

Data Day 2023



Analítica y Ciencia de Datos y Machine Learning para Seguridad, Ergonomía y Bienestar

Juan Baldemar Garza V. PhD.

Practitioner

.... Y académico....

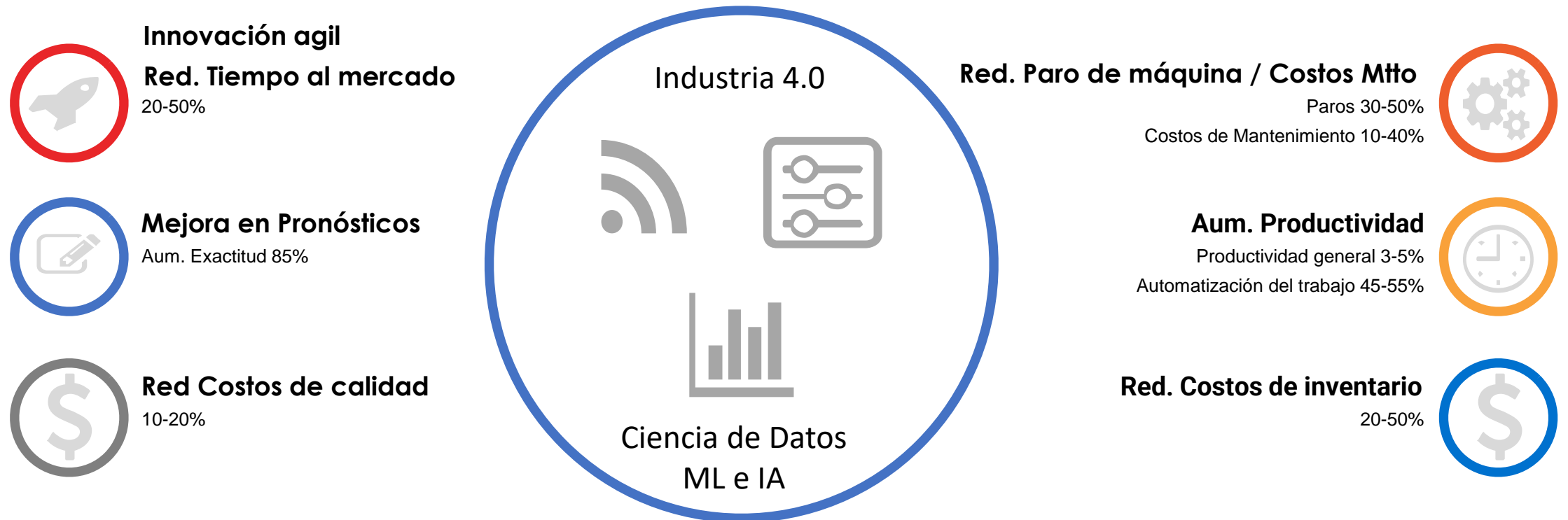
... Interesado en la Mejora Continua, Ciencia de Datos, IA, Tecnología,
Estrategia 4.0

y.....

... MEJORES EMPRESAS...

Contexto de DS en las empresas

Fuente: McKinsey & Company. Industry 4.0 at McKinsey's model factories.



¿Y la Seguridad Industrial?

Basado en la literatura se sabe que las condiciones ambientales como temperatura, humedad, presión atmosférica, iluminación, ruido, carga de trabajo, nivel de energía, estrés, fatiga y aspectos ergonómicos pueden tener un impacto notorio sobre la probabilidad de cometer un acto inseguro en la actividad laboral que se traduzca en un accidente de trabajo.

O bien tener un impacto en la productividad o eficiencia.

El caso de este estudio es presentar un par de proyectos realizado en una empresa de manufactura haciendo uso de la Ciencia de Datos, Analítica y Machine Learning para anticipar este tipo de situaciones.

¿Qué es Seguridad Industrial?

Ofrecer un entorno seguro donde el operador cuente con la suficiente confianza de realizar su trabajo, teniendo el apoyo de la compañía



Integridad Personal



Monitoreo de la Salud de los empleados



Bienestar del empleado



Anticipar Eventos

Caso 1

- Al tener un modelo predictivo se hace una diferencia entre **tener una actitud reactiva** en la que frecuentemente es **tarde para actuar**, a pasar a una **actitud proactiva, anticipándose a eventos**.
- La analítica predictiva permite extraer **conclusiones confiables** sobre eventos futuros mediante un análisis de **hechos históricos y actuales** para **predecir eventos**.



¿Cómo puedo anticipar eventos?

En una empresa de manufactura

2020

18 accidentes

(Incapacitantes y no Incapacitantes)



2021

25 accidentes

(Incapacitantes y no Incapacitantes)

En consecuencia...

 **40%**

Pagos de Sustitución del personal + pagos a clínicas privadas

 **10%**

Prima de Índice de Riesgo



Problemática

General

“Proveer un modelo predictivo para estimar probabilidades de riesgos de seguridad para la toma de decisiones en una empresa de manufactura.”

Particulares

- **Identificar las variables predictivas relevantes para la elaboración del modelo, mediante la recopilación de datos y análisis bibliográfico.**
- **Diseñar y desarrollar un modelo predictivo de riesgos de seguridad.**
- **Implementar el modelo en una plataforma para realizar predicciones correspondientes.**
- **Establecer criterios para la toma de decisiones mediante los resultados del modelo.**



Cross- Industry Standard Process for Data Mining



• Metodología

Seguridad Industrial 88%

Predictive Analytics 64%

Data Mining 58%

Machine Learning 58%

Big Data 58%

Business Analytics 52%

33 Artículos

Importancia de los tópicos de la literatura recabada

#	Variable	Apariciones en Artículos
1	Ruido	24%
2	Temperatura	
3	Postura y ergonomía	21%
4	Equipo de protección adecuado	
5	Horas de capacitación	18%
6	Condiciones inseguras y métricas de seguridad	
7	Iluminación	15%
8	Mantenimiento de las máquinas a tiempo	
9	Experiencia del trabajador	12%
10	Edad del operador	
11	Conocimiento del operador sobre la máquina	9%
12	Factores personales/psicológicos	
13	Demanda excesiva (sobrecarga)	6%
14	Horas de trabajo extra	
15	Manipulación de químicos	3%
16	Rotación de personal	
17	Ventilación	3%
18	Cantidad de errores cometidos por operadores	
19	Equipo respiratorio adecuado	3%
20	Existen procedimiento de trabajo seguro	
21	Herramienta en condiciones adecuadas	3%
22	Hidratación	
23	Inspecciones rutinarias	3%
24	Falta de comunicación adecuada	
25	Poco tiempo para realizar tareas	3%
26	Registros médicos	
27	Supervisor cumple su trabajo	3%
28	Velocidades inseguras	
29	Tipo de contrato del empleado	3%

Comprensión de los datos

- Ruido (Co)
- **Hidratación (D)**
- Iluminación (Co)
- Horas extras (D)
- Temperatura (Co)
- **EPP adecuado (C)**
- Enfermedades (D)
- **Espacio Suficiente (C)**
- Edad del operador (D)
- Años en la máquina (D)
- **Contacto Sust. Químicas (D)**
- **Fatigada del operador (D)**
- **Herramienta adecuada (C)**
- Antigüedad en la empresa (D)
- **Conocimientos suficientes (D)**
- **Preocupación del empleado (C)**
- Horas de descanso trabajado (Co)
- Horas de capacitación de seguridad (Co)
- Horas de capacitación de producción (Co)
- **Tiempo suficiente para realizar tareas (D)**
- **Iluminación percibida por el empleado (D)**

