Trabajo de Fin de Máster

Minería de datos y relación entre tipos de inteligencias.

Autor: José Pérez-Parras Toledano
Tutora: María Del Carmen Pegalajar Jiménez

Universidad de Granada

Indice

1. Introducción y objetivos
2. Análisis de los datos
3. Relación entre las variables
4. Creación de los experimentos para predicción
5. Resultados de Regresión
6. Aprendizaje no supervisado
7. Resultados de Clasificación
8. Conclusiones

1 Introducción y objetivos

Repaso de los objetivos principales en nuestro trabajo

Minería de datos a partir de encuestas psicológicas

- Datos recolectados por la **Universidad Rovira i Virgili**
- **Joan Boada** y **Elena Pérez** para la Tesis "La inteligencia energética, más allá de la inteligencia emocional"
- Encuestas que miden distintos aspectos de los individuos.
- Recolección de aspectos sociales y psicológicos.

Objetivos principales

- Análisis de las variables del problema
- Estudio correlacional en relación con la Inteligencia Energética
- Predicción de algunas variables con métodos de **regresión**
- Estudio con técnicas de aprendizaje no supervisado
- Predicción de algunas variables con métodos de **clasificación**

Tecnologías usadas

Python 3.6

matplotlib para la visualización

scikit-learn para los modelos y métricas

Jupyter notebooks para la ejecución

2 Análisis de los datos

Breve análisis de los datos

Primer análisis del problema

856 Personas encuestadas

47 Variables a analizar

19 Variables de carácter social

28 Variables de tipo psicológico

Variables del problema

Variables sociales

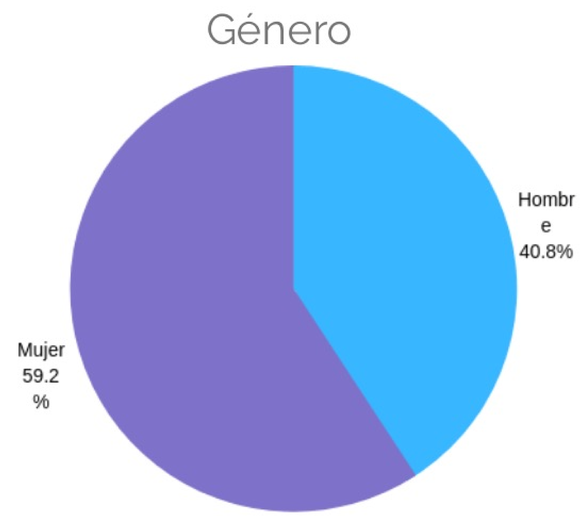
- Género
- Edad
- Estado civil
- Situación laboral-ocupacional
- Nivel de formación académica
- Unidad de Negocio-División
- Antigüedad en empresa-institución
- Cargo
- Número de profesionales a cargo
- Antigüedad en el cargo
- Antigüedad en profesión
- Grupo profesional (CIUO)
- Sector
- Pertenencia a Banca Internacional
- Número de empleados en empresa
- Factura anual aproximada
- Ámbito de actuación de empresa
- Situación de la empresa en el último año

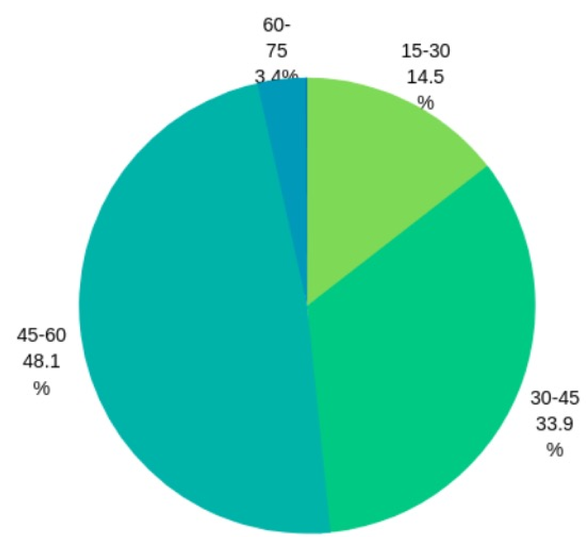
Variables psicológicas

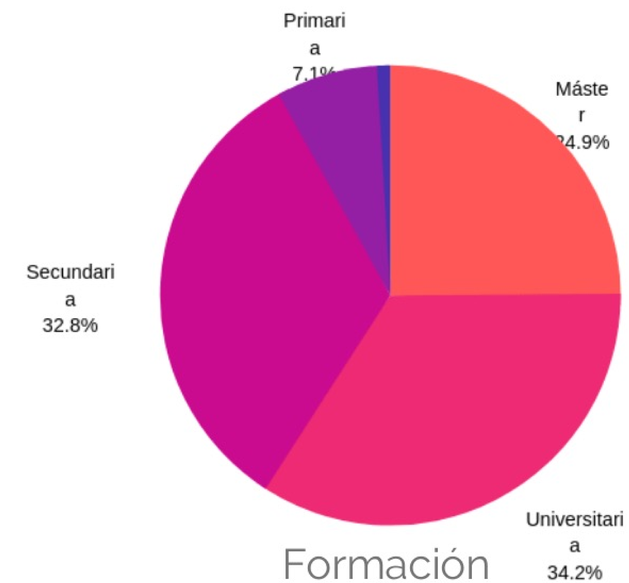
- **FLow-Fluir**
- Autoeficacia
- Autoestima
- Extraversión
- Estabilidad emocional
- Responsabilidad
- Amabilidad
- Apertura a la experiencia
- **Flurishing-Florecimiento**
- Vigor
- Dedicación
- Absorción
- Engagement total
- Satisfacción con la vida
- Percepción
- Comprensión
- Regulación
- Inteligencia emocional total
- **Self Drive Direction**
- **Consciencia energética**
- **Energía total**
- Vitalidad emocional
- Estrés organizacional
- Estrés físico
- Abandono
- Riesgo salud
- Calidad personal y organizacional total

