

Desarrollo de un modelo predictivo en base a variables sociodemográficas y de salud que derivan en ansiedad y depresión



Universidad
Internacional
de Valencia

De:



Planeta Formación y Universidades

01

Motivación Y Objetivos



Motivación

- La ansiedad y la depresión son dos tipos de trastornos dentro de las enfermedades mentales.
- Tienen un efecto nocivo en el comportamiento de las personas.
- Sus síntomas no son de fácil detección y muchas veces están relacionados con el entorno y las condiciones personales del individuo.
- Con la aparición del COVID-19 en el año 2020 estos trastornos han empezado a aumentar considerablemente sin que los sistema de salud estén preparados para afrontar esta situación.
- Aplicación de la ciencia de datos en la detección de enfermedades mentales y en especial en trastornos como la depresión y la ansiedad.

"No hay salud, sin salud mental."
Dr. Tedros Adhanom Ghebreyesus
Director general
Organización Mundial de la Salud



Objetivos

- **GENERAL:**
- Evaluar y analizar un conjunto de indicadores de salud, factores sociales y de estilo de vida que permitan la detección temprana de la depresión y la ansiedad.
- **ESPECIFICOS:**
- Implementar un proceso de extracción de conocimiento KDD, (del inglés *Knowledge Discovery in Databases*).
- Desarrollar un modelo predictivo sobre indicadores socioeconómicos y de salud que busque pronosticar la presencia de estos trastornos.
- Encontrar patrones asociados con la ubicación geográfica y que permita identificar dónde se ubican la mayoría de los casos de ansiedad y depresión.
- Evaluar si hay alguna influencia del COVID-19 en el comportamiento de los indicadores de la ansiedad y la depresión.

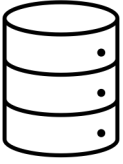
02

La Ciencia de Datos en La Salud Mental



Fuentes de Investigación

#	Autor(es)	Título	Año	Aplicación
1	Sau y Bhakta	Detección de ansiedad y depresión entre trabajadores del mar utilizando tecnología de aprendizaje automático.	2019	<ul style="list-style-type: none"> Investigación basada en encuestas. Variables sociodemográficas de y salud. Uso de algoritmos de clasificación.
2	Richter, Fishbain, Fruchter, Richter-Levin y Okon-Singer	Sistema de soporte de diagnóstico basado en aprendizaje automático para diferenciar entre trastornos clínicos de ansiedad y depresión.	2021	<ul style="list-style-type: none"> Investigación basada en variables clínicas. Uso de métodos de clasificación.
3	Rodríguez Muñoz, Cebriá Andreu, Corbella Santomá, Segura Bernal y Sobreques Soriano	Rasgos de personalidad y malestar psíquico asociados a los pacientes hiperfrecuentadores de consultas de Atención Primaria.	2003	<ul style="list-style-type: none"> Perfil de pacientes hiperfrecuentadores en los sistemas de salud. Uso de datos sociodemográficos y de salud. Uso de métodos estadísticos. Uso de una metodología de análisis de datos.



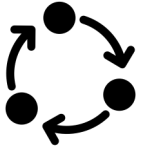
Encuesta Europea de Salud (EESA-2020)

Población	Personas adultas mayores de 15 años
Ámbito geográfico	Nacional (España)
Trabajo de campo	15 de julio de 2019 – 24 de julio 2020
Tamaño muestral	22.072 entrevistas realizadas