Features

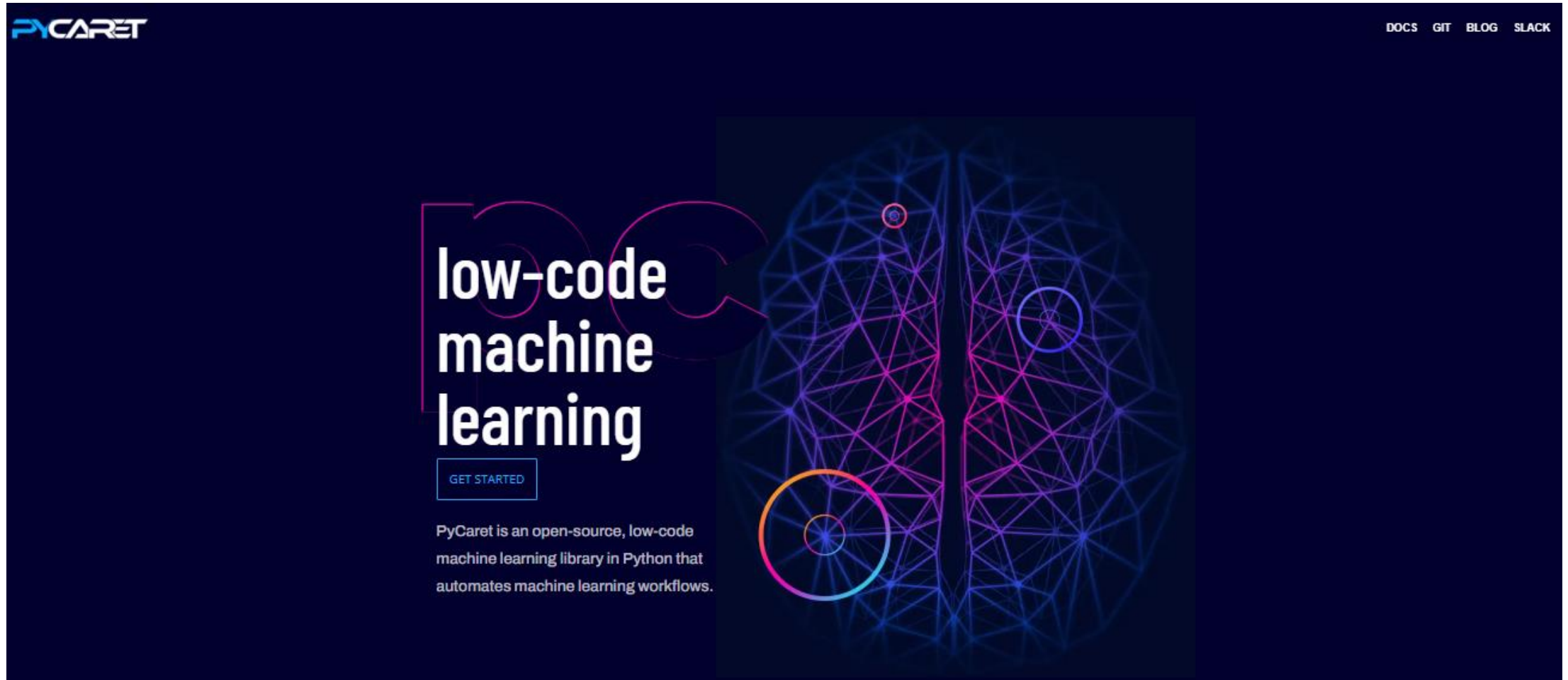
- + Cliente
- + Dep. Seguridad
- + Dep. Médico

Sugeridas

- Categoría del empleado (D)
- Calificación examen teórico (D)
- Nivel de riesgo por máquina (D)
- Calificación examen práctico (D)
- Actividades correctivas cerradas (D)

Modelado



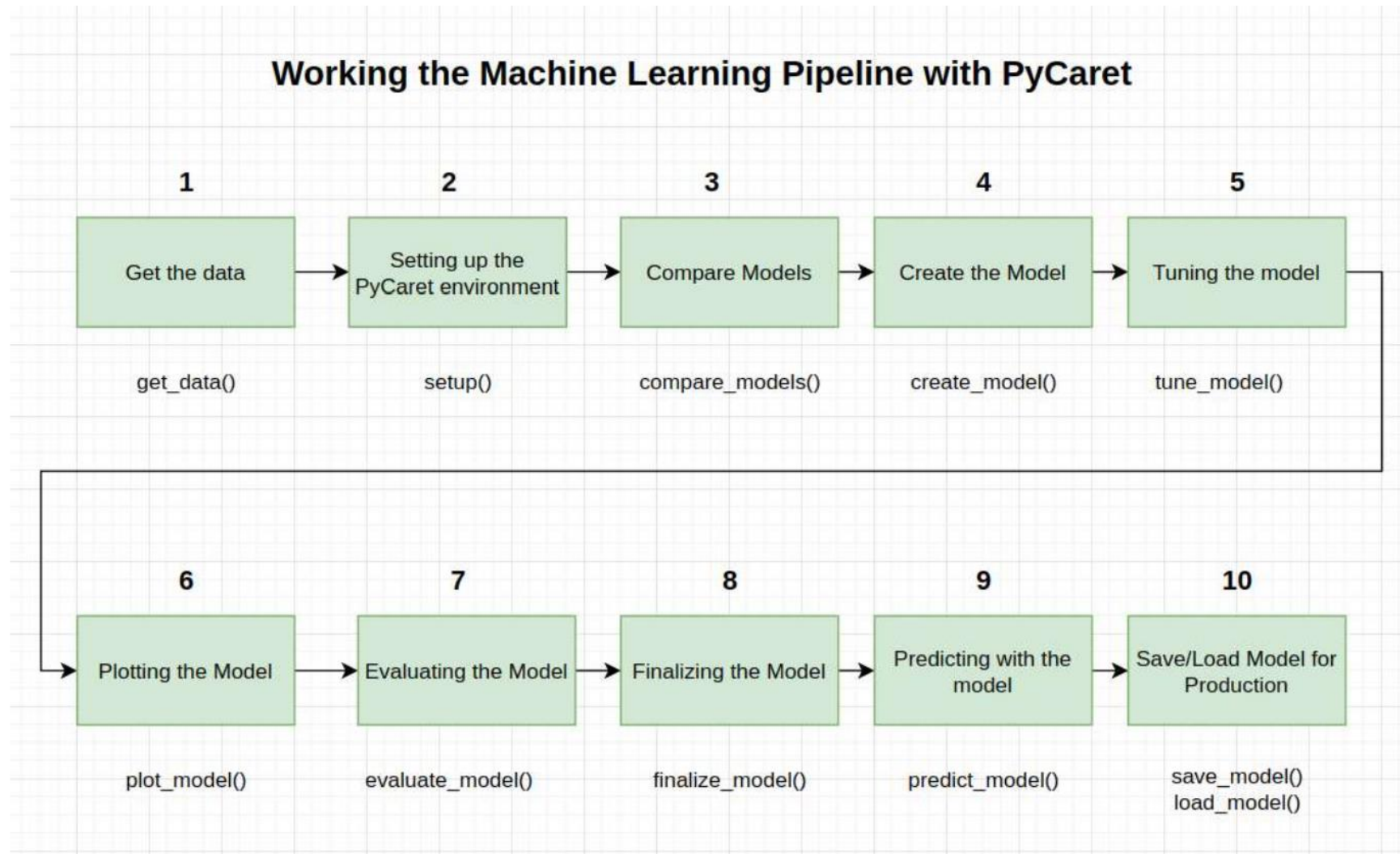


The image shows a screenshot of the PyCaret website landing page. The page has a dark blue background. In the top left corner, the 'PYCARET' logo is displayed in white. In the top right corner, there are links for 'DOCS', 'GIT', 'BLOG', and 'SLACK'. The main content area features the text 'low-code machine learning' in a large, white, sans-serif font. Below this text is a 'GET STARTED' button with a white border and blue text. Underneath the button, a short paragraph describes PyCaret as an open-source, low-code machine learning library in Python that automates machine learning workflows. To the right of the text, there is a stylized graphic of a human brain composed of a network of glowing blue and purple lines, with several circular highlights in various colors (red, blue, orange) overlaid on it.

low-code machine learning

[GET STARTED](#)

PyCaret is an open-source, low-code machine learning library in Python that automates machine learning workflows.