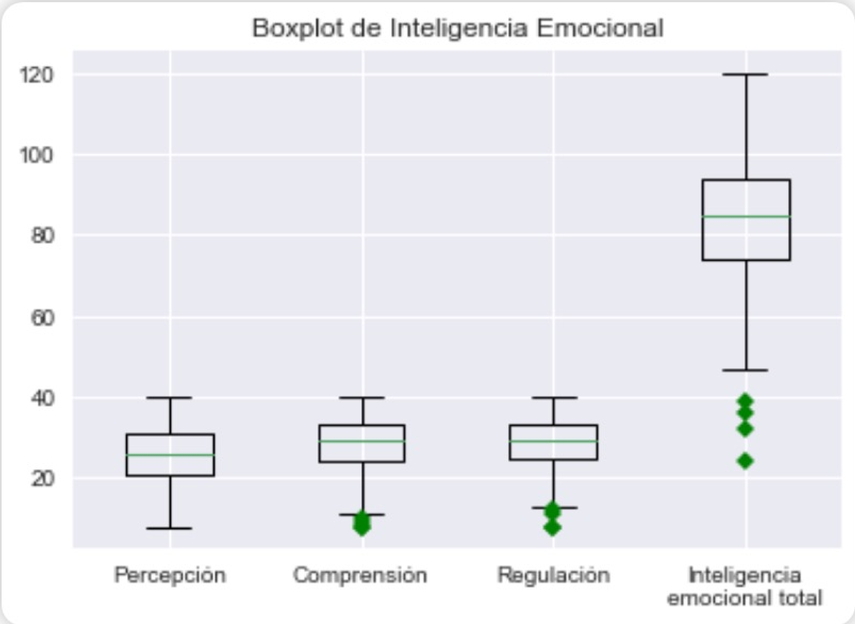
Algunos análisis de los aspectos sociales

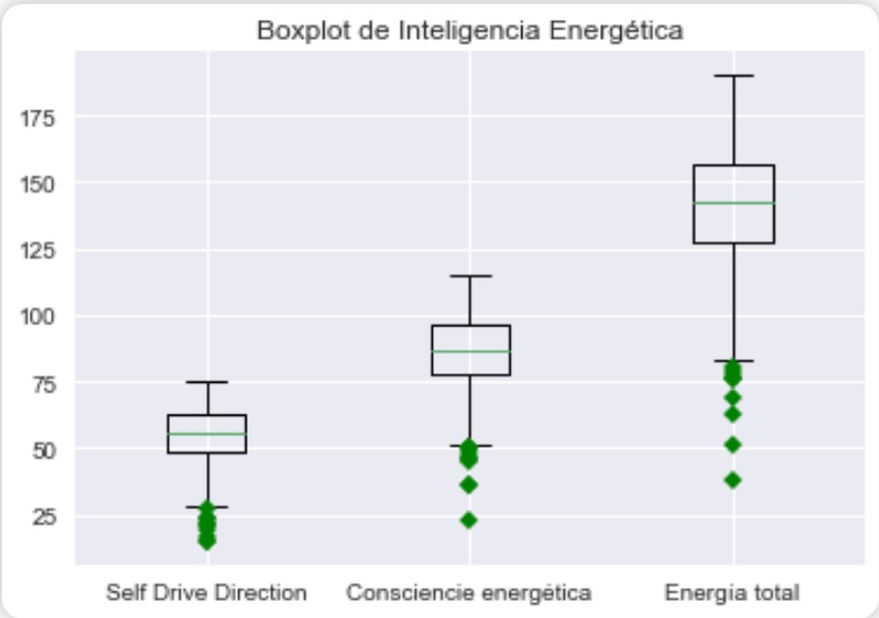
- Igualdad de encuestados en cuanto a género.
- Mayor representación de encuestados en una edad entorno a los 50 años.
- Alta formación entre los encuestados.
- La mayoría de ellos en estado de empleo.









