

# Técnicas Machine Learning para explicar el CEC en la generación de energía eléctrica



## Resumen

En este póster se implementaron estrategias de Machine Learning para brindar un análisis alternativo a la base de datos extraída del artículo; *Evaluación de mezclas biodiésel-diésel en la generación de energía eléctrica* (2011) [3] permitiendo construir modelos y de esta forma explicar el consumo específico de combustible en el funcionamiento de una planta generadora de energía eléctrica. Se crearon modelos utilizando las metodologías: Árboles de regresión, Bosque aleatorios y modelos de regresión lineal, entre los cuales se seleccionó el mejor modelo utilizando el error de validación cruzada, que fue calculado por medio de el software estadístico R.

## Introducción

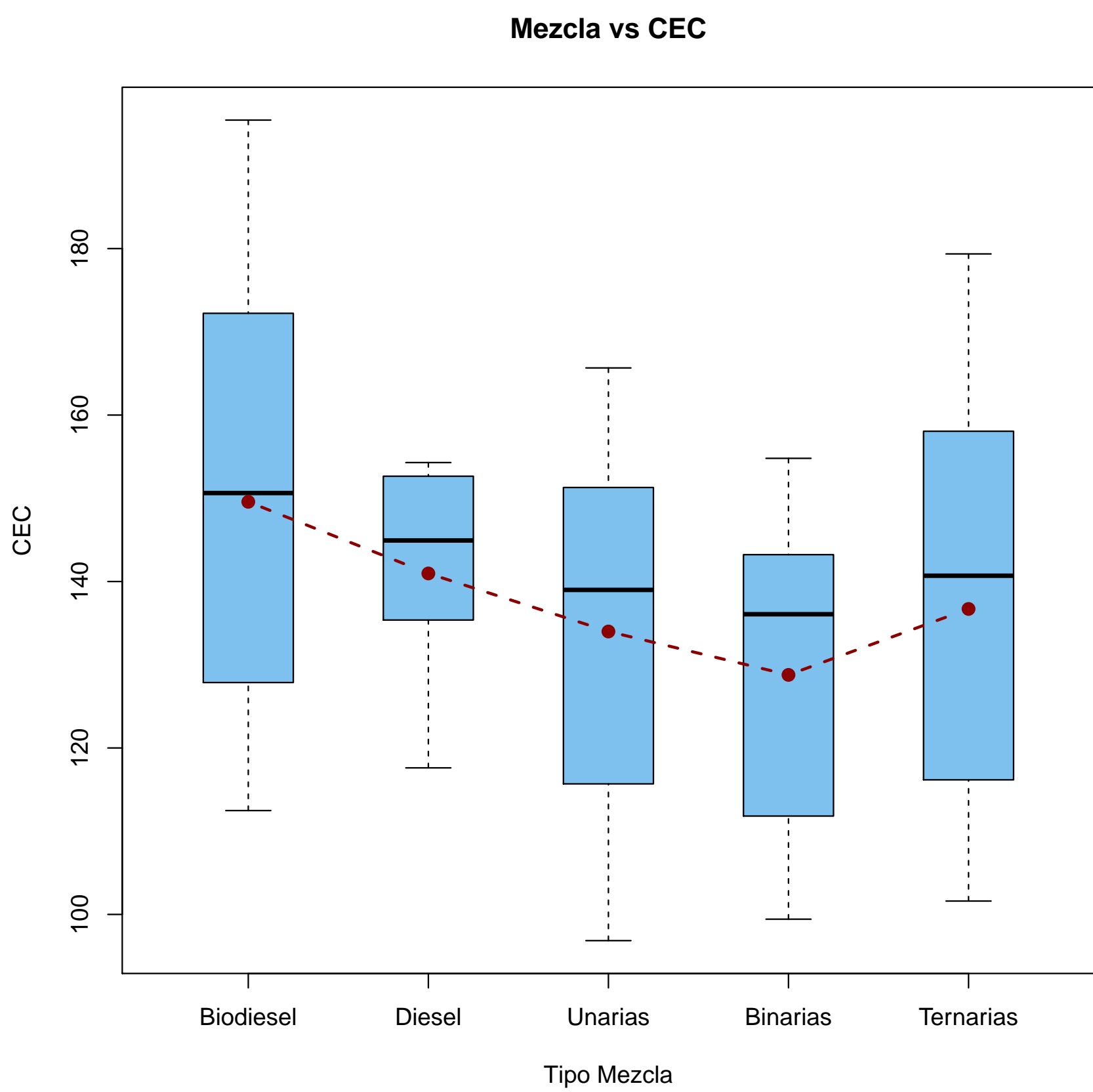
El biodiésel es un combustible líquido producido a partir de los aceites vegetales y grasas animales. El objetivo de este póster es mostrar el efecto de tres diferentes biodiésels mezclados con diésel (combustible que se obtiene a partir de la destilación y la purificación del petróleo crudo) en el funcionamiento de una planta de generación de energía eléctrica, simulando el consumo energético de una vivienda constituida por diez bombillos, una estufa eléctrica de una hornilla, una plancha y un ventilador, con el fin de determinar el consumo específico de combustible.