In [6]: `df.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 855 entries, 0 to 854
Data columns (total 34 columns):
#   Column      Non-Null Count  Dtype
---  -
0   ID           855 non-null   object
1   EDAD         855 non-null   int64
2   A_MAQ        855 non-null   int64
3   ANT          855 non-null   int64
4   CAT_EMP      855 non-null   int64
5   HORAS_CAP    855 non-null   int64
6   CAL_EXTEOR   855 non-null   int64
7   CAL_EXPRAC   855 non-null   int64
8   ENF          855 non-null   int64
9   RIESG        855 non-null   float64
10  ILLUM_TOM    855 non-null   int64
11  RUIDO        855 non-null   int64
12  HORAS_EXTR   855 non-null   float64
13  HORAS_DES    855 non-null   float64
14  LIQUID       855 non-null   int64
15  EPP          855 non-null   int64
16  HERRAM       855 non-null   int64
17  PREOCUP      855 non-null   int64
18  SUST_QUI     855 non-null   int64
19  ILLUM_ENC    855 non-null   int64
20  ESPA_SUF     855 non-null   int64
21  FATIGA       855 non-null   int64
22  TIEMPO_SUF   855 non-null   int64
23  CONOC        855 non-null   int64
24  TEMP         855 non-null   float64
25  ACC          855 non-null   int64
26  ACA          855 non-null   int64
27  CI           855 non-null   int64
28  AI           855 non-null   int64
29  CI_1         855 non-null   int64
30  AI_1         855 non-null   int64
31  AI_LOG       855 non-null   int64
32  CI_LOG       855 non-null   int64
33  ACA_LOG      855 non-null   int64
dtypes: float64(4), int64(29), object(1)
memory usage: 227.2+ KB
```


Pycaret – Python

```
In [21]: ▶ best = compare_models()
```

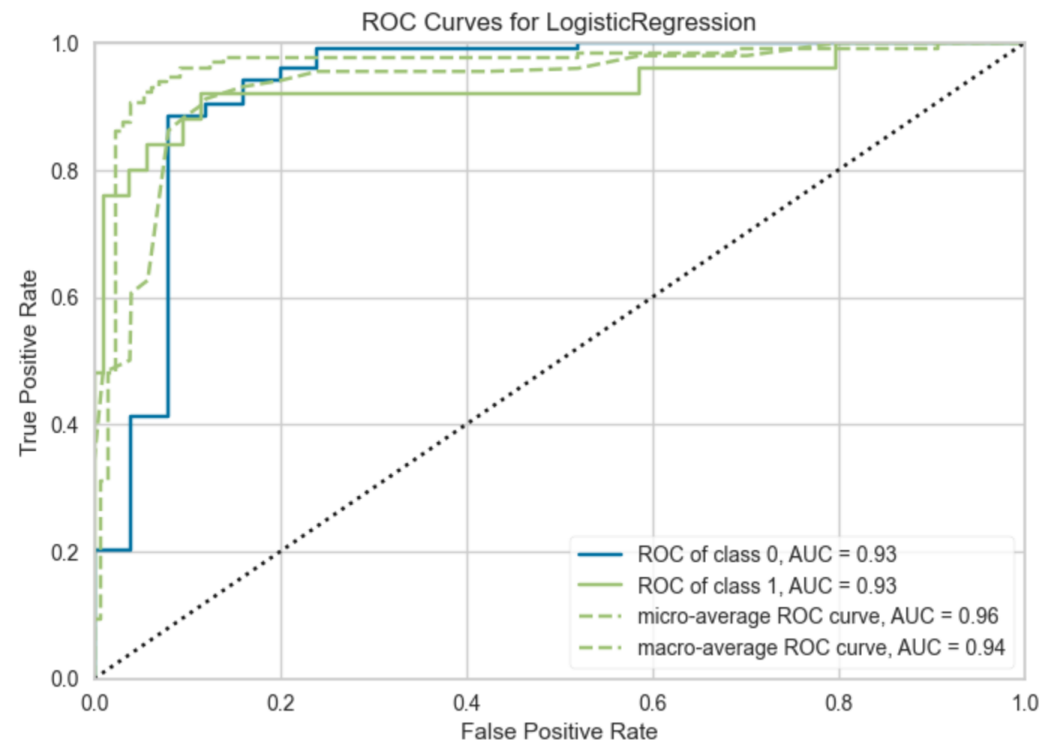
	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	TT (Sec)
nb	Naive Bayes	0.9466	0.8814	0.7300	1.0000	0.8374	0.8071	0.8254	0.0130
ridge	Ridge Classifier	0.9366	0.0000	0.7467	0.9405	0.8207	0.7835	0.7986	0.0120
lda	Linear Discriminant Analysis	0.9299	0.8619	0.7467	0.9038	0.8059	0.7643	0.7770	0.0120