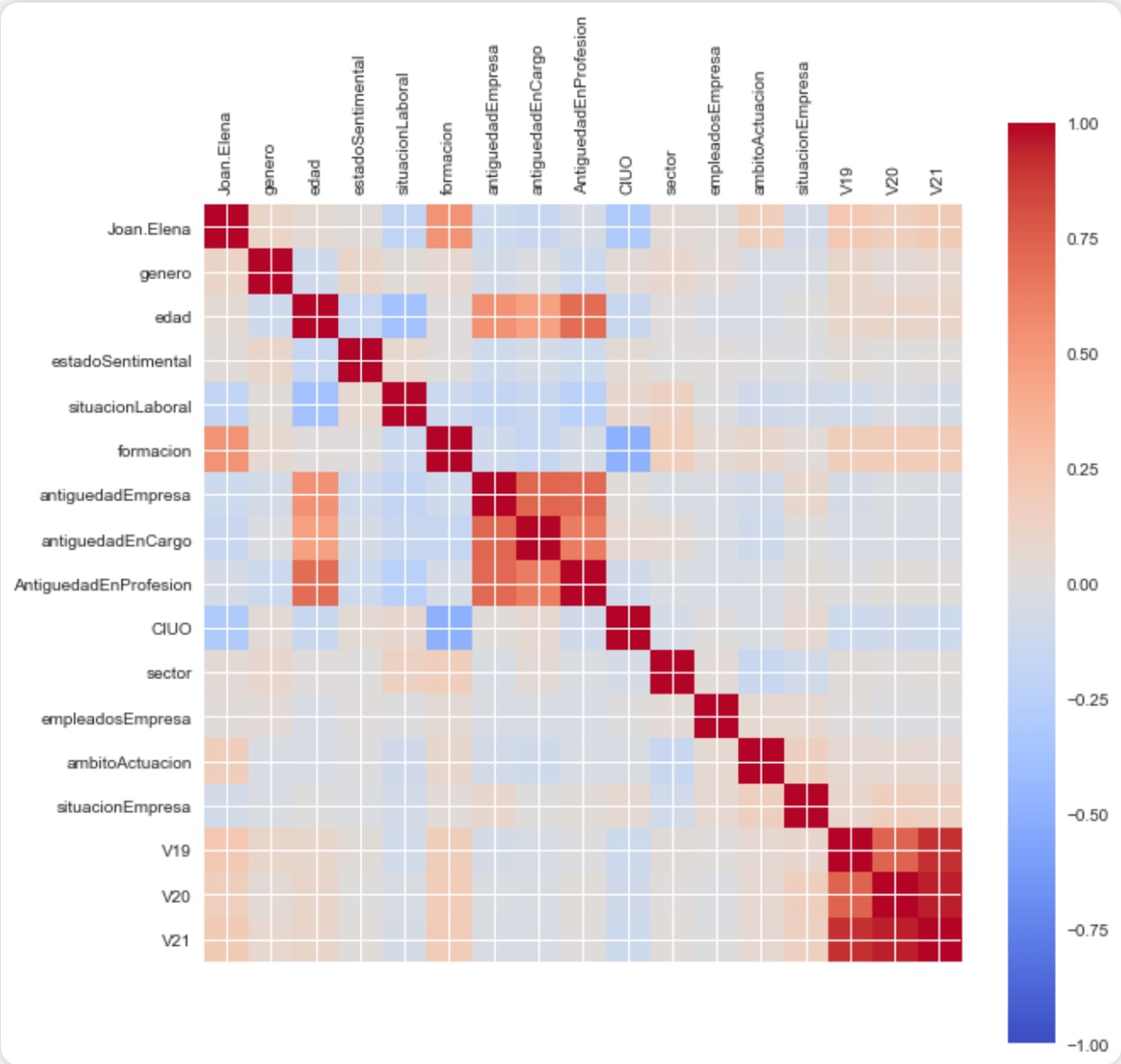


Análisis de los aspectos psicológicos

- Todas las variables psicológicas siguen el rango formal establecido.
- Existe una mayor concentración de encuestados en los rangos altos de las mismas.