## Estadística descriptiva

En el Cuadro 1 se presentan las variables utilizadas en la creación de los modelos, cabe resaltar que se revisaron los residuales y los diagnósticos para balanceo e influencia de la base de datos y no se encontró problemas.

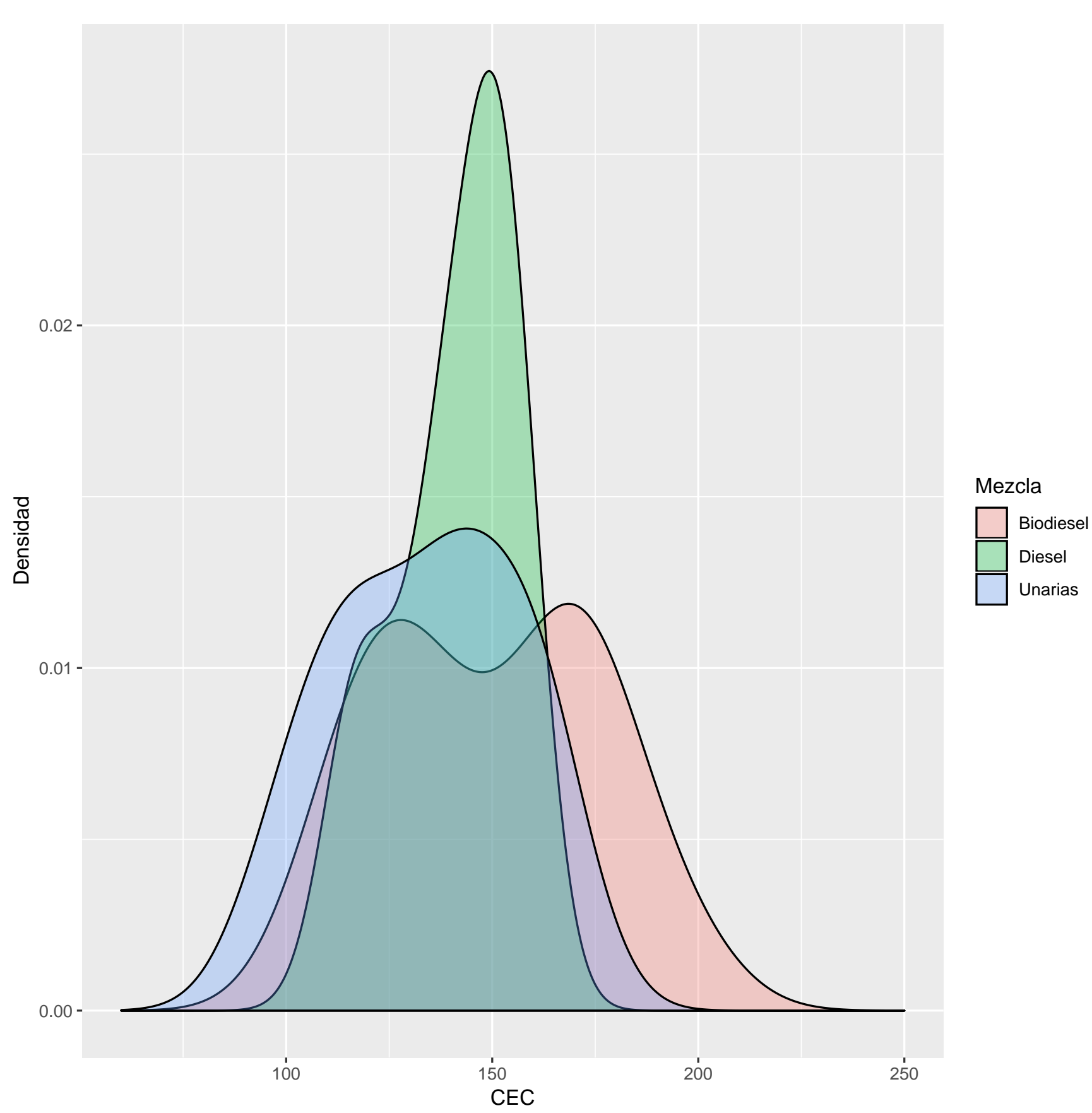
Variable	Tipo de variable
CEC [g/kwh]	Númerica
Velocidad [rpm]	Númerica
Tiempo [H]	Númerica
Carga [5 levels]	Categórica
Mezclas [5 levels]	Categórica

Cuadro 1: Clasificación variables



**Figura 1:** CEC discriminado por mezcla. La Figura 1 permite comparar la variación en el consumo específico de combustible para cada una de las mezclas de combustible.