xgboost	Extreme Gradient Boosting	0.9299	0.8794	0.7467	0.9133	0.8106	0.7689	0.7816	0.0380
rf	Random Forest Classifier	0.9266	0.8885	0.7300	0.8971	0.7971	0.7534	0.7641	0.0880
lightgbm	Light Gradient Boosting Machine	0.9266	0.8640	0.7467	0.8848	0.7970	0.7533	0.7656	0.2090
catboost	CatBoost Classifier	0.9266	0.8853	0.7300	0.8900	0.7946	0.7510	0.7609	2.0990
lr	Logistic Regression	0.9132	0.8474	0.7467	0.8411	0.7768	0.7243	0.7354	0.5780
gbc	Gradient Boosting Classifier	0.9099	0.8571	0.7300	0.8267	0.7642	0.7095	0.7187	0.0460
et	Extra Trees Classifier	0.9099	0.8770	0.7300	0.8233	0.7589	0.7047	0.7164	0.0690
ada	Ada Boost Classifier	0.9032	0.8585	0.7300	0.8048	0.7451	0.6871	0.7020	0.0390
dt	Decision Tree Classifier	0.8697	0.8299	0.7633	0.6444	0.6892	0.6084	0.6194	0.0160
qda	Quadratic Discriminant Analysis	0.8061	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0140
dummy	Dummy Classifier	0.8061	0.5000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0110
svm	SVM - Linear Kernel	0.7154	0.0000	0.4800	0.3677	0.3590	0.2482	0.2710	0.0150
knn	K Neighbors Classifier	0.6422	0.6366	0.5733	0.2955	0.3859	0.1741	0.1923	0.3450

Pycaret – Python

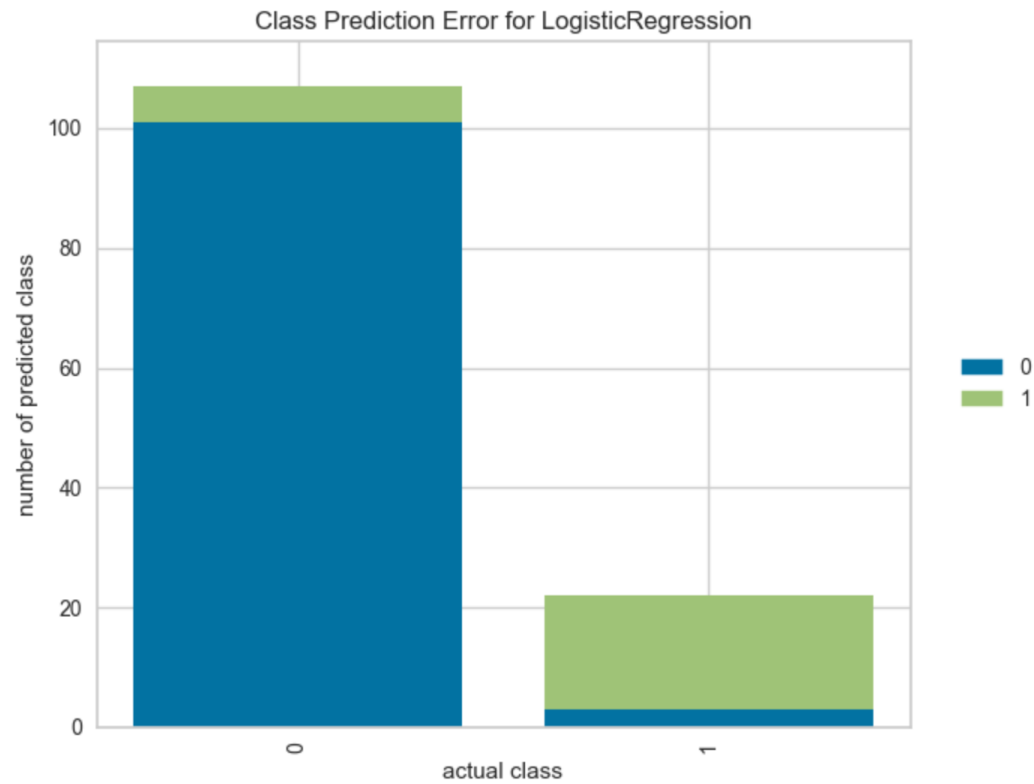
```
In [22]: ► lr= create_model('lr')
```

	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC
Fold							
0	0.9000	0.6319	0.5000	1.0000	0.6667	0.6154	0.6667
1	0.9333	0.9583	0.8333	0.8333	0.8333	0.7917	0.7917
2	0.9000	0.7917	0.6667	0.8000	0.7273	0.6667	0.6708
3	0.8667	0.8125	0.6667	0.6667	0.6667	0.5833	0.5833
4	0.9333	0.9097	0.6667	1.0000	0.8000	0.7619	0.7845
5	0.7667	0.8403	0.6667	0.4444	0.5333	0.3860	0.4001
6	0.9667	0.8542	0.8333	1.0000	0.9091	0.8889	0.8944
7	0.9333	0.8333	0.8333	0.8333	0.8333	0.7917	0.7917
8	0.9667	1.0000	1.0000	0.8333	0.9091	0.8889	0.8944
9	0.9655	0.8417	0.8000	1.0000	0.8889	0.8688	0.8764
Mean	0.9132	0.8474	0.7467	0.8411	0.7768	0.7243	0.7354
Std	0.0580	0.0952	0.1327	0.1705	0.1185	0.1533	0.1496

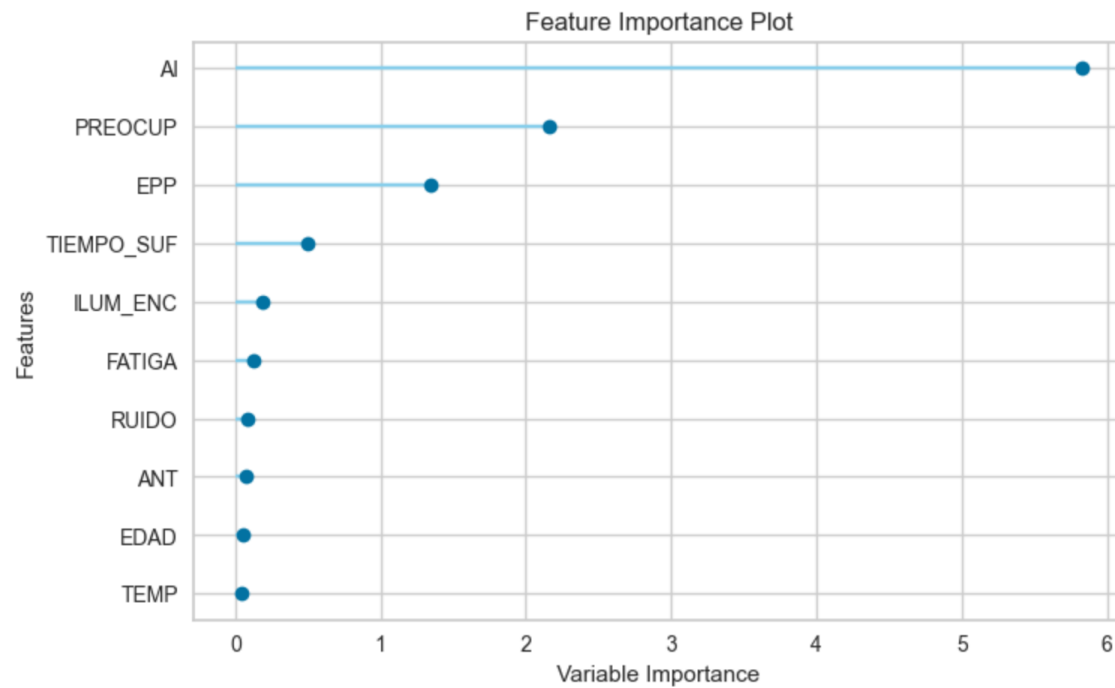
```
In [24]: plot_model(lr)
```



