

# Introducción a la Ciencia de Datos

Maestría en Ciencias  
de la Computación

Dr. Irvin Hussein López Nava



Dimensionality reduction



Feature selection



The slide features several decorative blue squares of varying shades (dark blue and light blue) scattered across the background. The main title is centered and reads:

## 2.5 Selección de Características



# ¿Qué es la selección de Características?

# Definiciones

## ChatGPT

Es un proceso en el análisis de datos y el aprendizaje automático que implica la elección de un subconjunto relevante de características. El objetivo eliminar las características que son redundantes o irrelevantes.

<https://chat.openai.com/>

## Wikipedia

Es el proceso de selección de un subconjunto de características relevantes para su uso en la construcción de modelos. Debe distinguirse de la extracción de características.

[https://en.wikipedia.org/wiki/Feature\\_selection](https://en.wikipedia.org/wiki/Feature_selection)

## Google Bard

Es un proceso de identificación y selección de un subconjunto de características. El objetivo es mejorar el rendimiento de un modelo de aprendizaje automático, como un modelo de clasificación, regresión o agrupamiento.

<https://bard.google.com/>



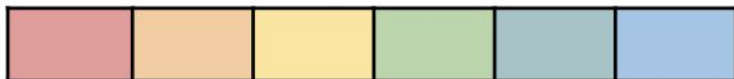
# ***Feature selection***

- Es el proceso de **seleccionar** las características más importantes para ingresar en los algoritmos de aprendizaje automático.
- Las técnicas de selección se emplean para **reducir la cantidad de variables** de entrada mediante la eliminación de características redundantes o irrelevantes.
  - Redundante e irrelevante son dos nociones distintas, ya que una característica relevante puede ser redundante en presencia de otra característica relevante con la que está fuertemente correlacionada.

Features



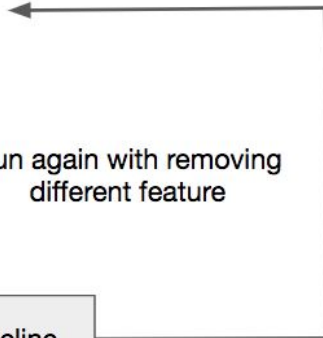
Remove Single Feature



Model Training

Compare to baseline

Run again with removing  
different feature



# *Occam's Razor*

- En los problemas de ciencia de datos de la vida real, es casi raro que todas las variables del conjunto de datos sean útiles para construir un modelo.
- Añadir variables redundantes reduce la capacidad de **generalización** del modelo y también puede reducir la precisión global de un clasificador. Además, añadir más variables a un modelo aumenta su complejidad global.
- Según el principio de Parsimonia "**Occam's Razor**", la mejor explicación de un problema es la que implica el menor número posible de supuestos.



# Beneficios

- Modelos más simples.
- Tiempos de entrenamiento más cortos.
- Reducción de varianza.
- Evitar la maldición de la alta dimensionalidad.
  - A medida que aumenta la dimensionalidad y la cantidad de características, el volumen del espacio aumenta tan rápido que los datos disponibles se vuelven limitados.

# Métodos de selección

- Los algoritmos de selección de características se clasifican como **supervisados**, que se pueden usar para datos etiquetados; o **no supervisados**, que se pueden usar para datos **no** etiquetados.
- Las técnicas supervisadas se clasifican como métodos de filtrado (filters), métodos envolventes (wrappers), métodos integrados (embedded), o métodos híbridos.

# Métodos *Filter*

- Seleccionan características en función de **estadísticas**.
- Se aplica una métrica seleccionada para identificar atributos irrelevantes y se realiza una selección recursiva de características.
- Los métodos de filtro son **univariados**, en los que se establece una lista de clasificación ordenada de características para informar la selección final del subconjunto de características; o **multivariante**, que evalúa la relevancia de las características en su conjunto, identificando características redundantes e irrelevantes.