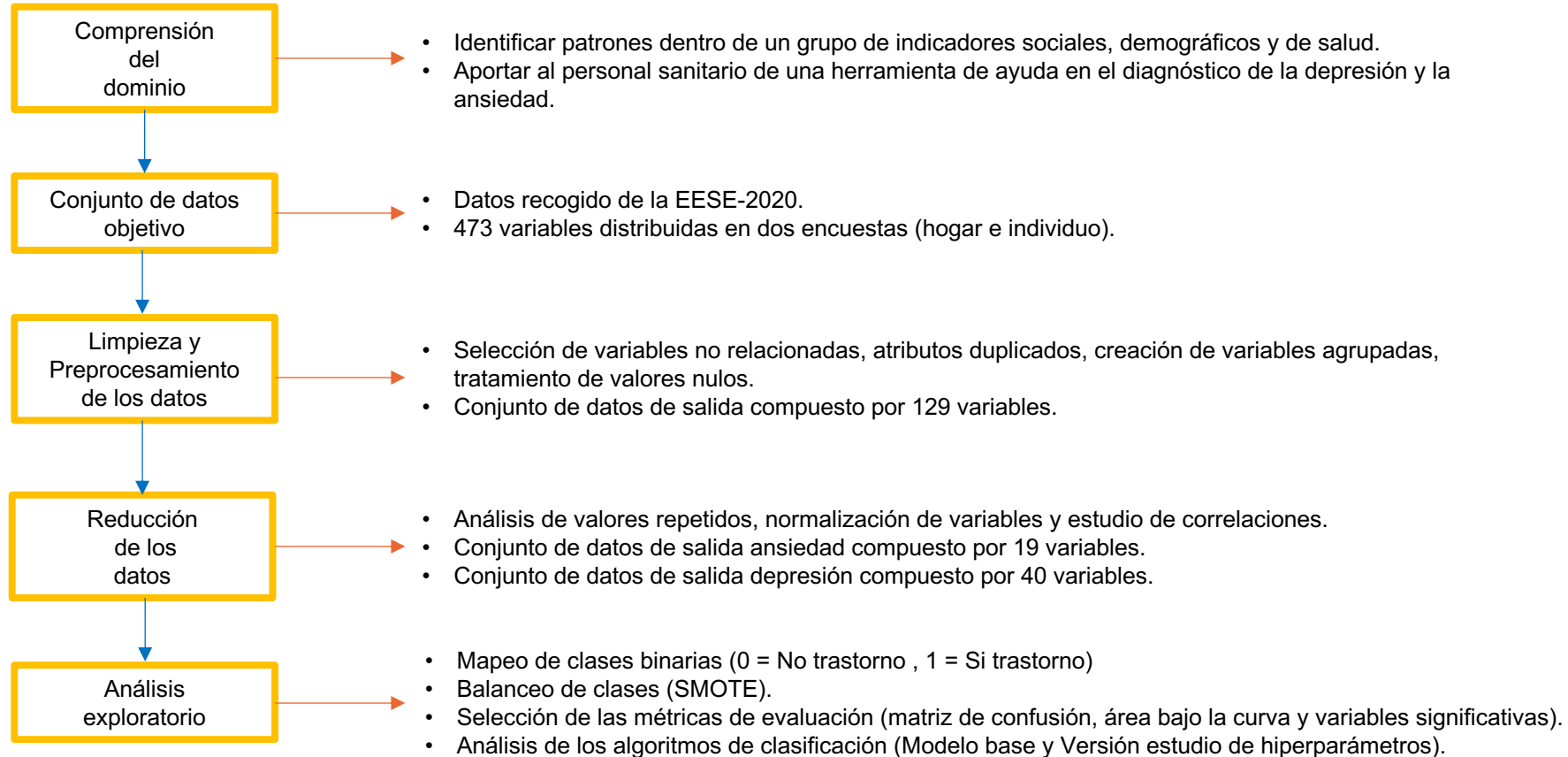
- Modelo conjunto de encuesta impulsado por la Oficina Europea de Estadística (Eurostat).
- Integrada en la Plan Estadístico Nacional de España, y ejecutada en 2009, 2014 y 2020.
- Recoge información sanitaria y datos sociodemográficos.
- Consta de cuatro módulos:
 - Módulo sociodemográfico
 - Módulo del estado de la salud
 - Módulo de asistencia sanitaria
 - Módulo de determinantes de la salud