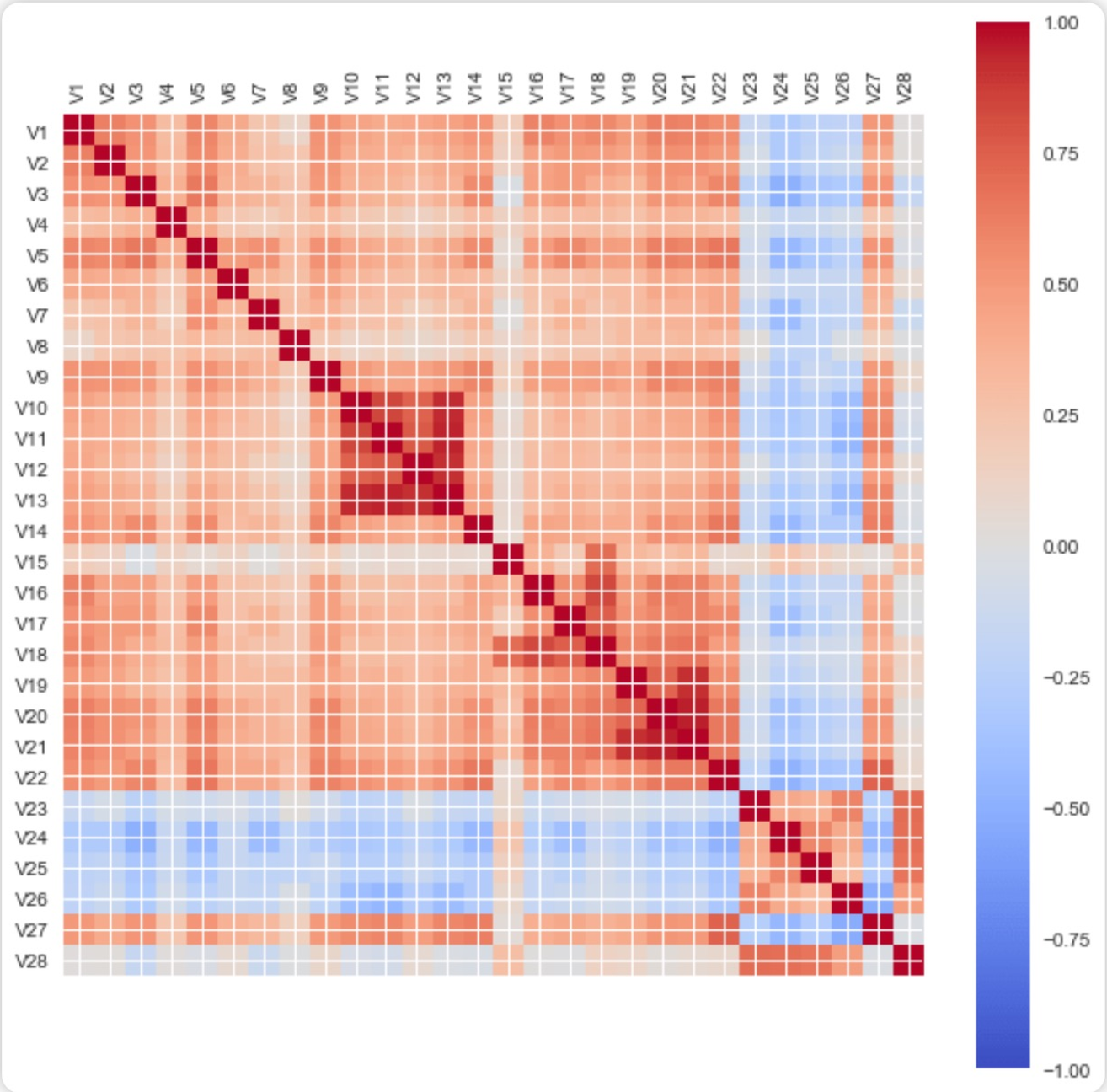
3 Relaciones entre las variables

Relaciones entre las variables con el uso de heatmap.



Correlación con las variables sociales

- No existe apenas relación entre las variables energéticas y las sociales.
- Relación un poco pronunciada con la formación.
- Mayor relación con **edad** y **género**.



Correlación con las variables psicológicas

- Mayor relación entre las propias variables de carácter psicológico
- Se ven claramente los clusters de variables.
- Relación muy pronunciada con las variables de tipo emocional.
- Otras variables con gran correlación serán la **Autoeficacia**, **Autoestima** y **Estabilidad emocional**.

4 Creación de los experimentos

Creación de los experimentos junto a los modelos y métricas usadas

Estudios propuestos por los expertos psicólogos

Se han realizado los experimentos mencionados sobre estos estudios

Dp 1.- Género
Dp 2.- Edad
Dp 3.- Estado civil
Dp 4.- Situación
Dp 5.- Nivel de formación académica
Dp 7.- Antigüedad en la empresa-institución
Dp 9.- Número total de profesionales/empleados a su cargo
Dp 10.- Antigüedad en el cargo
Dp 11.- Antigüedad en esta profesión
V2.- Autoeficacia
V3.- Autoestima
V4.- Extraversión
V5.- Estabilidad emocional
V6.- Responsabilidad
V7.- Amabilidad
V8.- Apertura a la experiencia
V10.- Engagement 1. Vigor
V11.- Engagement 2. Dedicación
V12.- Engagement 3. Absorción
V13.- Engagement total
V14.- Satisfacción con la vida
V15.- Inteligencia emocional 1. Percepción.
V16.- Inteligencia emocional 2. Comprensión
V17.- Inteligencia emocional 3. Regulación
V18.- Inteligencia emocional total
V22.- VCPO 1. Vitalidad emocional
V23.- VCPO 2. Estrés organizacional
V24.- VCPO 3. Estrés emocional
V25.- VCPO 4. Estrés físico

V1.- Flow-Fluir

Dp 1.- Género
Dp 2.- Edad
Dp 3.- Estado civil
Dp 4.- Situación
Dp 5.- Nivel de formación académica
Dp 7.- Antigüedad en la empresa-institución
Dp 9.- Número total de profesionales/empleados a su cargo
Dp 10.- Antigüedad en el cargo
Dp 11.- Antigüedad en esta profesión
V2.- Autoeficacia
V3.- Autoestima
V4.- Extraversión
V5.- Estabilidad emocional
V6.- Responsabilidad
V7.- Amabilidad
V8.- Apertura a la experiencia
V10.- Engagement 1. Vigor
V11.- Engagement 2. Dedicación
V12.- Engagement 3. Absorción
V13.- Engagement total
V14.- Satisfacción con la vida
V15.- Inteligencia emocional 1. Percepción
V16.- Inteligencia emocional 2. Comprensión
V17.- Inteligencia emocional 3. Regulación
V18.- Inteligencia emocional total
V22.- VCPO 1. Vitalidad emocional
V23.- VCPO 2. Estrés organizacional
V24.- VCPO 3. Estrés emocional
V25.- VCPO 4. Estrés físico