```
In [25]: plot_model(lr, plot="error")
```



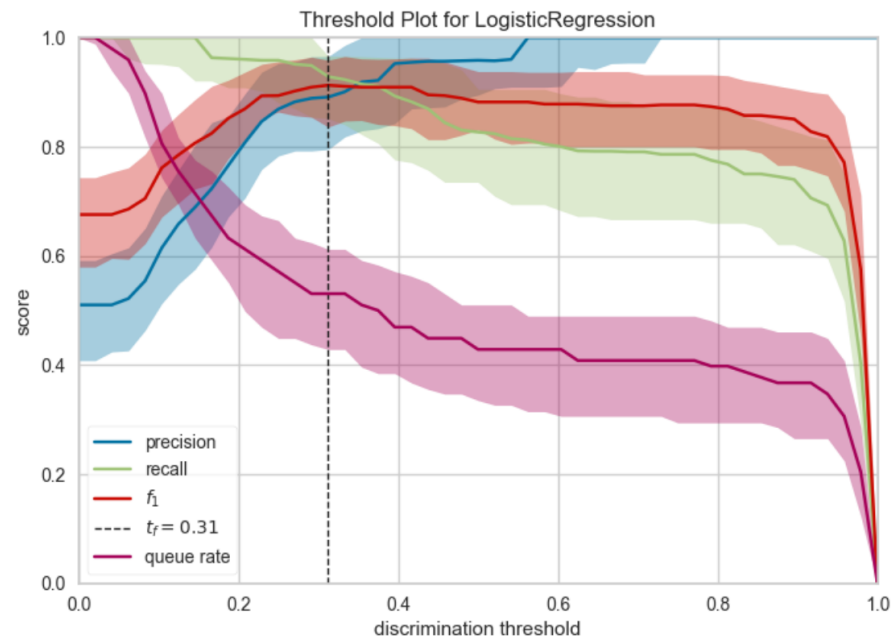
```
In [26]: plot_model(lr, plot="feature")
```



Pycaret – Python

In [27]: `evaluate_model(lr)`

Plot Type:	Pipeline Plot	Hyperparameters	AUC	Confusion Matrix	Threshold	Precision Recall	Prediction Error	Class Report	Feature Selection
	Learning Curve	Manifold Learning	Calibration Curve	Validation Curve	Dimensions	Feature Importance	Feature Importance...	Decision Boundary	Lift Chart
	Gain Chart	Decision Tree	KS Statistic Plot						



In [28]: `predict_model(lr)`

	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC
0	Logistic Regression	0.9302	0.9296	0.7600	0.8636	0.8085	0.7661	0.7684

Out[28]:

	EDAD	ANT	RUIDO	EPP	PREOCUP	ILUM_ENC	FATIGA	TIEMPO_SUF	TEMP	AI	ACA_LOG	prediction_label	prediction_score
74	29	7	83	1	0	4	4	4	34.000000	0	0	0	0.9228
335	31	1	84	1	0	4	2	5	35.599998	0	0	0	0.7704
68	45	19	84	1	1	4	5	3	35.299999	0	0	0	0.7267
264	26	4	88	1	0	4	3	4	37.700001	1	1	1	0.9627
360	28	1	85	1	0	4	2	4	35.099998	0	0	0	0.8776
...
4	31	1	84	0	1	2	4	4	34.900002	1	1	1	0.9935
283	35	7	85	1	0	3	3	3	34.900002	0	0	0	0.9217
150	48	21	86	1	0	5	3	4	35.400002	0	0	0	0.9384