03

Desarrollo del Proyecto



Metodología (Proceso KDD)





04

Evaluación y Resultados



Rendimiento de los modelos

Conjunto de datos ansiedad

Clasificador	Modelo Base						GridSearchCV					
	Accuracy	Recall 0	Recall 1	F1-Score 0	F1-Score 1	AUC	Accuracy	Recall 0	Recall 1	F1-Score 0	F1-Score 1	AUC
Regresión Logística	0.63	0.68	0.59	0.65	0.62	0.68	0.63	0.67	0.59	0.65	0.62	0.68
Árbol de Decisiones	0.90	0.89	0.92	0.90	0.91	0.90	0.72	0.84	0.60	0.75	0.68	0.80
SVM	0.84	0.78	0.89	0.83	0.84	0.91	0.93	0.88	0.99	0.93	0.94	0.97
Naive Bayes	0.63	0.74	0.52	0.67	0.58	0.69	0.63	0.74	0.52	0.67	0.58	0.69
SGD	0.63	0.64	0.62	0.63	0.63	0.68	0.63	0.68	0.58	0.65	0.61	0.58
Random Forest	0.96	0.99	0.93	0.96	0.96	0.99	0.96	0.99	0.93	0.96	0.96	0.99
Extra Trees	0.98	0.98	0.98	0.98	0.98	1	0.97	0.97	0.98	0.97	0.97	1
XgBoost	0.92	0.98	0.87	0.93	0.92	0.97	0.95	0.99	0.91	0.95	0.95	0.98
CatBoost	0.95	0.99	0.91	0.95	0.95	0.98	0.92	0.97	0.86	0.92	0.91	0.97
Evaluación del Modelo Final												
Stacked Model	Random Forest + SVM + XgBoost						0.96	0.98	0.94	0.96	0.96	0.99
Extra Trees							0.97	0.97	0.98	0.97	0.97	1

Conjunto de datos depresión

Clasificador	Modelo Base						GridSearchCV					
	Accuracy	Recall 0	Recall 1	F1-Score 0	F1-Score 1	AUC	Accuracy	Recall 0	Recall 1	F1-Score 0	F1-Score 1	AUC
Regresión Logística	0.80	0.82	0.78	0.80	0.79	0.88	0.80	0.82	0.78	0.80	0.79	0.88
Árbol de Decisiones	0.92	0.92	0.93	0.92	0.92	0.92	0.84	0.82	0.97	0.84	0.85	0.92
SVM	0.82	0.85	0.79	0.83	0.82	0.91	0.96	0.94	0.99	0.96	0.97	0.99
Naive Bayes	0.77	0.79	0.74	0.77	0.76	0.85	0.77	0.79	0.74	0.77	0.76	0.85
SGD	0.80	0.74	0.85	0.78	0.81	0.88	0.80	0.78	0.82	0.80	0.81	0.88
Random Forest	0.96	0.97	0.95	0.96	0.96	0.99	0.96	0.97	0.95	0.96	0.96	0.99
Extra Trees	0.97	0.96	0.99	0.97	0.97	1	0.96	0.94	0.99	0.96	0.96	1
XgBoost	0.95	0.96	0.93	0.95	0.95	0.99	0.95	0.98	0.93	0.96	0.95	0.99
CatBoost	0.96	0.98	0.93	0.96	0.95	0.99	0.95	0.96	0.94	0.95	0.95	0.99
Evaluación del Modelo Final												
Stacked Model	Random Forest + SVM + XgBoost						0.98	0.98	0.97	0.98	0.98	1
Extra Trees							0.96	0.94	0.99	0.96	0.96	1

