



# Análisis de curvas de presión de cabeza en pozos surgentes

Por **Adriana Romero** (Practia), **Ignacio Alvarez Claramunt** y **Jose Luis Barros** (YPF), **Julio César Rodríguez Martino** (Practia) y **Gabriel Horowitz** (Y-TEC)

Este trabajo fue presentado en las *VI Jornadas de Geotecnología*, en el marco del *10° Congreso de Exploración y Desarrollo de Hidrocarburos del IAPG* (Noviembre de 2018, Mendoza).

**U**n pozo está en surgencia natural (pozo surgente) cuando la presión en el fondo es suficiente para impulsar su producción hasta la superficie. Los pozos surgentes hacen uso de algunas restricciones en la superficie para regular la velocidad del flujo expulsado, de manera que el rendimiento general de un pozo es función de varias variables. Por ejemplo, el tamaño de la tubería, el tamaño del estrangulador, la presión en la cabeza, el tamaño de la línea de flujo y la densidad de perforación, entre



**En el presente trabajo se realizó el análisis de las curvas de presión en la cabeza utilizando ciencia de datos, con el objetivo de predecir el comportamiento de la curva de presión e identificar, de manera temprana, las anomalías que podrían presentarse. Y así trabajar en su rápida corrección.**

otros. Esto implica que el cambio de cualquiera de estas variables alterará el rendimiento del pozo.

Una de las técnicas para el análisis de la producción es estudiar el comportamiento de la declinación que presenta la presión en la cabeza del pozo, ya que, en condiciones de flujo crítico, el caudal es una función de la misma. A partir de esta gráfica se puede identificar el comportamiento de cada pozo y determinar la presencia de parafina u otros materiales que afecten su producción o, en el peor de los

casos, causen la suspensión del proceso de extracción.

Este trabajo realizó el análisis de las curvas de presión en la cabeza utilizando ciencia de datos, con la finalidad de hacer la predicción del comportamiento de la curva de presión y la identificación temprana de las anomalías que se pudieran presentar con el fin de aplicar una rápida corrección.

Se tomaron los datos de 130 pozos surgentes correspondientes al NOC Oil Sur. Se inició con un proceso de filtrado de la curva de presión, con dos objetivos específicos: primero eliminar valores atípicos de la serie de tiempo y el segundo suavizar la curva de manera que se pueda realizar predicciones a futuro.

A continuación, se aplicó la metodología ARIMA (*autoregressive integrated moving average*) con la finalidad de predecir valores de la curva. Se basa en los valores pasados de la serie de tiempo correspondiente a la curva de presión para inferir los valores futuros, se aprovechó la característica de tendencia de la curva para aplicar este algoritmo.

Luego, para identificar la anomalía se diseñó un modelo basado en la declinación de la curva. La curva de declinación de presión es un exponencial descendente, por lo cual la primera y la segunda derivada indican la tendencia (ascendente-descendente) y curvatura (cóncava o convexa), respectivamente. Una vez que se tienen estos valores a la mano se procede a clasificarlos según la anomalía: parafinamiento, incrustación y obstrucción.

Finalmente, el modelo se está probando en la actualidad en la sala de control en Loma Campana, y entrega una probabilidad de ocurrencia de alguna de las anomalías nombradas cada hora.

## Metodología

### Origen de datos

Para este estudio se dispone de datos de presión en la cabeza de 130 pozos de NOC Oil Sur. Estos datos censados presentan períodos en los cuales la amplitud de la curva varía de manera abrupta, generando ruido. Esto dificulta su análisis sin antes realizar una limpieza de los datos, quitando estos picos y dejando una curva suavizada para luego trabajar en un modelo predictor.

La figura 1 muestra la curva de presión de cabeza desde el 1 de enero de 2018 hasta el 22 de marzo de 2018. Este pozo, en particular, presenta una elevación en la curva de presión.

Con el fin de eliminar el ruido (valores atípicos) de la serie temporal, se aplica un filtro *Running Median* como se observa en la figura 2. Este filtro se utiliza comúnmente para suavizar las curvas con picos muy altos dentro de la serie de tiempo.