# Ejemplos de métodos *Filter*

- **Correlación de Pearson:** se utiliza como medida para cuantificar la dependencia lineal entre dos variables continuas X e Y.
- **Análisis de varianza ANOVA:** proporciona una prueba estadística de si las medias de varios grupos son iguales o no.
- **Chi-Cuadrado:** prueba estadística aplicada a los grupos de características categóricas para evaluar la probabilidad de correlación o asociación entre ellos utilizando su distribución de frecuencia.

# Ejemplos de métodos *Filter*

- **Ganancia de información:** se calcula mediante la diferencia entre la entropía antes y después de conocer el valor de una característica. La entropía mide la impureza o incertidumbre en un conjunto de datos.  
$$\text{InfoGain}(\text{Class}, \text{Attribute}) = H(\text{Class}) - H(\text{Class} \mid \text{Attribute}).$$
- **Radio de la ganancia de información:** es una relación entre la ganancia de información y la información intrínseca. Intenta reducir el sesgo hacia los atributos multivaluados teniendo en cuenta el número y el tamaño de las ramas al elegir un atributo.

$$\text{GainR}(\text{Class}, \text{Attribute}) = (H(\text{Class}) - H(\text{Class} \mid \text{Attribute})) / H(\text{Attribute}).$$

# Métodos *Wrapper*

- Consideran la selección de un conjunto de características como un **problema de búsqueda**, en el que se evalúa su calidad con la preparación, evaluación y comparación de una combinación de características con otras combinaciones de características.
  - Este método facilita la detección de posibles interacciones entre variables.
- Estos métodos se centran en subconjuntos de características que ayudan a mejorar la calidad de los resultados del algoritmo de agrupamiento utilizado para la selección.

# Ejemplos de métodos *Wrapper*

- **Forward selection:** es un método iterativo en el que se comienza sin tener características en el modelo. En cada iteración, se agrega la característica que mejora el modelo hasta que la adición de una nueva variable no mejora el rendimiento del modelo.
- **Backward elimination:** se comienza con todas las características y se elimina la característica menos significativa en cada iteración, lo que mejora el rendimiento del modelo. Se repite hasta que no se observa ninguna mejora.



# Ejemplos de métodos *Wrapper*

- **Recursive feature elimination:** es un algoritmo de optimización voraz (greedy) que tiene como objetivo encontrar el subconjunto de características con mejor rendimiento. Crea modelos repetidamente y deja de lado la característica de mejor o peor rendimiento en cada iteración. Construye el siguiente modelo con las características restantes hasta agotar todas las características. Luego clasifica las características según el orden de su eliminación.

# Diferencias entre Filters y Wrappers

- Los **filters** miden la relevancia de las características por su correlación con la variable dependiente, mientras que los **wrappers** miden la utilidad de un subconjunto de características entrenando un modelo en él.
- Los **filters** son mucho más rápidos en comparación con los **wrappers**, ya que no implican el entrenamiento de los modelos.
- Por otro lado, los **wrappers** también son computacionalmente muy costosos.

# Diferencias entre Filters y Wrappers

- Los **filters** usan métodos estadísticos para la evaluación del subconjunto de características, mientras que los **wrappers** usan validación cruzada.
- Los **filters** pueden no encontrar el mejor subconjunto de características en muchas ocasiones, pero los **wrappers** sí pueden lograrlo.
- El uso del subconjunto de características de los **wrappers** hace que el modelo sea más propenso a sobreajustarse en comparación con el uso del subconjunto de características de los **filters**.