V9.- Flourishing-Florecimiento

Dp 1.- Género
Dp 2.- Edad
Dp 3.- Estado civil
Dp 4.- Situación
Dp 5.- Nivel de formación académica
Dp 7.- Antigüedad en la empresa-institución
Dp 9.- Número total de profesionales/empleados a su cargo
Dp 10.- Antigüedad en el cargo
Dp 11.- Antigüedad en esta profesión
V2.- Autoeficacia
V3.- Autoestima
V4.- Extraversión
V5.- Estabilidad emocional
V6.- Responsabilidad
V7.- Amabilidad
V8.- Apertura a la experiencia
V10.- Engagement 1. Vigor
V11.- Engagement 2. Dedicación
V12.- Engagement 3. Absorción
V13.- Engagement total
V14.- Satisfacción con la vida
V15.- Inteligencia emocional 1. Percepción
V16.- Inteligencia emocional 2. Comprensión
V17.- Inteligencia emocional 3. Regulación
V18.- Inteligencia emocional total
V22.- VCPO 1. Vitalidad emocional
V23.- VCPO 2. Estrés organizacional
V24.- VCPO 3. Estrés emocional
V25.- VCPO 4. Estrés físico

V21. Energía Total.

Dp 1.- Género
Dp 2.- Edad
Dp 3.- Estado civil
Dp 4.- Situación
Dp 5.- Nivel de formación académica
Dp 7.- Antigüedad en la empresa-institución
Dp 9.- Número total de profesionales/empleados a su cargo
Dp 10.- Antigüedad en el cargo
Dp 11.- Antigüedad en esta profesión
V2.- Autoeficacia
V3.- Autoestima
V4.- Extraversión
V5.- Estabilidad emocional
V6.- Responsabilidad
V7.- Amabilidad
V8.- Apertura a la experiencia
V10.- Engagement 1. Vigor
V11.- Engagement 2. Dedicación
V12.- Engagement 3. Absorción
V13.- Engagement total
V14.- Satisfacción con la vida
V15.- Inteligencia emocional 1. Percepción
V16.- Inteligencia emocional 2. Comprensión
V17.- Inteligencia emocional 3. Regulación
V18.- Inteligencia emocional total
V22.- VCPO 1. Vitalidad emocional
V23.- VCPO 2. Estrés organizacional
V24.- VCPO 3. Estrés emocional
V25.- VCPO 4. Estrés físico

V19.- Energía.1. Self Drive Direction.

Se mostrará **uno** como ejemplo

Dp 1.- Género
Dp 2.- Edad
Dp 3.- Estado civil
Dp 4.- Situación
Dp 5.- Nivel de formación académica
Dp 7.- Antigüedad en la empresa-institución
Dp 9.- Número total de profesionales/empleados a su cargo
Dp 10.- Antigüedad en el cargo
Dp 11.- Antigüedad en esta profesión
V2.- Autoeficacia
V3.- Autoestima
V4.- Extraversión
V5.- Estabilidad emocional
V6.- Responsabilidad
V7.- Amabilidad
V8.- Apertura a la experiencia
V10.- Engagement 1. Vigor
V11.- Engagement 2. Dedicación
V12.- Engagement 3. Absorción
V13.- Engagement total
V14.- Satisfacción con la vida
V15.- Inteligencia emocional 1. Percepción
V16.- Inteligencia emocional 2. Comprensión
V17.- Inteligencia emocional 3. Regulación
V18.- Inteligencia emocional total
V22.- VCPO 1. Vitalidad emocional
V23.- VCPO 2. Estrés organizacional
V24.- VCPO 3. Estrés emocional
V25.- VCPO 4. Estrés físico

V19.- Energía.1. Self Drive Direction.

Todos los estudios realizados

1. Predicción de las **5 variables** a estudiar propuestas por los psicólogos.

2. **Variables de entrada** propuestas por el **equipo de psicólogos**.