Con la finalidad de tener una serie de tiempo lo más suave posible, sin perder las características de la curva, se aplicó un filtro pasabajos (descarta los valores altos). Con esto las predicciones que se realizarán serán lo más cercanas al comportamiento de la presión. En la figura 3 se muestra el resultado de este filtro.

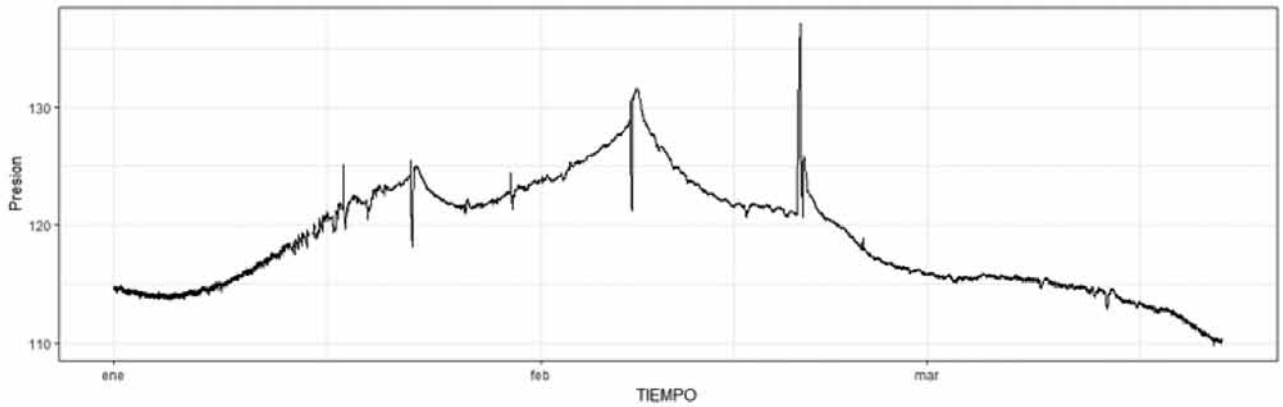


Figura 1. Presión de cabeza de pozo surgente.

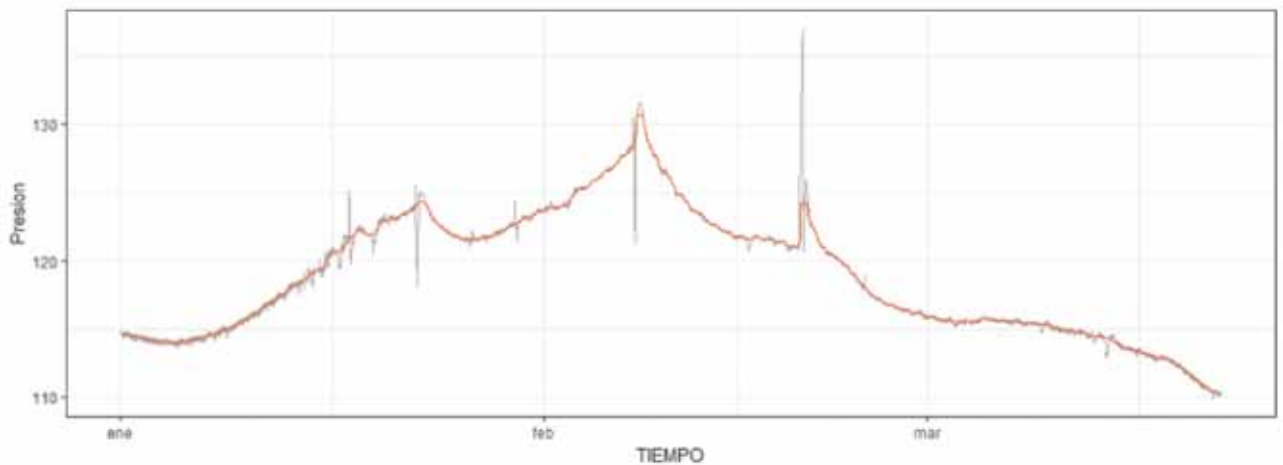


Figura 2. Curva de presión después de aplicar *Running Median*.