211	36	6	84	1	0	5	2	5	37.200001	0	0	0	0.8074
2	31	1	83	1	0	4	1	4	34.500000	0	0	0	0.8230

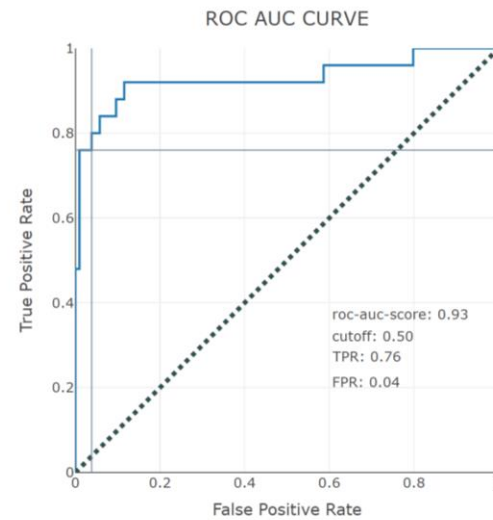
Python – Explainer Dashboard

Model performance metrics

metric	Score
accuracy	0.93
precision	0.864
recall	0.76
f1	0.809
roc_auc_score	0.93
pr_auc_score	0.882
log_loss	0.252

ROC AUC Plot

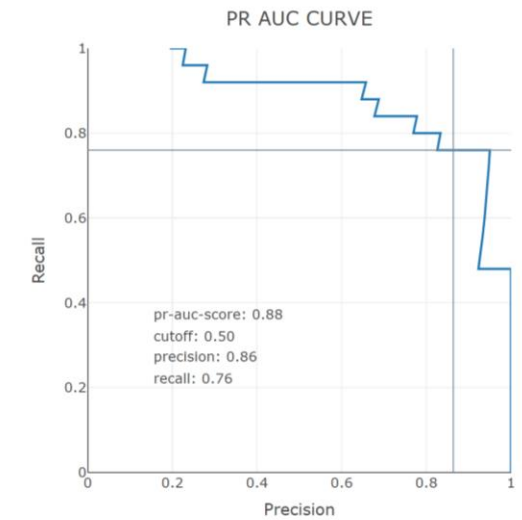
Trade-off between False positives and false negatives



Popout

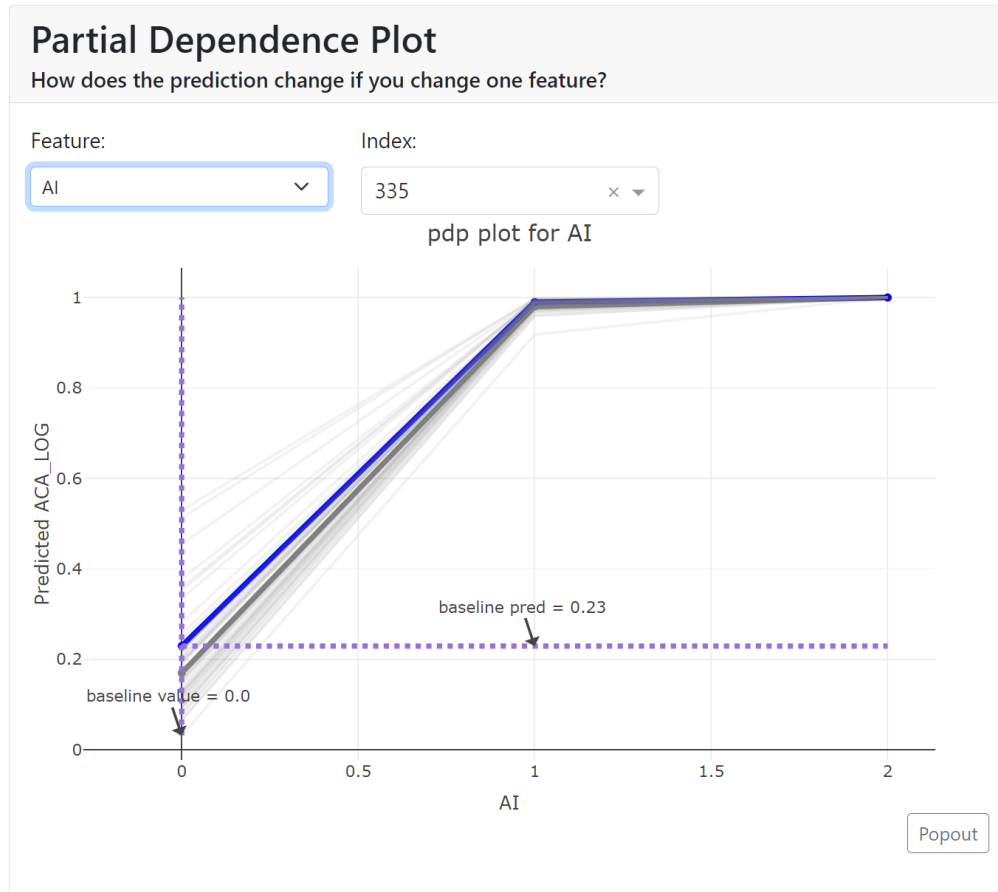
PR AUC Plot

Trade-off between Precision and Recall

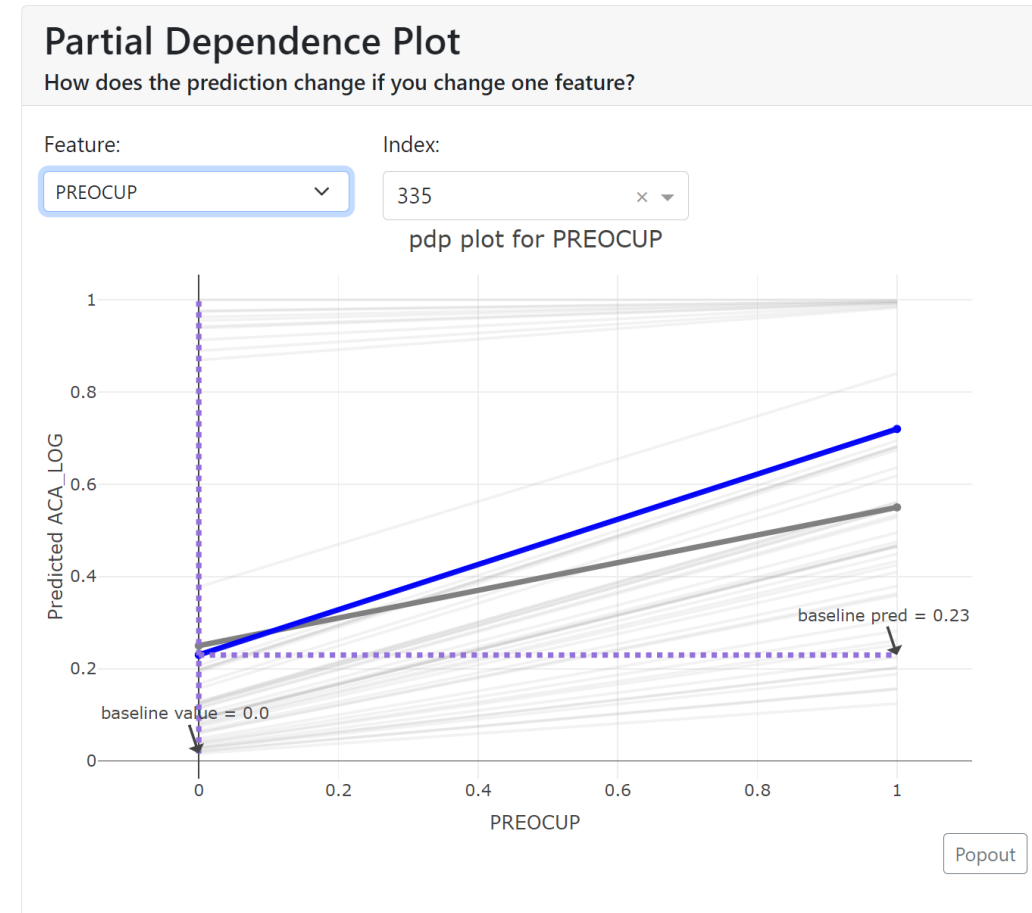


Popout

Python – Explainer Dashboard



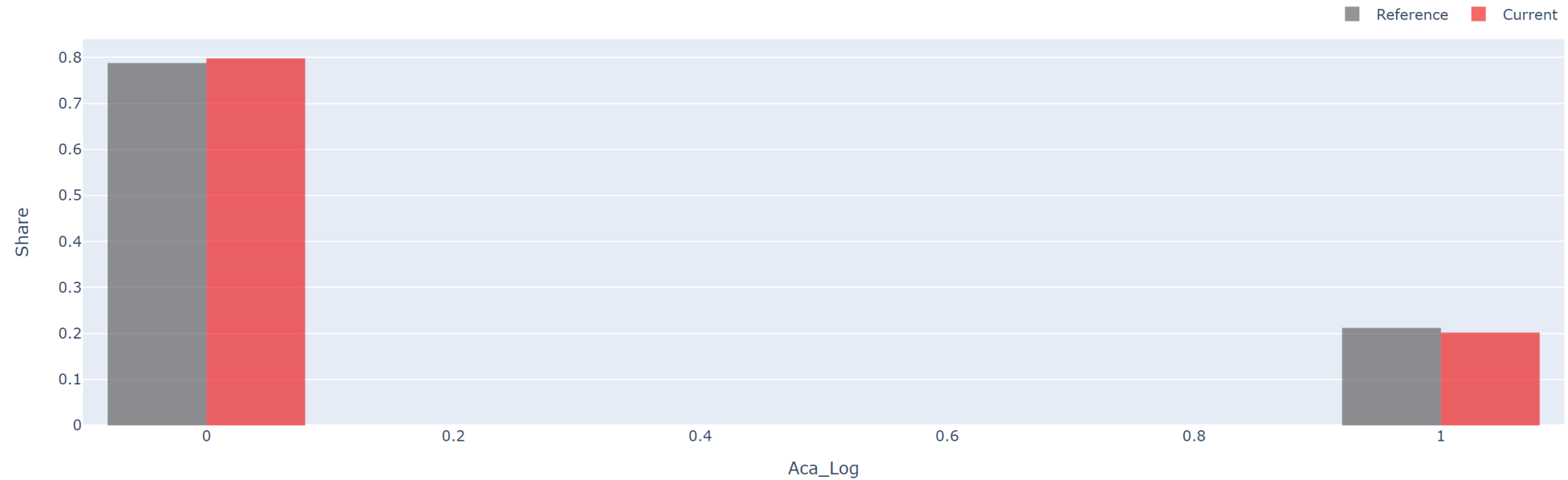
Acto Inseguro



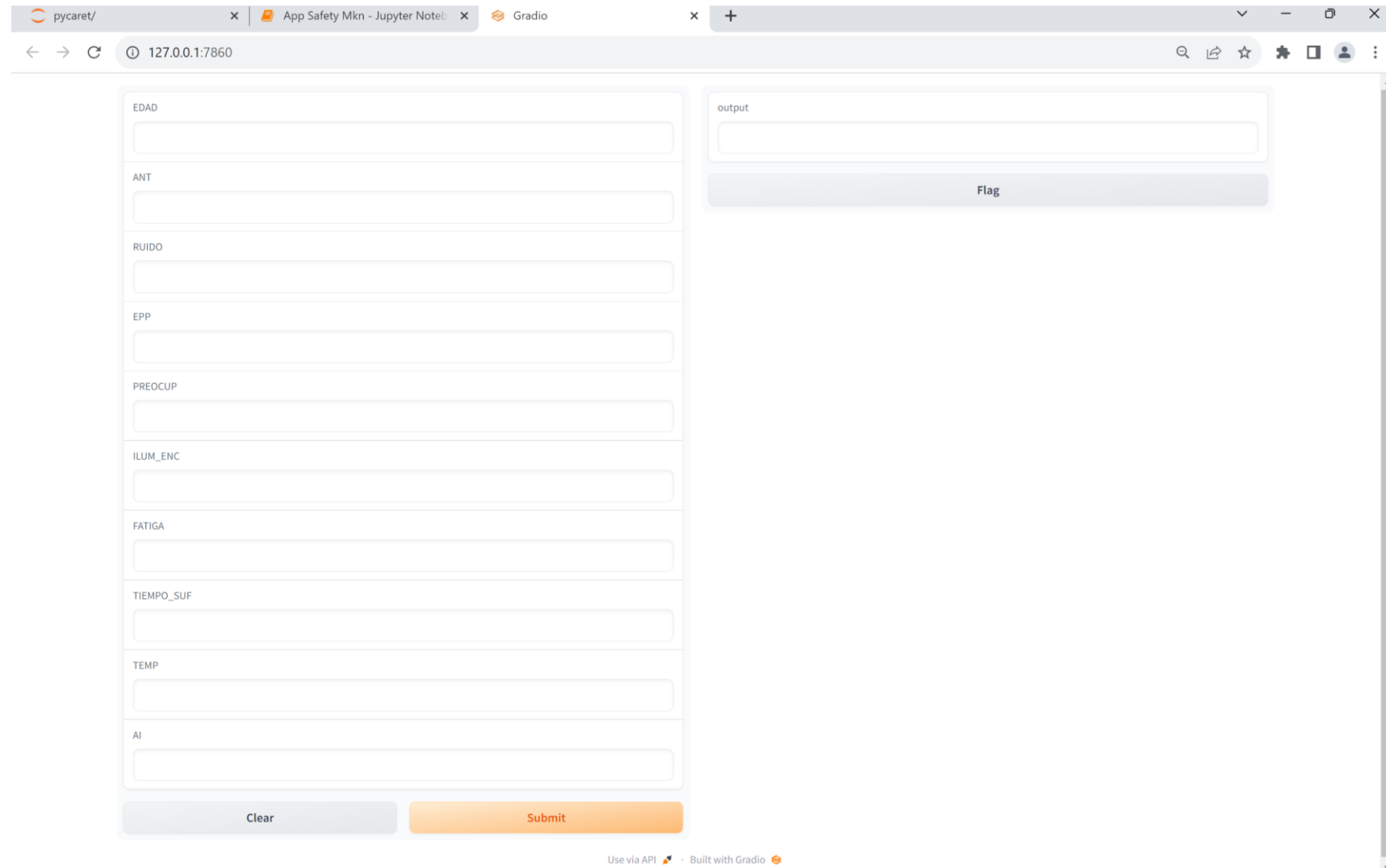
Preocupación

Python – Drift Target

Target Drift: not detected, drift score=0.696348 (Z-test p_value)



Python – Gradio – Test user



EDAD

ANT

RUIDO

EPP

PREOCUP

ILUM_ENC

FATIGA

TIEMPO_SUF

TEMP

AI

output

Flag

Clear

Submit

Use via API · Built with Gradio

Python – Fast API

FastAPI 0.1.0 OAS3

/openapi.json

default

POST /predict Predict

Parameters Try it out

No parameters

Request body *required* application/json

Example Value | Schema