3. Predicción usando **regresión** con **las variables propuestas** del problema.

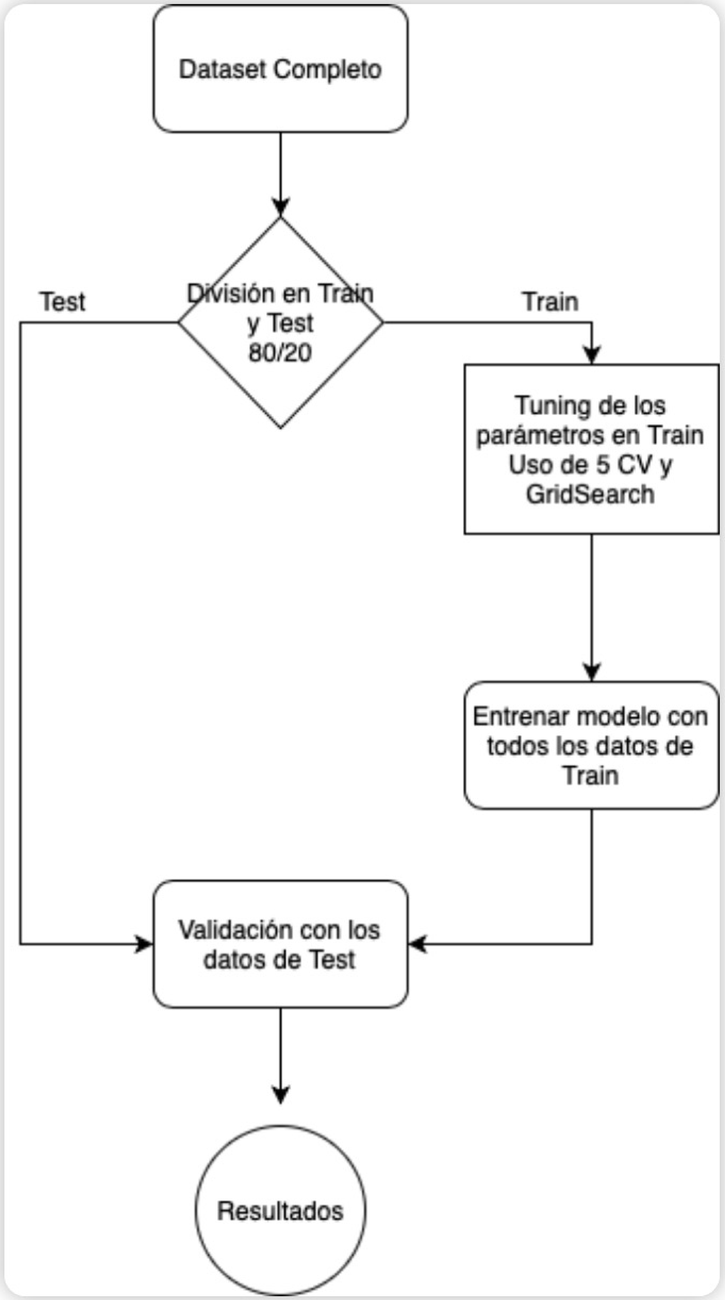
4. Predicción usando **regresión** con las **variables psicológicas** del problema.

5. Estudio del problema con técnicas de **aprendizaje no supervisado**

6. Creación de intervalos en **todas las variables a predecir**.

7. Predicción usando **clasificación multiclase** con **las variables propuestas** del problema.

8. Predicción usando **clasificación multiclase** con las **variables psicológicas** del problema.



Proceso de los experimentos

- Normalización de los datos.
- División en train y test.
- Búsqueda en rejilla para los mejores parámetros en train.
- 5 cross-validation para train.
- Creación del modelo con los mejores parámetros.
- Obtención de resultados en test.

Modelos de predicción para regresión y clasificación usados

1. **Regresión lineal.**
2. **Support Vector Machine (SVM)**
3. **Nearest Neighbour (kNN)**
4. **Decision Tree**
5. **Random Forest**
6. **Multi-Layer Perceptron**

6 Resultados en Regresión

Breve resumen para los resultados obtenidos por nuestros modelos en el apartado de regresión

Resultados para la variable Energía total

	Train			Test			
	MSE ↓	**R2 ↑**	**EVC ↑**	**MSE ↓**	**R2 ↑**	**EVC ↑**	**Params**
Linear Regression	0,008478	0,604443	0,608108	0,010314	0,492449	0,493395	None
Nearest Neighbor Regression	0,010809	0,504114	0,510620	0,012304	0,394555	0,394796	n_neighbors: 10
SVM Regressor	0,008460	0,607358	0,609927	0,010368	0,489823	0,491073	Kernel: rbf, C: 1
Decision Tree Regressor	0,017556	0,173524	0,176278	0,019163	0,057026	0,065202	Criterion: mae
Random Forest Regressor	0,008421	0,609131	0,611884	**0,009711**	**0,522119**	**0,525499**	Criterion: mae, n_estimators: 500
MLP Regressor	0,008334	0,611917	0,614761	0,009982	0,508823	0,510500	Hidden_layers: 5x2, Activation: Identity, Solver: lbfgs

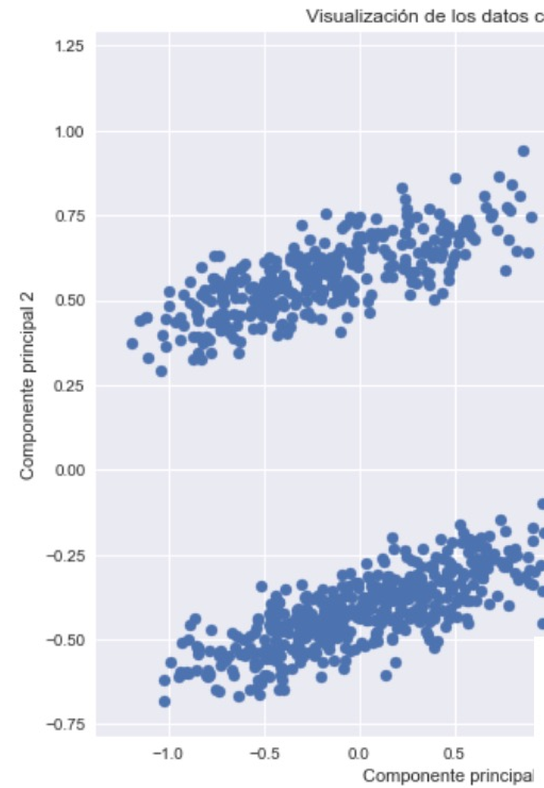
- Mejor modelo: Random Forest.
- Mejores resultados usando solamente las variables psicológicas.
- Amplio rango de configuraciones probadas en la red neuronal.