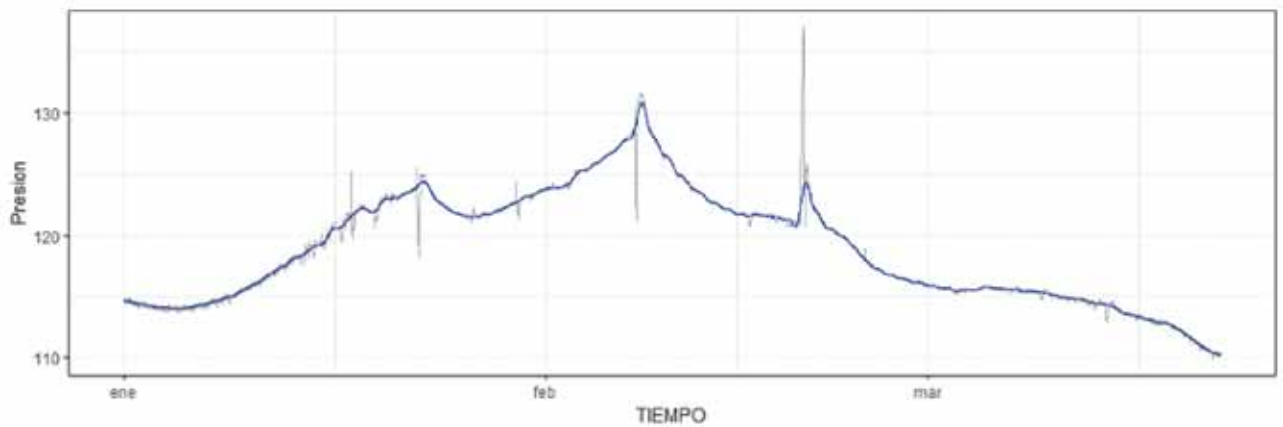


Figura 3. Curva de presión después de aplicar filtro pasabajos.

## Modelos ARIMA

Al aplicar la metodología ARIMA (*autoregressive integrated moving average*) se construye un modelo predictor de la serie de tiempo correspondiente a la presión de cabeza. A partir de este modelo se realizará la predicción de un día y medio hacia adelante. El modelo se puede ajustar nuevamente con cada nueva medición y tener una gráfica actualizada.

En la figura 4 se muestra la curva original y la curva estimada utilizando el modelo obtenido utilizando la metodología ARIMA.

## Diagnóstico de anomalías utilizando análisis de tendencias

Las perturbaciones que se presentan a causa de parafina, erosión o interferencia en la serie suavizada se observan en las gráficas de las derivadas de las curvas. En la figura 5 se muestra la clasificación según la tendencia y curvatura de la curva de presión, esto se realiza para asignar una clase a cada una de las anomalías.

Para el cálculo de las derivadas se procedió de la siguiente manera:

- Se toman los valores de presión. Estas lecturas se realiza cada 10 min.