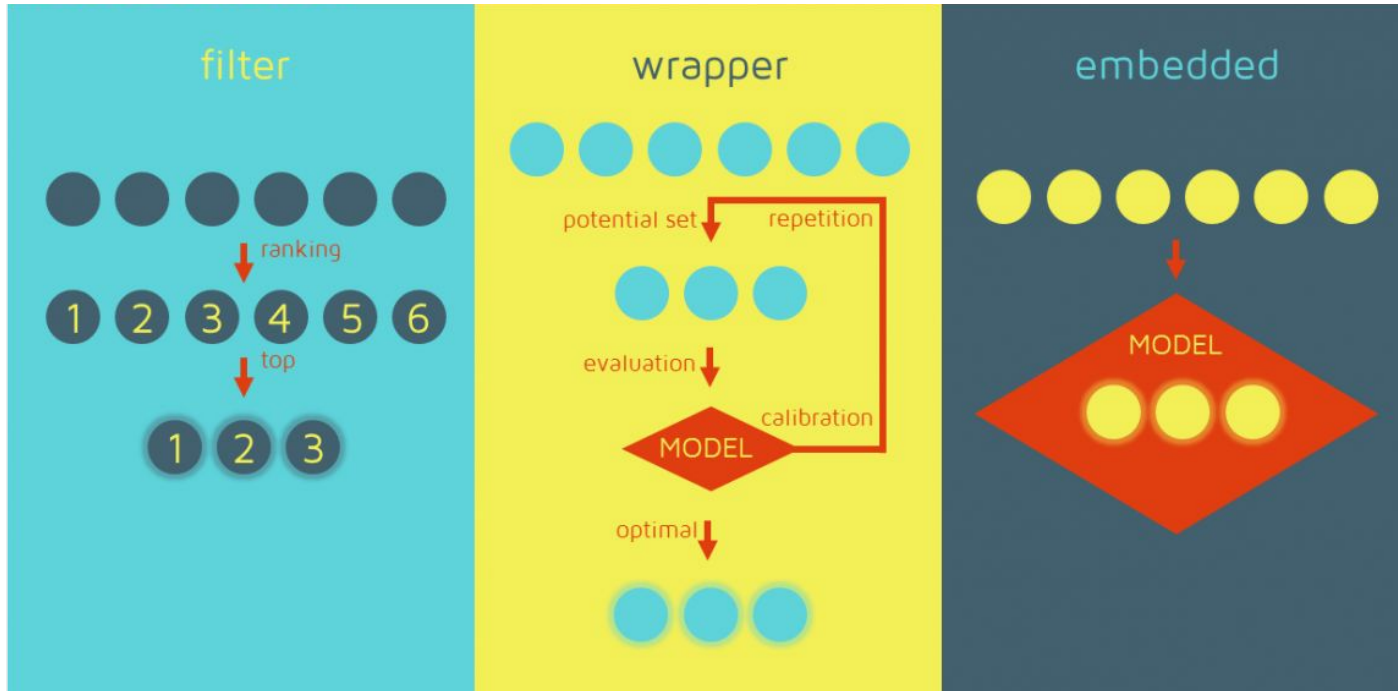
# Métodos *Embedded*

- Integra el algoritmo de selección de características como parte del algoritmo de aprendizaje, en el que la clasificación y la selección de características se realizan simultáneamente.
- Las características que contribuyen más a cada iteración del proceso de entrenamiento del modelo se extraen cuidadosamente.

# Ejemplos de métodos *Embedded*

- **LASSO regression:** realiza la regularización L1 que agrega una penalización equivalente al valor absoluto de la magnitud de los coeficientes.
- **RIDGE regression:** realiza la regularización L2 que agrega una penalización equivalente al cuadrado de la magnitud de los coeficientes.

# En resumen



# ¿Cómo elegir el método de selección?

- **Entrada numérica – salida numérica.** Problema de regresión de selección de características: coeficiente de correlación, como el coeficiente de correlación de Pearson (regresión lineal) o el coeficiente de rango de Spearman (para no lineal).
- **Entrada numérica – salida categórica.** Problema de clasificación de selección de características: coeficiente de correlación, teniendo en cuenta el objetivo categórico, como el coeficiente de correlación ANOVA (para lineal) o el coeficiente de clasificación de Kendall (no lineal).



# ¿Cómo elegir el método de selección?

- **Entrada categórica – salida numérica.** Problema de modelado predictivo de regresión (poco común): coeficiente de correlación, como el coeficiente de correlación ANOVA (para lineal) o el coeficiente de clasificación de Kendall (no lineal), pero a la inversa.
- **Entrada categórica – salida categórica.** Problema de modelado predictivo de clasificación: coeficiente de correlación, como la prueba Chi-Squared (tablas de contingencia) o información mutua, que es un método poderoso que es independiente de los tipos de datos.