7 Análisis con aprendizaje no supervisado

Se realizan diversos análisis con aprendizaje no supervisado para intentar obtener más información

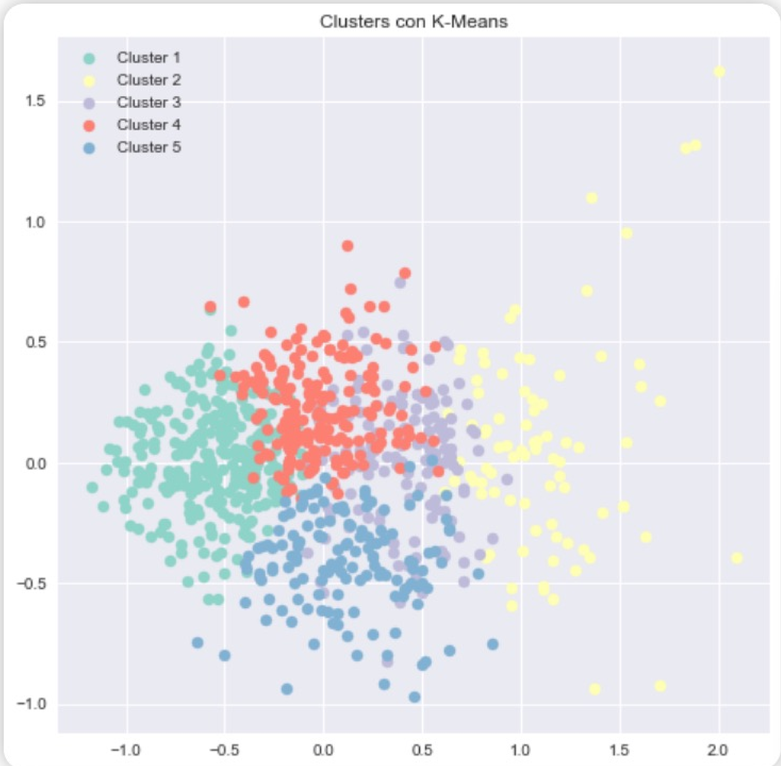
Visualización de los datos con PCA

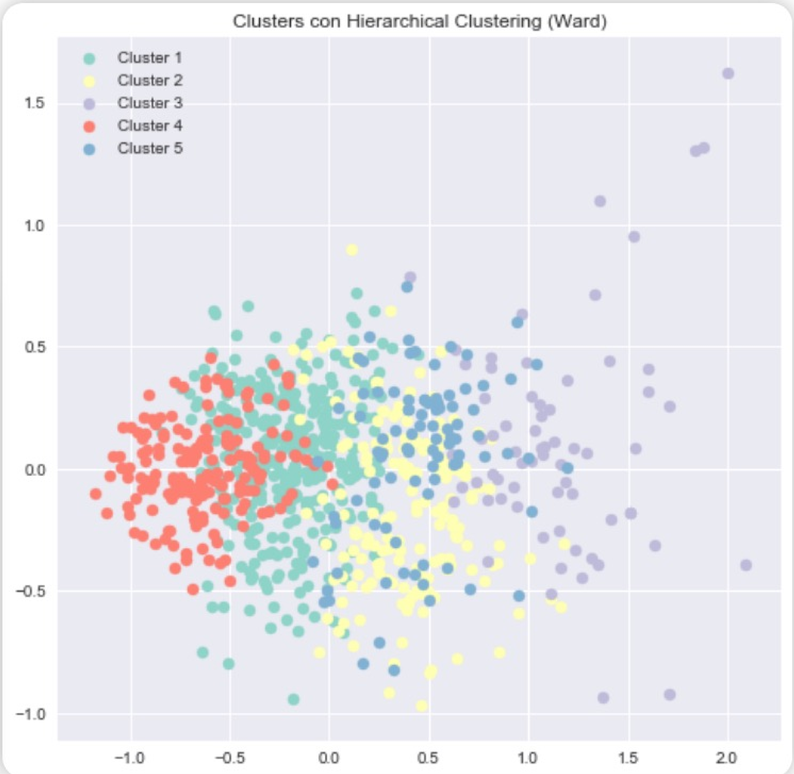
- Eliminación de la variable género debido a los clusters creados.
- Eliminación de las variables sociales.
- Se observa una gran agrupación de las instancias.