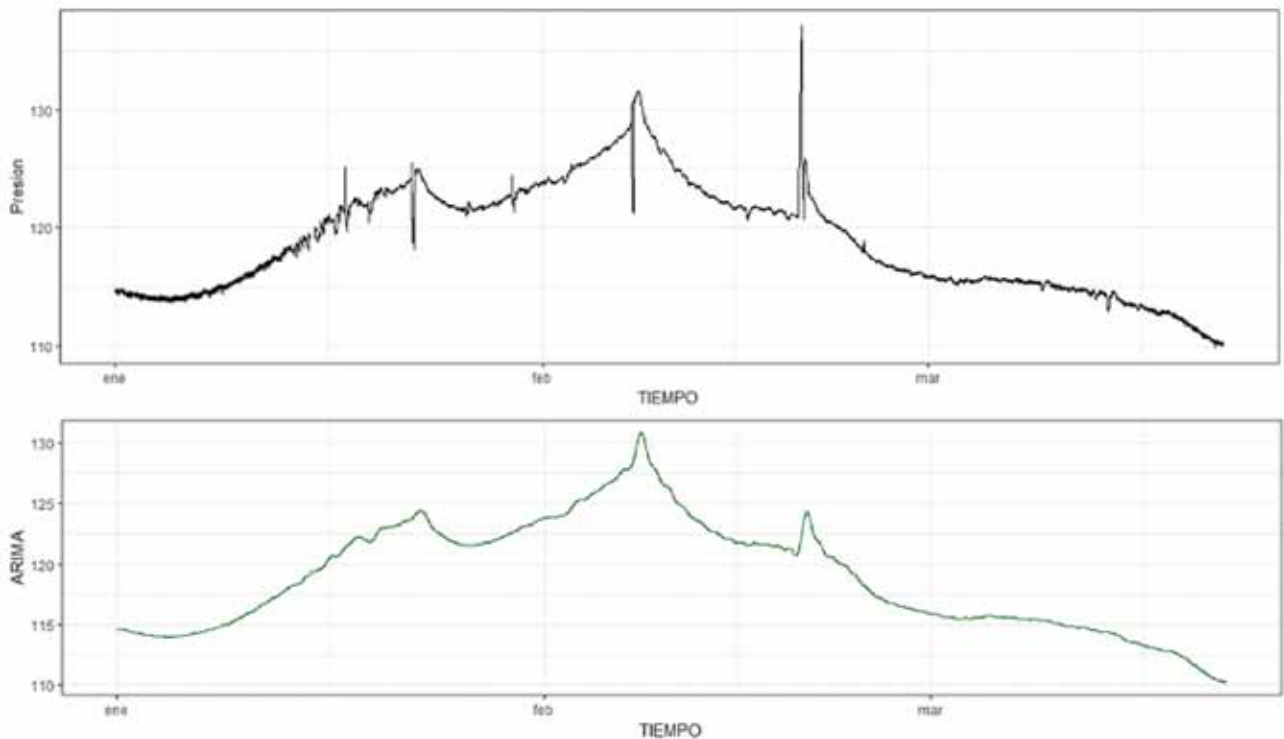


Figura 4. Arriba: curva original de presión de cabeza. Abajo: curva estimada mediante ARIMA.

- Se toman ventanas de 12 h (72 mediciones de 10 min) y se calcula una función que describa la curva de presión en esas 12 h. Se calculan la primera y segunda derivada de esa función.
- Dependiendo de los signos de la primera y segunda derivada se asigna una letra, según la figura 5.
  - Letra A: Comportamiento normal
  - Letra B: Parafinado
  - Letra C: Interferencia
  - Letra D: Incrustación

Una vez asignada la letra (clase) a cada ventana de 12 h, tomando en cuenta que no solamente basta con asignar

una letra por los valores de la derivada, sino también se debe tomar en cuenta la duración del evento, se asigna la clase definitiva. Para este estudio, si la tendencia de la curva se mantiene por dos días (cuatro letras iguales consecutivas) se establece que efectivamente existe un problema y se asigna una clase como se muestra en la figura 6.

Con la finalidad de etiquetar el tipo de anomalía que se presentará, se realiza un proceso de aprendizaje supervisado, es decir, que a partir de un conjunto de datos con una clase asignada, el modelo “aprende” a distinguir los diferentes casos de anomalías presentadas.

En la tabla 1 se detalla la distribución inicial de las clases dentro del conjunto de datos.

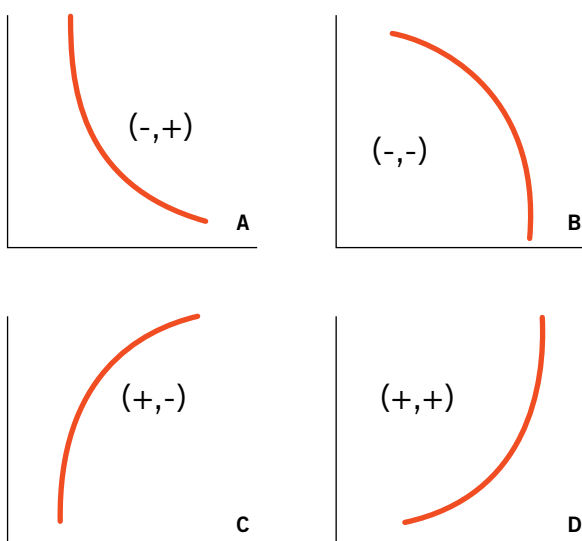


Figura 5. Clases para los casos, según el valor de sus derivadas.