## Estadística descriptiva



**Figura 2:** Densidad del CEC discriminado por tipo de Mezcla.

La figura 2 nos permite identificar cómo se comporta el Consumo Específico de Combustible de acuerdo al tipo de Mezcla usado para la generación de electricidad, en él se observa el motivo por el cual no se utiliza del todo biocombustible pues muestra la variabilidad que hay en el Consumo cuando se usa.

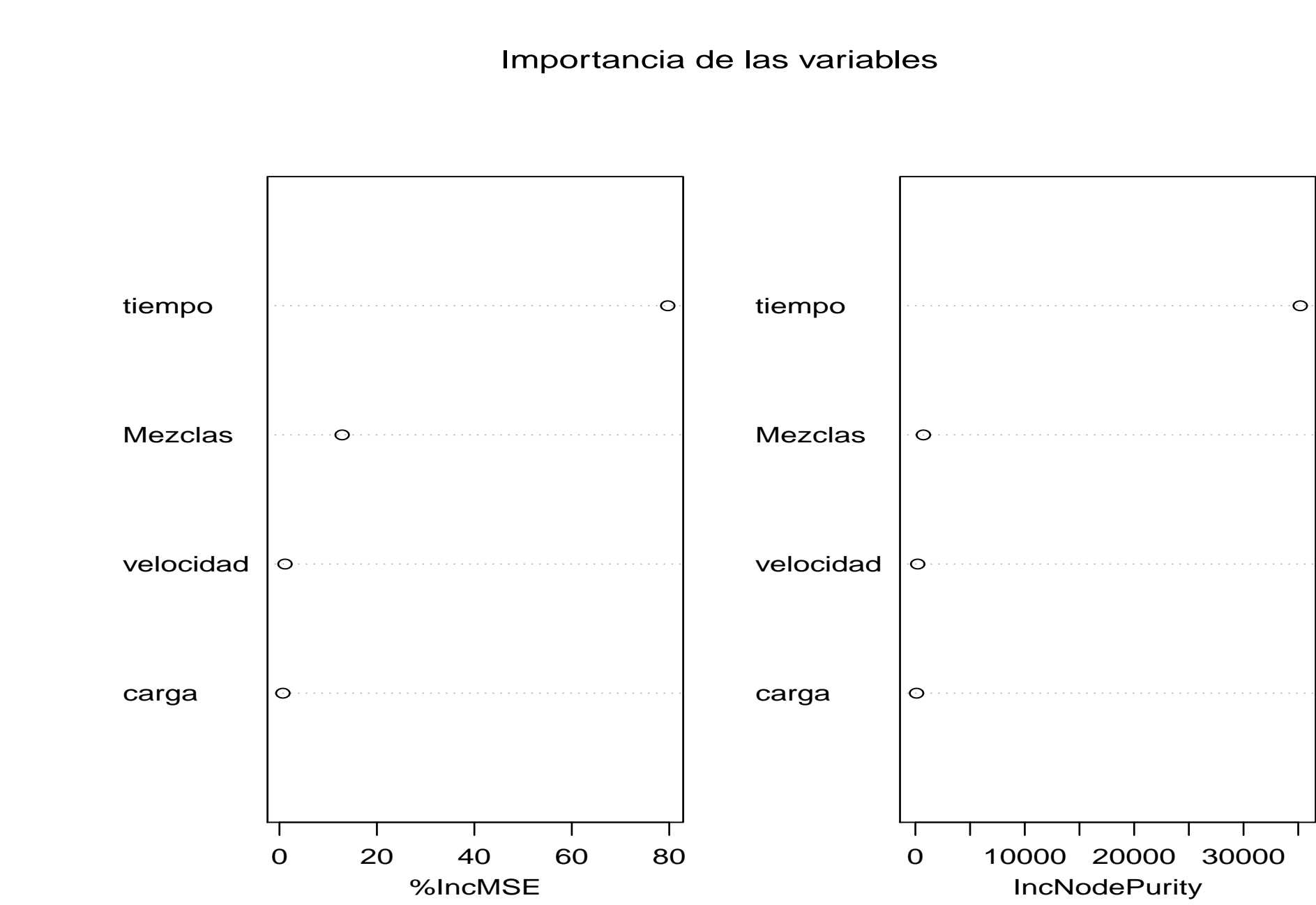
## Metodología

Tanto la estadística descriptiva como la creación de los modelos se hicieron por medio del software R. Se utilizaron los paquetes **tree**, **gamlss** y **randomForest** para la construcción de los modelos y el paquete **DMwR** para la comparación de estos. Se usó como métrica de comparación la validación cruzada **loocv** la cual consiste en realizar N iteraciones donde en cada iteración se deja un solo dato en el conjunto de prueba. Los resultados obtenidos se presentan en la siguiente tabla en la cual la columna **PVC** representa el promedio de los errores de validación cruzada y **DVC** representa la varianza en esos errores.

Metodología	PVC	DVC
Arboles de regresión	62.0499	133.335
Bosques aleatorios	4.965193	10.2352
GAMLSS	6.000875	16.77285

Cuadro 2: Resumen modelos

## Resultados



**Figura 3:** Importancia de variables

Basándonos en la figura 3 generada por el modelo **Bosques aleatorios** podemos notar cómo influye cada variable en la precisión del modelo y en la disminución de su impureza.

## Resultados

- ▶ Con el modelo de **bosque aleatorio** se puede determinar que las variables que mas influyen a la hora de explicar el CEC de cierto combustible son el tiempo y el tipo de mezcla.
- ▶ El modelo de **árboles de regresión** también arrojó que las variables más útiles en la explicación del CEC son el tiempo y el tipo de mezcla.