Matriz de Confusión para el Stack de Modelos

0 = No 1 = Si	0	1
0	3969	87
1	229	3827

Matriz de Confusión para Extra Trees

0 = No 1 = Si	0	1
0	3922	134
1	91	3965

Matriz de Confusión para el Stack de Modelos

0 = No 1 = Si	0	1
0	3907	85
1	110	3881

Matriz de Confusión para Extra Trees

0 = No 1 = Si	0	1
0	3746	246
1	54	3937



Informe de características principales

Informe Ansiedad

EESE-2020

viu

ANSIEDAD

Comunidad Autónoma

Todas

Rango de Edad

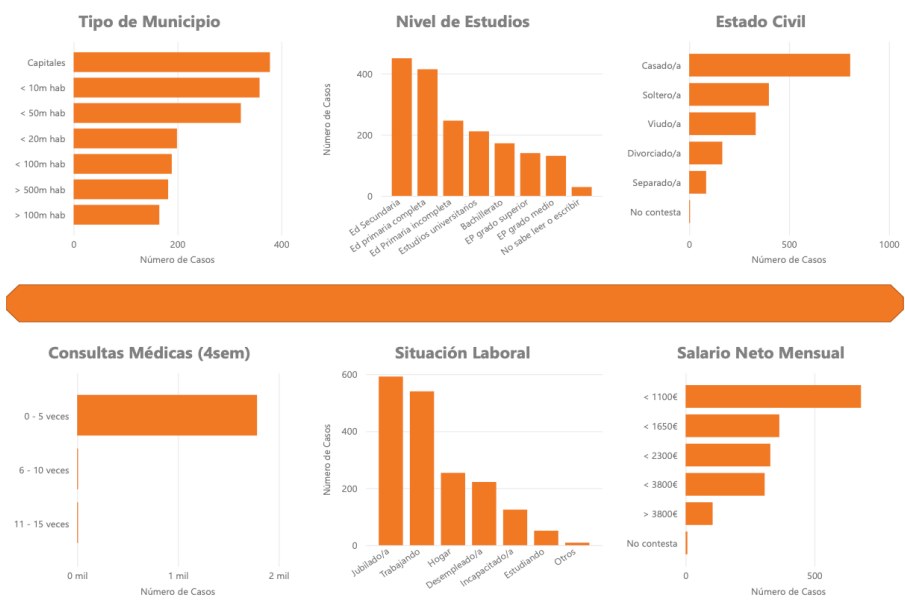
Todas

Número de Muestras

22072

Número de Casos Positivos

1793



Informe Depresión

EESE-2020

viu

DEPRESION

Comunidad Autónoma

Todas

Rango de Edad

Todas

Número de Muestras

22072

Número de Casos Positivos

2115





Comparativa 2014 vs 2020 (I)