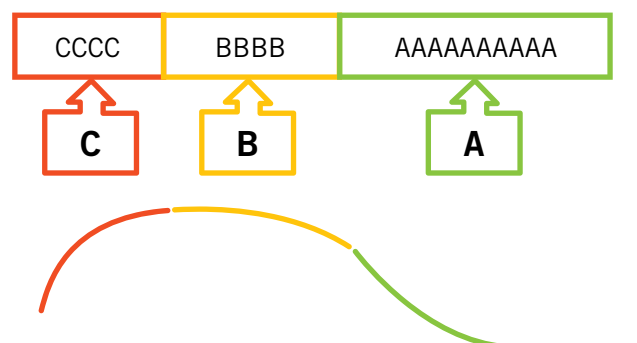


Figura 6. Asignación de clases a la curva basado en la duración de la tendencia.



Clase A	3945
Clase B	2907
Clase C	308
Clase D	429

Tabla 1. Distribución de clases.

Allí se observa que el conjunto de datos se encuentra desbalanceado. Esto significa que no se posee un número similar de observaciones para cada clase.

Los clasificadores (modelos de clasificación) no resuelven *per se* el problema de desbalanceo de clases, sino que son muy sensibles a las proporciones de las diferentes clases. Como consecuencia, estos modelos tienden a favorecer la clase con la mayor proporción de observaciones.

Para realizar el balanceo de las clases en el conjunto de datos se utiliza la técnica SMOTE (técnica de sobremuestreo de minorías sintéticas). Si bien esta técnica está pensada para el sobremuestreo de la clase minoritaria, en este caso al tener más de dos clases, se submuestra la clase mayoritaria a fin de no crear demasiadas muestras sintéticas, sino contar con la mayor parte de mediciones reales. Esto logra un mejor rendimiento del clasificador.

Este proceso se lleva a cabo en cada una de las validaciones (iteraciones) que realiza el algoritmo de entrenamiento del modelo, de manera que no se sobreentrena el modelo (*overfitting*). De esta forma se espera que las estimaciones sean más precisas, tomando en consideración que las lecturas de presión que lleguen al modelo serán desbalanceadas (no siempre se presentan casos de parafinado o interferencia de pozo).

El modelo desarrollado consiste en un ensamble de modelos: *Distributed Random Forest* (DRF), *Deep Learning* y *Gradient Boost Machine* (GBM). (Nota de la Redacción: DRF es una herramienta de clasificación y regresión que con los datos genera un “bosque de árboles” en lugar de un árbol único; más árboles reducirán la varianza al tomar la predicción promedio para una predicción final.

GBM obtiene resultados predictivos a través de aproximaciones cada vez más refinadas al construir en paralelo cada árbol de regresión.

Un ensamble de modelos es una técnica de combinación de dos o más modelos llamados bases. Cada uno de los modelos base tienen una “comprensión” diferente al momento de clasificar. Cada uno de ellos realiza estimaciones variadas en función de su propio enfoque del problema.

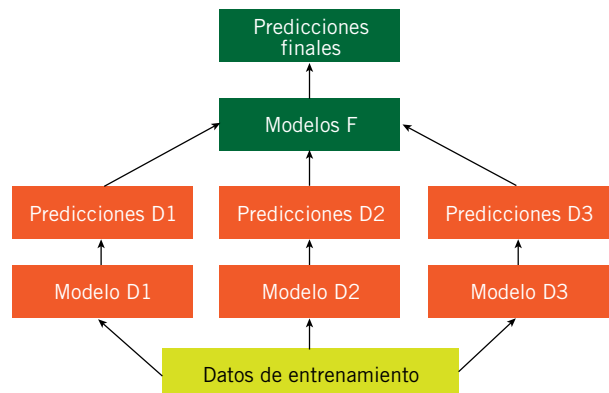


Figura 7. Esquema de *stacking* de modelos.

