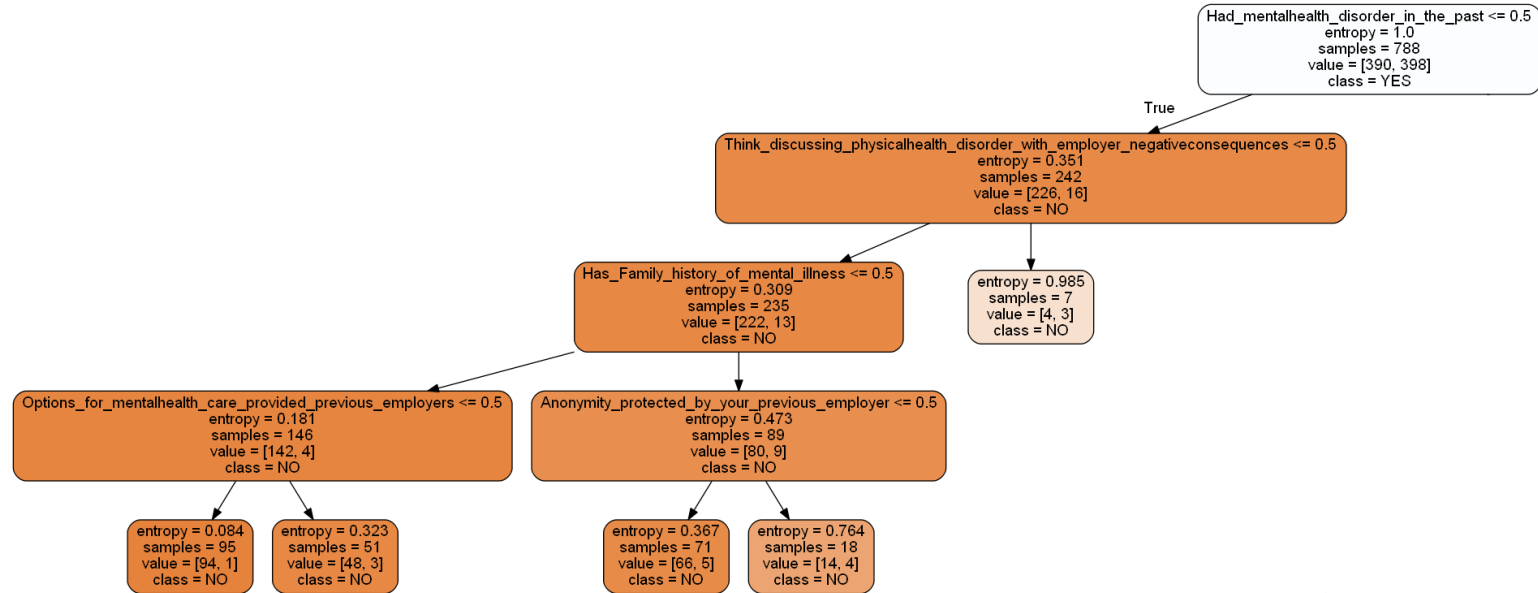
La revisión del aprendizaje del algoritmo con datos de entrenamiento y test brinda los siguientes resultados de precisión con el modelo inicial:

- Train: 0.806
- Test: 0.766

La matriz de confusión de las predicciones del grupo Test. La diagonal de esta matriz se lee: arriba a la izquierda True Negatives y abajo a la derecha True Positives.