Creación de clusters

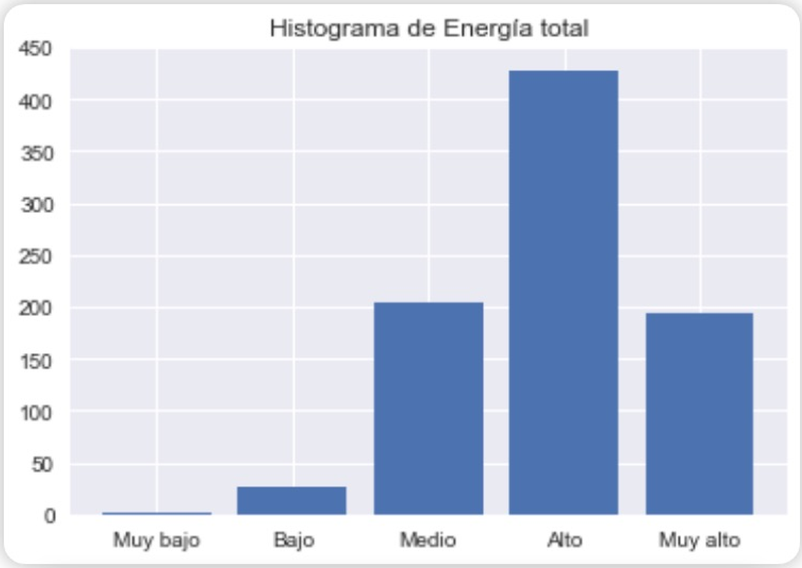
Obtención de malos resultados en las métricas para los algoritmos.

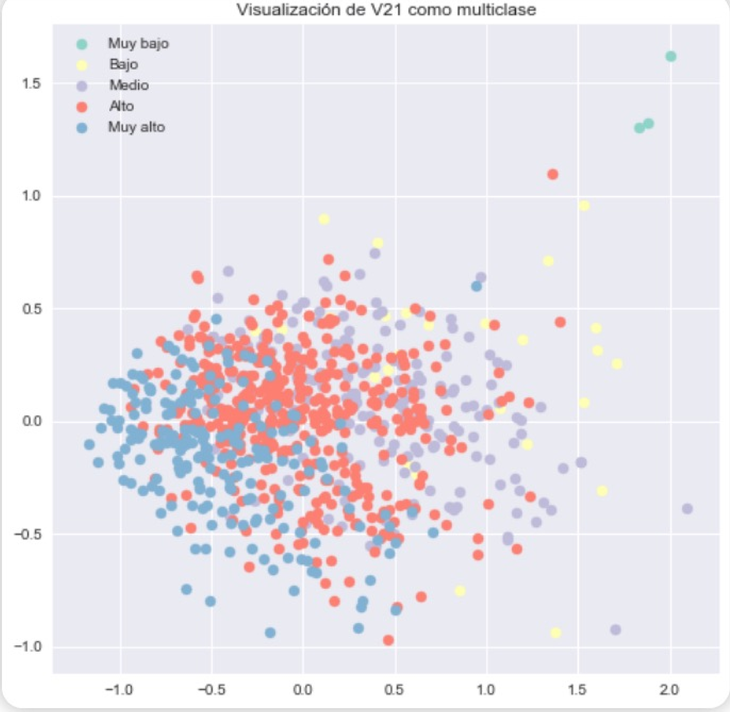




8 Aplicación de técnicas de clasificación

Resultados obtenidos por las técnicas de clasificación





Creación de intervalos para Energía Total

- Poca representación de los valores bajos.
- Una buena división en los valores altos.
- División general establecida para todas las variables a predecir.

Resultados obtenidos

	Train				Test				
	Accuracy ↑	**Recall ↑**	**Precision ↑**	**Hamming Loss ↓**	**Accuracy ↑**	**Recall ↑**	**Precision ↑**	**Hamming Loss ↓**	**Params**
SVM	0,640425	0,426733	0,460471	0,359575	0,662791	0,471259	0,517594	0,337209	kernel: poly, C: 10
kNN	0,162816	0,110833	0,128177	0,087104	0,616279	0,417864	0,488040	0,383721	k: 17
Decision Tree	0,144747	0,149641	0,147519	0,104853	0,575581	0,420340	0,411512	0,424419	Criterion: entropy
Random Forest	0,649098	0,504732	0,609018	0,350902	**0,691860**	**0,542893**	**0,784028**	**0,308140**	Criterion: entropy, n_estimators: 100
MLP	0,660842	0,518109	0,547735	0,339158	0,686047	0,504949	0,519287	0,313953	Hidden_layers: 5x20, Activation: tanh, Solver: Adam

- Casi **70%** de Accuracy en test.
- Mejor modelo Random Forest.
- Buena puntuación en *Precision*.
- Buena puntuación de la red neuronal.

9 Conclusiones finales y trabajo futuro

Conclusiones finales

1. Uso de los árboles de **Random Forest** para obtener conclusiones
2. Poca importancia de las **variables sociales** en los aspectos psicológicos
3. Gran relación entre la **Inteligencia emocional** y la energética
4. Gran importancia de la **Autoeficacia**, **Autoestima** y **Responsabilidad** en las predicciones
5. Uso de la **Calidad Personal y Organizacional** del individuo en la diferenciación entre clases similares
6. **Mejora de resultados** en las predicciones **sin variables sociales**

Trabajo futuro

1. Mejora de los resultados en **clasificación** y **regresión**
2. Uso de técnicas de **Oversampling** en el proceso
3. Uso de técnicas de **Feature Selection**
4. Aplicar las **conclusiones** obtenidas en este estudio a mejoras en el **trabajo futuro**
5. Obtener un **mayor número** de muestras de otros sectores de la población.

¡Gracias!