Conjunto de datos ansiedad

Matriz de Confusión para el Stack de Modelos

0 = No 1 = Si	0	1
0	3960	172
1	360	3772

Matriz de Confusión para Extra Trees

0 = No 1 = Si	0	1
0	3700	432
1	253	3879

Conjunto de datos depresión

Matriz de Confusión para el Stack de Modelos

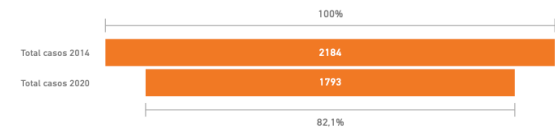
0 = No 1 = Si	0	1
0	3879	149
1	332	3696

Matriz de Confusión para Extra Trees

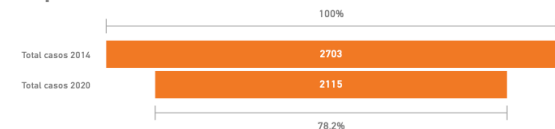
0 = No 1 = Si	0	1
0	3728	300
1	147	3881

EESE-2020 Análisis de Tendencias

Ansiedad 2014 vs 2020



Depresión 2014 vs 2020



Ratio Casos de Ansiedad sobre Total de Muestras

9,56 %
2014

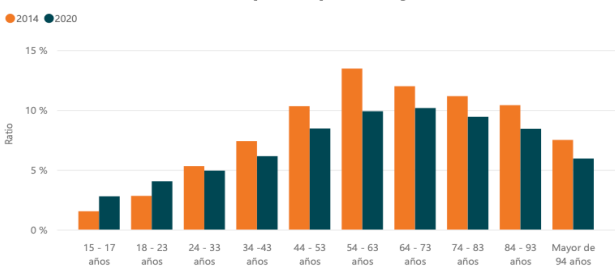
8,12 %
2020

Ratio Casos de Depresión sobre Total de Muestras

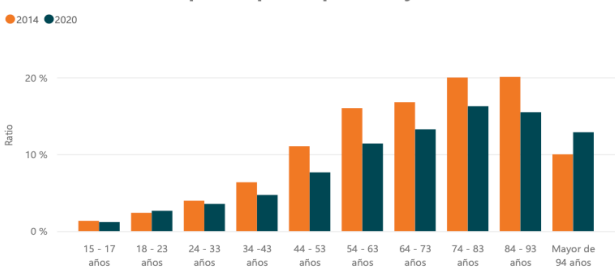
11,83 %
2014

9,58 %
2020

Ansiedad por Grupo Etario y Año



Depresión por Grupo Etario y Año





Comparativa 2014 vs 2020 (II)

Variables significativas 2014

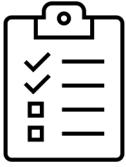
Ansiedad	Depresión
() número de apariciones en los modelos	
edad (5)	tipo_hogar (5)
lugar_consulta4sem (5)	nivel_estudios (5)
nivel_estudios (5)	situacion_laboral (4)
altura (5)	edad (4)
frecuencia_actfisica (5)	altura (2)
peso (5)	frecuencia_actfisica (2)
	tipo_municipio (2)
	apoyosocial_interes (1)

Variables significativas 2020

Ansiedad	Depresión
() número de apariciones en los modelos	
tipo_municipio (5)	consumo_medicamentos (5)
nivel_estudios (4)	limitación_salud6m (5)
salario_neto (4)	nivel_depresion (4)
cantidad_consultas4sem (3)	estado_salud12m (4)
situacion_laboral (2)	prevalencia_depresion (3)
comunidad_autonoma (2)	colesterol (1)
hemorroides (2)	dolor_4semanas (1)
estado_civil (1)	frecuencia_consumopescado (1)
altura (1)	apoyosocial_ayuda (1)
edad (1)	dificultad_concentracion (1)

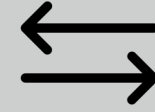
05

Conclusiones y Trabajos Futuros



Conclusiones

- El uso de una metodología basada en el proceso KDD permitió encontrar patrones relacionados con la ansiedad y la depresión.
- Los algoritmos basados en conjuntos de arboles de decisión (ensembles) son los modelos que mejor generalizaron el problema de clasificación planteado. En especial, el clasificador Extra Trees.
- Se pudo concluir que sí existe una marcada relación de variables sociales y económicas con la ansiedad.
- Respecto a la depresión se obtuvo que existe relación con datos asociados a condicionantes del estado de la salud.
- Con los datos utilizados no se encontró una tendencia en aumento de los casos entre 2014 y 2020 para la totalidad del conjunto. Pero sí para dos grupos específicos: menores de 23 años (ansiedad) y mayores de 85 años (depresión).



Trabajos Futuros

- Replicar este trabajo con encuestas realizadas por profesionales de salud mediante estudios clínicos.
- Enriquecimiento de variables del comportamiento del individuo a través del uso de las nuevas tecnologías (internet y redes sociales).
- Uso de un mayor histórico de datos en el tiempo y sobre una misma población.
- Agregar datos relacionados al número de profesionales que actualmente hay en cada localidad y la distribución de los centros sanitarios.
- Implantar campañas de detección temprana y de concienciación en el ciudadano sobre los grupos identificados con mayor prevalencia mediante técnicas de agrupación.

¡Gracias por vuestra atención!



Universidad
Internacional
de Valencia