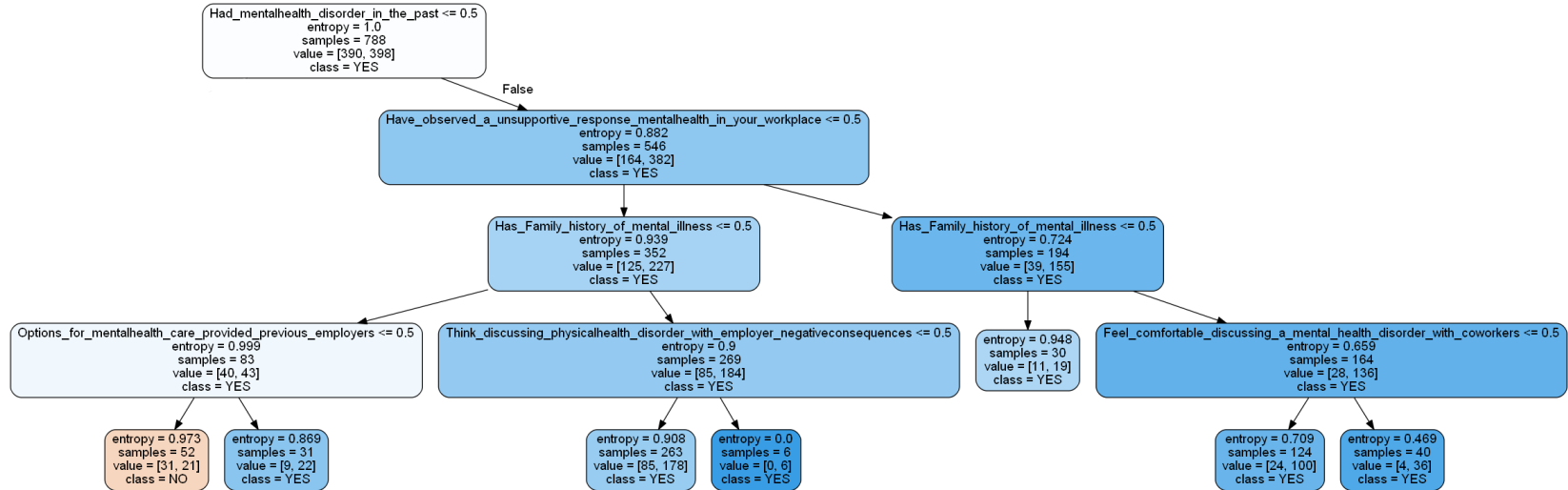
53	27
10	68

# Implementación: Extracción de conocimiento



**Gráfica 6.** Encuestados no diagnosticados con problemas de salud mental en el pasado

# Implementación: Extracción de conocimiento

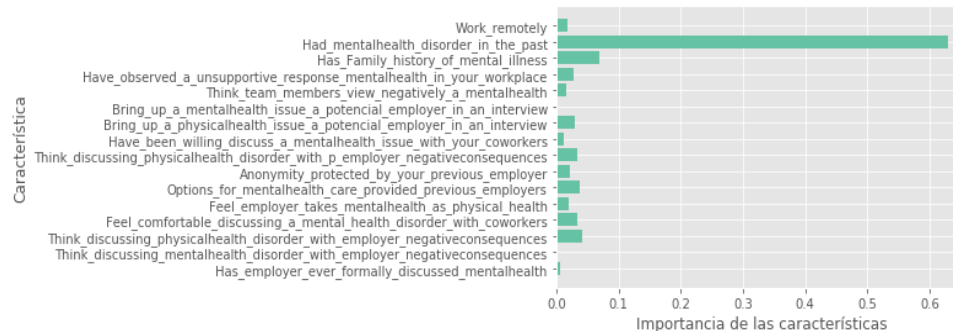


**Grafica 7.** Encuestados diagnosticados con problemas de salud mental en el pasado

# Implementación: Resultados

	Decision Tree Classifier in Python	Naive Bayers Tree	J48
Instancias correctamente clasificadas	76.58%	73.4177%	70.8861%
Instancias incorrectamente clasificadas	23.42%	26.5823%	29.1139%
Verdaderos positivos	45	57	55
Falsos negativos	33	14	16
Falsos Positivos	12	28	39
Verdaderos negativos	68	59	57

**Grafica 8.** Porcentajes de clasificación con árboles de decisión con WEKA



**Gráfica 9.** Importancia de variables en árbol de decisión del conjunto de datos original

# Implementación: Resultados

La interpretación de los resultados se segmenta en varios puntos de estudio:

- Resultados generales.
- Desordenes de salud mental con mayor diagnóstico.
- Problemas de salud mental por edad y género
- Antecedentes de salud mental y apertura para la búsqueda de ayuda.



## Conclusiones

De acuerdo al estudio realizado con los resultados de las encuestas publicadas por OSMI del año 2016, se ha llegado a las siguientes conclusiones:

- Entre los síntomas de prevalencia de salud mental, se encuentran los antecedentes de problemas de salud mental y también, influyen factores de antecedentes familiares. Este tipo de desórdenes tiene mayor impacto entre los 25 años y 40 años, en el caso de las personas con género diferente existen un 90% de probabilidad de presencia de algún tipo de desorden.
- Las estadísticas de apertura de los trabajadores para búsqueda de ayuda son muy reducidas, en muchos de los casos no es una opción considerada por las personas para manifestarla en sus ambientes de trabajo o en entrevistas de trabajo. Esta falta de apertura llega a tener una relación directa con respecto al alcance de las prestaciones que brindan las organizaciones en temas de salud.
- A nivel de empresas, se tiene información de prestaciones de los empleadores anteriores en donde no se conoce el tipo de empresa. De acuerdo a los datos estadísticos, existe una mayor presencia de personas diagnosticadas con algún tipo de desorden en empresas con más de 1000 empleados; sin embargo, a nivel de beneficios que brinden las empresas existe desconocimiento de los mismos por parte de los encuestados.

¡Gracias por su atención!