```
{
  "EDAD": 35,
  "ANT": 7,
  "RUIDO": 84,
  "EPP": 1,
  "PREOCUP": 0,
  "ILUM_ENC": 3,
  "FATIGA": 3,
  "TIEMPO_SUF": 3,
  "TIEMPO": 34.20000076293945,
  "AI": 0
}
```

Responses

Code	Description	Links
200	Successful Response	No links

Media type application/json

Controls Accept header.

Example Value | Schema

```
{
  "prediction": 0
}
```

Mediante los resultados que realizó el modelo predictivo, se toman decisiones que favorecen la seguridad industrial.

- Reducción de accidentes – **De 25 accidentes a 0.**
- Reducción de costos por accidentes.
- Aumento de la productividad.
- Mantener la integridad de la persona.
- Enfoque proactivo.

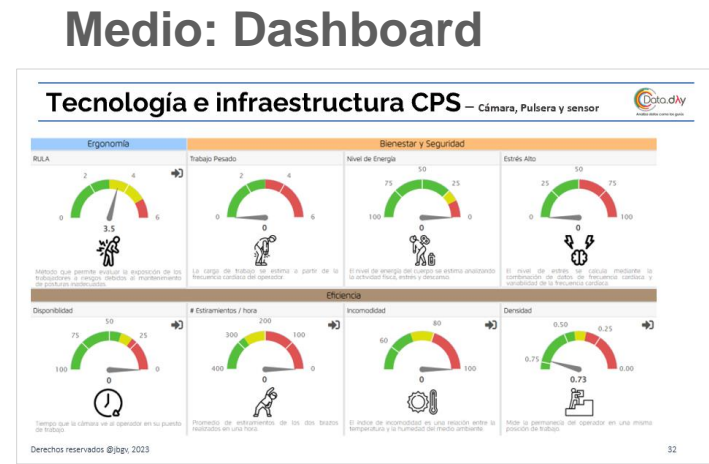
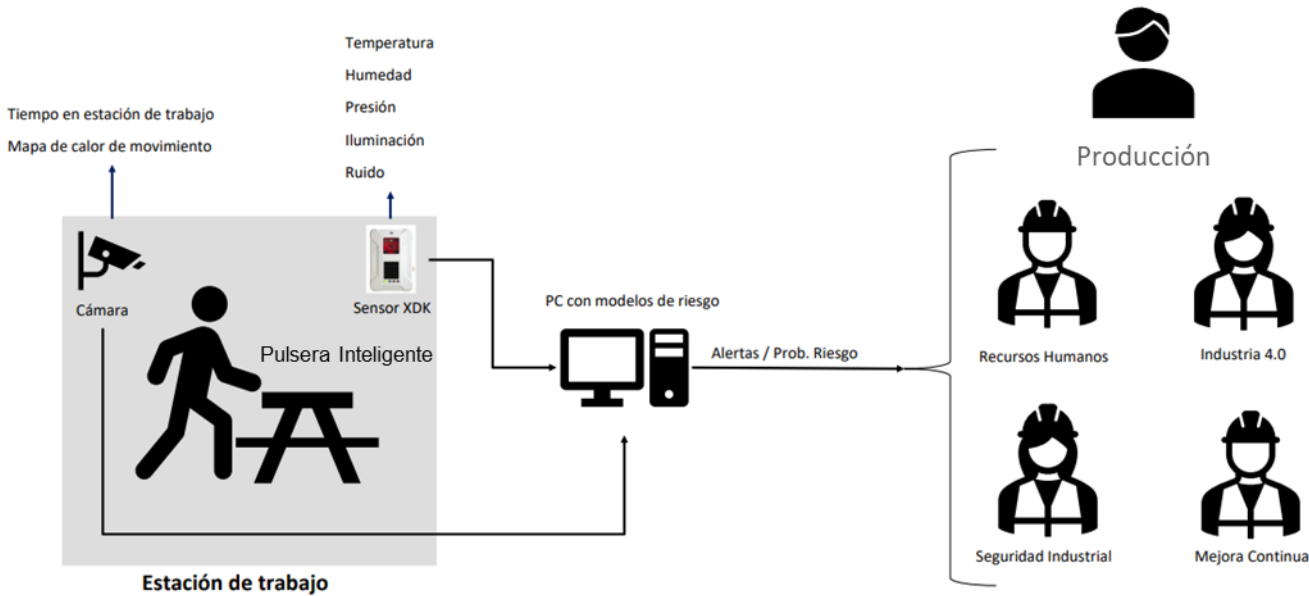
“Si eres capaz de predecir un accidente, eres capaz de prevenirlo.”

Caso 2 – Seguridad, Ergonomía y Bienestar

Tecnología e infraestructura CPS – Cámara, Pulsera y sensor

Basado en la literatura se sabe que las condiciones ambientales como temperatura, humedad, presión atmosférica, iluminación, ruido, tiempo de actividad de trabajo, estrés, carga de trabajo, nivel de energía de las personas y aspectos ergonómicos pueden tener un **impacto notorio sobre la eficiencia** en una estación de trabajo. Incluso **llevando a los colaboradores a incrementar el riesgo de una lesión y la probabilidad de cometer un acto inseguro** en la actividad laboral.

Objetivo contar con Información en tiempo real que nos permita anticipar riesgos:



Output

- Seguimiento EHS
- Seguimiento Producción

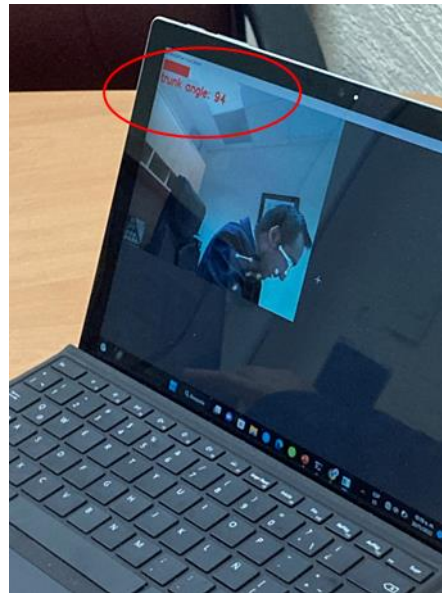
Se busca evaluar la tecnología para revisar factibilidad técnica y económica de transferir a otras líneas o estaciones de trabajo una solución

RULA método. ¿Quién lo desarrolló?

La adopción continuada o repetida de posturas inadecuadas durante el trabajo genera fatiga y a la larga puede ocasionar trastornos en el sistema musculoesquelético. Esta carga estática o postural es uno de los factores a tener en cuenta en la evaluación de las condiciones de trabajo, y su reducción es una de las medidas fundamentales a adoptar en la mejora de puestos.

Fue desarrollado por los doctores McAtamney y Corlett de la Universidad de Nottingham en 1993 (Institute for Occupational Ergonomics)

Este método es utilizado para evaluar la exposición de los trabajadores a factores de riesgo que pueden ocasionar trastornos en los miembros superiores del cuerpo: posturas, repetitividad de movimientos, fuerzas aplicadas, actividad estática del sistema musculoesquelético.



Exploración de algoritmos de malas posturas

Aplicable a oficinas o piso de trabajo. Ej. Baldemar con mala postura.

Método Rula – Seleccionado.

Se revisaron 2 métodos REBA y RULA – Método elegido fue RULA

RULA – Evaluación – Niveles de riesgos

A. Arm and Wrist Analysis

Step 1: Locate Upper Arm Position:

Step 1a: Adjust...
 If shoulder is raised: +1
 If upper arm is abducted: +1
 If arm is supported or person is leaning: -1

Step 2: Locate Lower Arm Position:

Step 2a: Adjust...
 If either arm is working across midline or out to side of body: Add +1

Step 3: Locate Wrist Position:

Step 3a: Adjust...
 If wrist is bent from midline: Add +1

Step 4: Wrist Twist:

If wrist is twisted in mid-range: +1
 If wrist is at or near end of range: +2

Step 5: Look-up Posture Score in Table A:
 Using values from steps 1-4 above, locate score in Table A

Step 6: Add Muscle Use Score
 If posture mainly static (i.e. held >10 minutes), Or if action repeated occurs 4X per minute: +1

Step 7: Add Force/Load Score
 If load < .4.4 lbs. (intermittent): +0
 If load 4.4 to 22 lbs. (intermittent): +1
 If load 4.4 to 22 lbs. (static or repeated): +2
 If more than 22 lbs. or repeated or shocks: +3

Step 8: Find Row in Table C
 Add values from steps 5-7 to obtain Wrist and Arm Score. Find row in Table C.

B. Neck, Trunk and Leg Analysis

Step 9: Locate Neck Position:

Step 9a: Adjust...
 If neck is twisted: +1
 If neck is side bending: +1

Step 10: Locate Trunk Position:

Step 10a: Adjust...