# Casos de estudio

# Ejemplo de enfoque automático

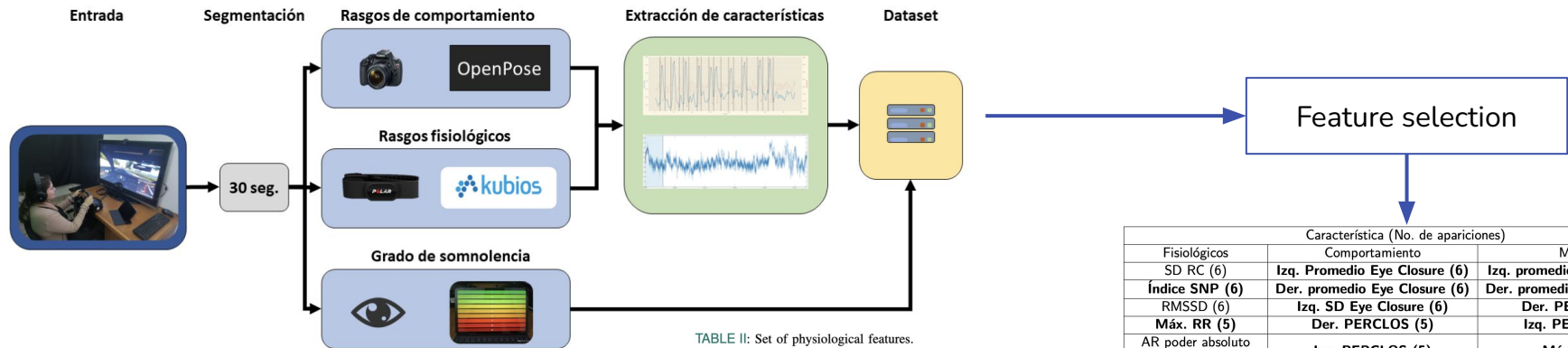


TABLE I: Set of behavioral features.

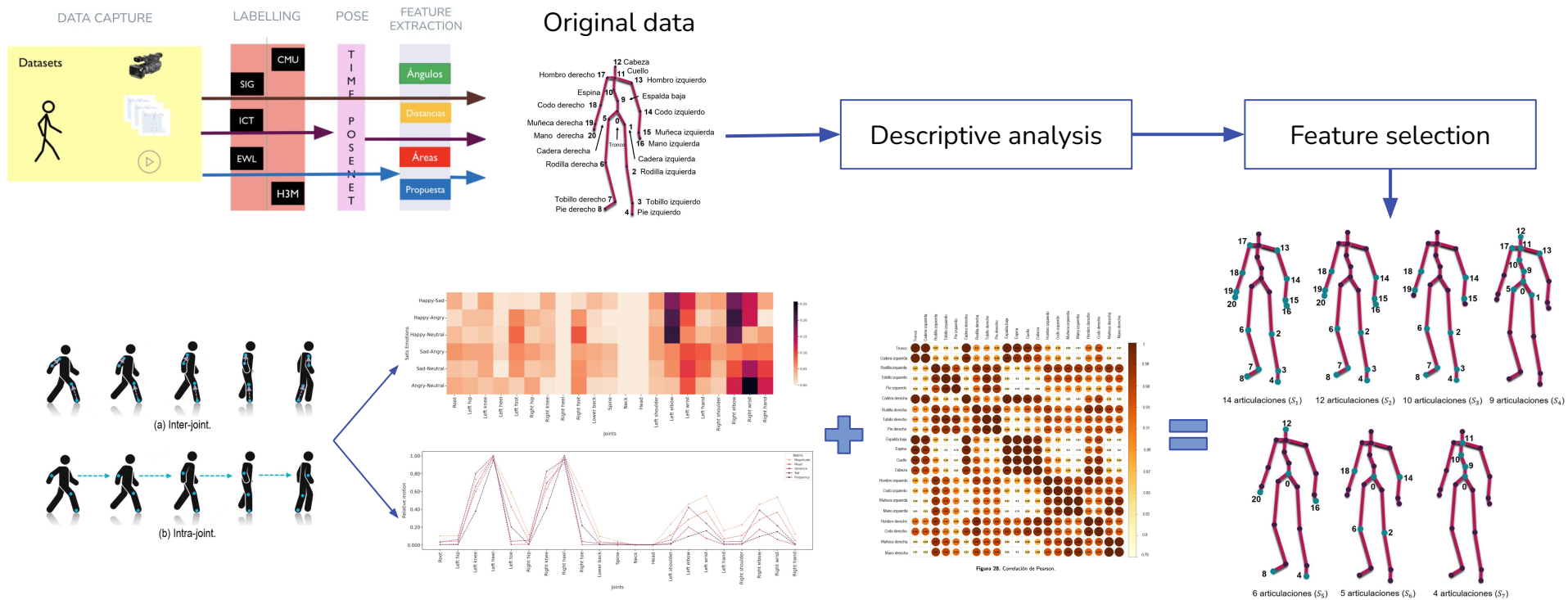
Feature	Description
<b>Time domain:</b>	
Blinks	The number of blinks
PERCLOS	Percentage of eye closure over a time window
Blinking frequency*	Distance between blinks over a time window
Blinking duration*	Time elapsed between start-end of a blink
Time to max eye-closing*	Time elapsed from the start of a blink to its maximum closing point
Time to max eye-opening*	Time elapsed from the maximum closing point to the end of a blink
<b>Amplitude domain:</b>	
Eye Closure*	The amplitude of the closing and opening of the eyes through time
<b>Velocity domain:</b>	
Eye Closure Velocity*	Maximum amplitude value for the eye closure
Eye Opening Velocity*	Minimum amplitude value for the eye-opening