- ▶ Aplicando el modelo de **regresión con gamlss**, se llegó a que la variable respuesta (CEC) tiene distribución (GA). Además dado que la función de enlace usada en la modelación es log, se puede obtener las expresiones para los parámetros estimados  $\mu$  y  $\sigma$  como se muestra a continuación; además se encontró que cada una de las variables escogidas en el modelo es significativa a un nivel del 5%.

$$\hat{\mu} = \exp(6,35 - 17,57\text{tiempo} + 37,92\text{tiempo}^2 + 0,08M_{\text{biodiesel}} + 0,033M_{\text{unarias}} + 0,031M_{\text{binarias}} + 0,035M_{\text{ternarias}})$$
$$\hat{\sigma} = \exp(-9,23 - 119,3\text{tiempo} - 551,7\text{tiempo}^2 - 0,35M_{\text{biodiesel}} - 2,23M_{\text{unarias}} - 2,78M_{\text{binarias}} - 2,54M_{\text{ternarias}})$$

## Conclusiones

- ▶ Las mezclas unarias creadas con 80 % de diésel y 20 % de biodiésel son las que presentan menor CEC. El consumo más alto de combustible se obtuvo cuando se usó mezclado diésel con aceite de higuera.
- ▶ No se presentaron diferencias significativas en las medias del Consumo Específico de Combustible cuando realizamos este análisis discriminando por el tipo de mezcla.
- ▶ El diésel usado al 100% es el combustible que presenta menor variabilidad en el consumo y el biodiésel es el que mayor variabilidad presenta.
- ▶ El mejor biodiésel es el proveniente del aceite de palma, ya que minimiza el consumo específico de combustible.

## Referencias

[1] R Core Team (2018). *R: A Language and Environment for Statistical Computing*. Vienna, Austria., <http://www.R-project.org/>

[2] Stasinopoulos, M; Rigby, R; Heller, G; Voudouris, V; & De Bastiani, F. *Flexible Regression and Smoothing Using GAMLSS in R*, ACHAPMAN&HALLBOOK

[3] Rojas, A; Chaparro, O; Andrés, C (2011). *Evaluación de mezclas biodiésel-diésel en la generación de energía eléctrica*, <http://www.redalyc.org/articulo.oa?id=47721739002>

## Contacto

- ▶ Marin B. Daniel & Agudelo G. Samuel. & Irál P. Rene
- ▶ [ydmarinb@unal.edu.co](mailto:ydmarinb@unal.edu.co);
- ▶ [saagudeloga@unal.edu.co](mailto:saagudeloga@unal.edu.co);
- ▶ [riral@unal.edu.co](mailto:riral@unal.edu.co)
- ▶ <https://github.com/ydmarinb/Poster/blob/master/CODIGO%20POSTER.R>