 If trunk is twisted: +1
 If trunk is side bending: +1

Step 11: Legs:
 If legs and feet are supported: +1
 If not: +2

Step 12: Look-up Posture Score in Table B:
 Using values from steps 9-11 above, locate score in Table B

Step 13: Add Muscle Use Score
 If posture mainly static (i.e. held >10 minutes), Or if action repeated occurs 4X per minute: +1

Step 14: Add Force/Load Score
 If load < .4.4 lbs. (intermittent): +0
 If load 4.4 to 22 lbs. (intermittent): +1
 If load 4.4 to 22 lbs. (static or repeated): +2
 If more than 22 lbs. or repeated or shocks: +3

Step 15: Find Column in Table C
 Add values from steps 12-14 to obtain Neck, Trunk and Leg Score. Find Column in Table C.

Scores

Table A: Wrist Score

Upper Arm	Lower Arm	Wrist Score						
		Wrist Twist 1	Wrist Twist 2	Wrist Twist 3	Wrist Twist 4			
1	1	1	2	2	2	3	3	3
1	2	2	2	2	2	3	3	3
1	3	2	3	3	3	3	3	4
2	1	2	3	3	3	3	4	4
2	2	3	3	3	3	3	4	4
2	3	3	4	4	4	4	5	5
3	1	3	3	4	4	4	5	5
3	2	3	4	4	4	4	5	5
3	3	4	4	4	4	4	5	5
4	1	4	4	4	4	4	5	5
4	2	4	4	4	4	4	5	5
4	3	4	4	4	5	5	6	6
5	1	5	5	5	5	5	6	6
5	2	5	6	6	6	6	7	7
5	3	6	6	6	7	7	7	8
6	1	7	7	7	7	8	8	9
6	2	8	8	8	8	9	9	9
6	3	9	9	9	9	9	9	9

Table B: Trunk Posture Score

Neck Posture Score	Legs											
	1	2	3	4	5	6						
1	1	1	2	3	3	4	5	5	6	6	7	7
2	2	3	3	4	4	5	5	5	6	7	7	7
3	3	3	3	4	4	5	5	6	6	7	7	7
4	5	5	5	6	6	7	7	7	7	8	8	8
5	7	7	7	7	8	8	8	8	8	8	8	8
6	8	8	8	8	8	8	9	9	9	9	9	9

Table C: Neck, Trunk, Leg Score

Wrist / Arm Score	Neck, Trunk, Leg Score							
	1	2	3	4	5	6	7+	
1	1	1	2	3	3	4	5	5
2	2	2	3	4	4	5	5	5
3	3	3	3	4	4	5	6	6
4	4	3	3	3	4	5	6	6
5	4	4	4	4	5	6	7	7
6	4	4	5	6	6	7	7	7
7	5	5	6	6	7	7	7	7
8+	5	5	6	7	7	7	7	7

Scoring: (final score from Table C)
 1-2 = acceptable posture
 3-4 = further investigation, change may be needed
 5-6 = further investigation, change soon
 7 = investigate and implement change

Información en tiempo real.

- Riesgo ergonómico RULA
- 1-2 Aceptable sin riesgo
- 3-4 Aceptable, pero se sugieren evaluar cambios en la operación
- 5-6 No aceptable, se sugieren cambios en corto plazo
- 7 No aceptable, cambios urgentes

RULA – Rapid Upper Limb Assessment - Evaluación de carga postural

Medición de bienestar en estación de trabajo.

Podemos evaluar el bienestar del colaborador con tres aspectos relevantes:

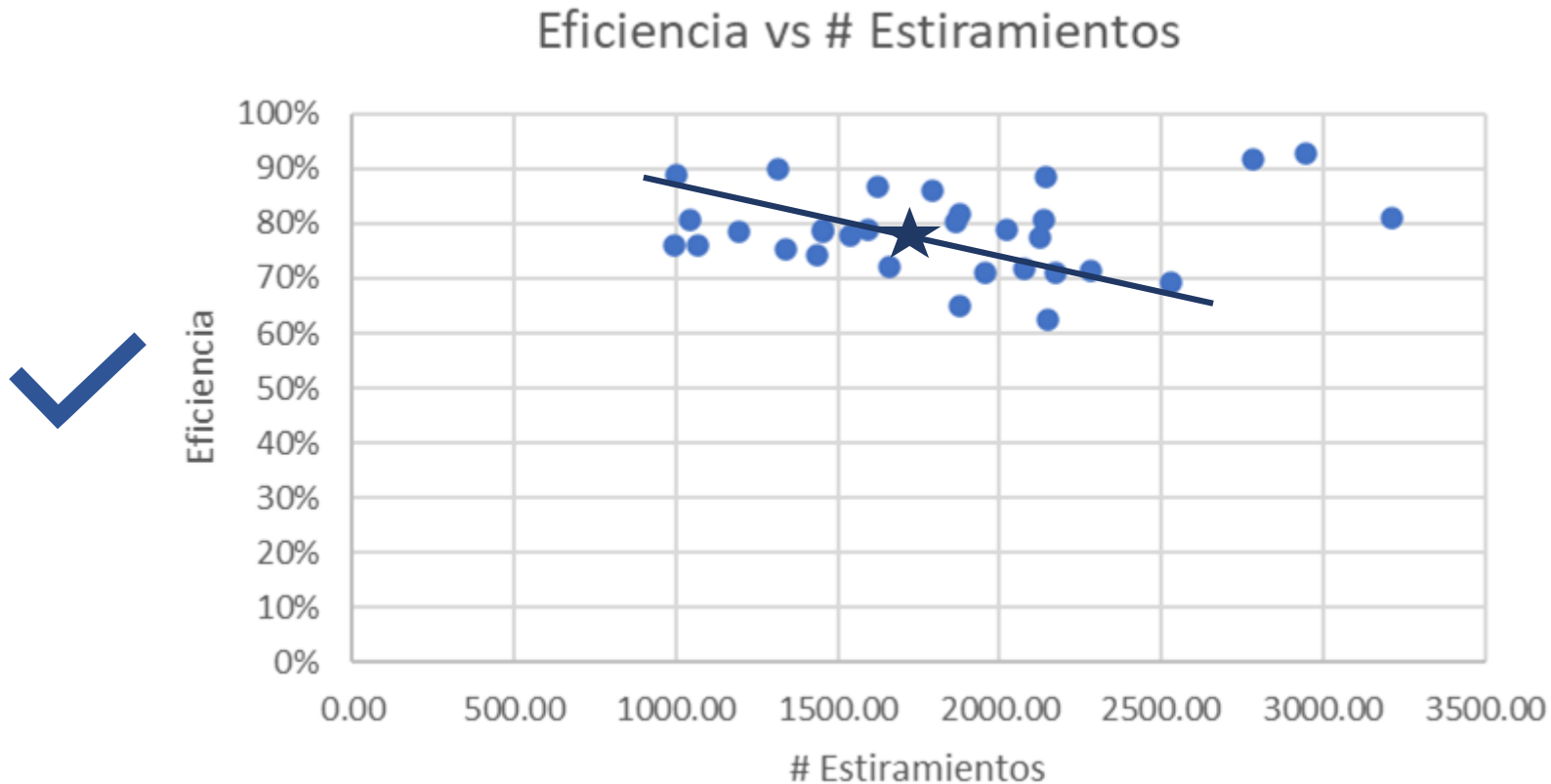
Carga de trabajo

Nivel de energía

Nivel de estrés

KPI	descripción	verde	amarillo	rojo
Carga de trabajo	La carga de trabajo se estima a partir de la frecuencia cardíaca, se considera trabajo moderado cuando hay más de 90 latidos por minuto y trabajo pesado cuando hay más de 110 latidos por minuto.	$x > 90$ (-50%)	$x > 90$ (+50%)	$x > 110$ (+3%)
Nivel de energía	La función Body Battery analiza la actividad física, estrés y descanso para informar sobre los niveles de energía del cuerpo.	$x \geq 25$	$25 > x > 15$	$x \leq 15$
Nivel de estrés	La función Estrés calcula los niveles de estrés (de 0 a 100) principalmente mediante la combinación de datos de frecuencia cardíaca y variabilidad de la frecuencia cardíaca.	$x > 50$ (-50%)	$x > 50$ (+50%)	$x > 75$ (+50%)

Nos alarmará cuando el operador tiene sobre carga de trabajo, nivel bajo de energía o alto estrés



- En el ciclo los estiramientos corresponden principalmente a la operación principal.
- Podemos ver que después de 1600 repeticiones la *eficiencia* disminuye por debajo del promedio que es 77%.
- En un cálculo muy general corresponde a 88 piezas, el promedio de piezas por persona por día es 112.

Mediante los resultados que desarrolló el modelo predictivo, se podrán tomar decisiones que favorezcan la seguridad, ergonomía y bienestar de los empleados.

- **Incremento estimado de aumento de eficiencia (5%)**
- **Reducción de lesiones por posturas incorrectas o actos inseguros.**
- **Bienestar de los empleados.**

- Se muestran 2 casos de uso aplicados a Manufactura de forma exitosa.
- Mucho potencial de Analítica, Data Science y ML en la Manufactura en México.
- Beneficios relevantes para las empresas y colaboradores.

Gracias

Juan Baldemar Garza V. PhD.