TABLE II: Set of physiological features.

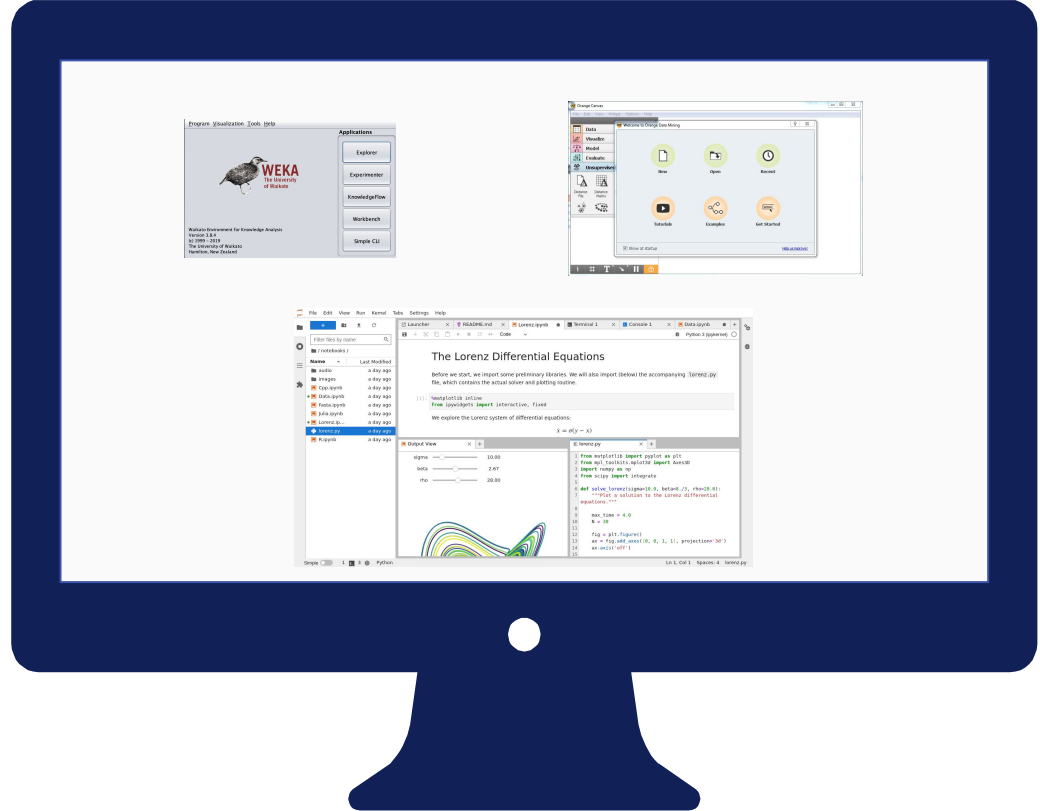
Feature	Description
<b>Time domain:</b>	
RR1*	RR intervals, same as Inter-beat-interval (IBI)
HR*	Heart rate signal
RMSDD	RMS differences between successive RRI
NN50	# of successive RRI pairs differ more than 50 ms
NN50	NN50 divided by the total number of RRI
HRV triang. index	The integral of the RRI histogram divided by the height of the histogram
TINN	Baseline width of the RRI histogram
Stress index	Square root of Baevsky's stress index
<b>Frequency domain:</b>	
Peak frequency	VLF, LF, and HF band peak frequencies
Abs. power $ms^2$	Abs. powers of VLF, LF, and HF bands
Abs. power $log$	Natural logarithm transformed values of absolute powers of VLF, LF, and HF bands
Relative power	Relative powers of VLF, LF, and HF bands
Normalized power LF/HF	Powers of LF and HF bands in normalized units
	Ratio between LF and HF band powers
<b>Nonlinear domain:</b>	
SD1	In the Poincaré plot, the standard deviation perpendicular to the line-of-identity
SD2	In the Poincaré plot, the standard deviation along the line of identity
SD1/SD2	Ratio between SD2 and SD1
<b>Kubios:</b>	
PNS index	Parasympathetic nervous system activity compared to normal resting values
SNS index	Sympathetic nervous system activity compared with normal resting values

Característica (No. de apariciones)		
Fisiológicos	Comportamiento	Múltiple
SD RC (6)	Izq. Promedio Eye Closure (6)	Izq. promedio Eye Closure (6)
Índice SNP (6)	Der. promedio Eye Closure (6)	Der. promedio Eye Closure (6)
RMSSD (6)	Izq. SD Eye Closure (6)	Der. PERCLOS (5)
Máx. RR (5)	Der. PERCLOS (5)	Izq. PERCLOS (5)
AR poder absoluto AF ms2 (5)	Izq. PERCLOS (5)	Máx. RR (4)
Promedio RR (5)	Der. mín. Eye Closure (5)	Der. mín. Eye Closure (4)
Poincaré SD1 (5)	Der. SD Eye Closure (4)	Índice SNP (4)
Mín. RR (5)	Der. parpadeos (4)	Promedio RC (4)
AR poder relativo BF pct (5)	Izq. duración mínima de parpadeo (3)	Promedio RR (3)
AR BF frq (5)	Izq. tiempo promedio al cerrado del ojo (3)	Der. parpadeos (3)
NN50 (5)	Izq. tiempo mínimo a la apertura del ojo (3)	Izq. SD Eye Closure (2)
AR relación BF AF(4)	Izq. velocidad promedio de la apertura del ojo (3)	Izq. mín. Eye Closure (2)
FFT poder absoluto MBF ms2 (4)	Der. tiempo mínimo a la apertura del ojo (2)	Der. SD Eye Closure (2)
Índice SNS (4)	Izq. tiempo promedio a la apertura del ojo (2)	NN50 (2)
Promedio RC (4)	Izq. mín. Eye Closure (2)	AR AF frq (2)

# Ejemplo de enfoque semi-automático



(Go to live)



# Gracias!

¿Alguna pregunta?

hussein@cicese.mx

<https://sites.google.com/view/husseinlopeznava>



**CREDITS:** This presentation was based on a template by [Slidesgo](#), and includes icons by [Flaticon](#).