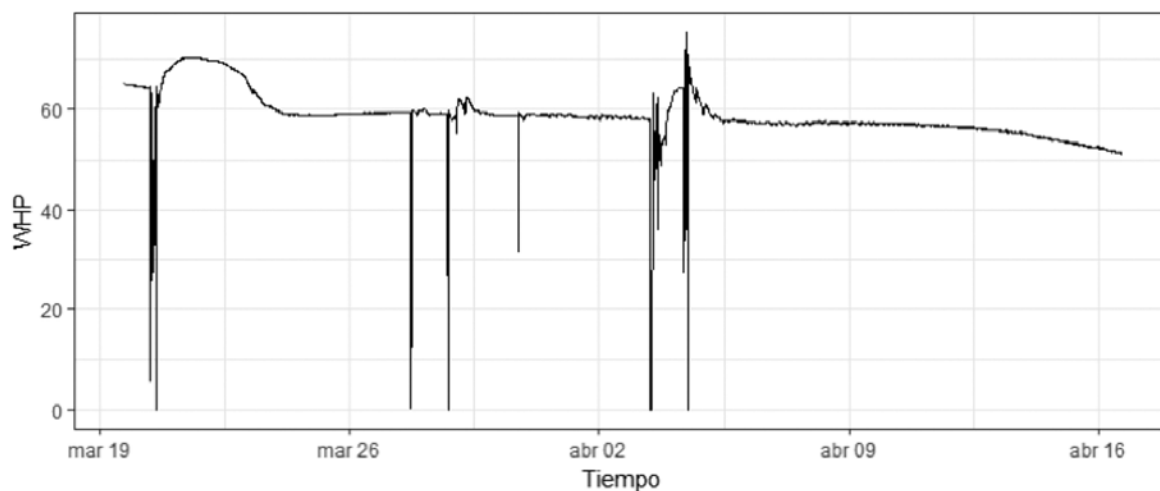


Figura 8. Curva de presión de cabeza. Pruebas de modelo.

Luego un metamodelo final considera las predicciones de los modelos base y define la clasificación. Esto hace que la decisión final sea más robusta, precisa y menos propensa a ser parcial.

La técnica de ensamble se denomina *stacking* (apilamiento). Consiste en apilar varias capas de modelos, una sobre otra, donde cada uno de los modelos pasa sus predicciones al modelo de la capa superior y este toma las decisiones basadas en las salidas de los modelos de las capas debajo de él.

El modelo de la capa superior se selecciona realizando varias pruebas moviendo el orden de los modelos de la pila y midiendo la precisión de las estimaciones en las diferentes combinaciones generadas. Al final se establece el orden de la pila en la cual la precisión sea mayor (Figura 7).

### Pruebas realizadas

Actualmente el modelo se encuentra funcionando en modo de prueba en la sala de control de Loma Campana, realiza predicciones de 130 pozos surgentes y entrega una respuesta cada hora.

En la figura 8 se muestra la curva de presión del pozo SOIL-0314, en un período de un mes. Se puede destacar que al final se observa la declinación de la curva de manera cóncava decreciente.

En la figura 9 se observan las probabilidades de ocurrencia de cada clase. Se resaltan los valores reales medidos y las predicciones realizadas en ese lapso.

En este caso se puede observar que las predicciones a futuro benefician la clasificación de la clase "A", pero esta clasificación se puede corregir cada vez que ingrese un valor real de presión.

### Conclusiones

El trabajo presentado muestra un modelo que estima fallas en pozos surgentes utilizando análisis de tendencias.

Para este trabajo se utilizaron los datos históricos de la curva de presión en la cabeza del pozo con datos correspondientes a 130 pozos surgentes en Loma Campana.

	Tiempo	predict	A	B	C	D
1	2018-04-15 11:10:01	A	0.8898727	0.09421792	0.009776676	0.006132668
2	2018-04-15 12:10:01	B	0.2508556	0.73049948	0.009373589	0.009271323
3	2018-04-15 13:10:01	B	0.2078054	0.77427505	0.008747060	0.009172515
4	2018-04-15 14:10:01	B	0.2179081	0.76766365	0.007925867	0.006502411
5	2018-04-15 15:10:01	B	0.2267282	0.76059322	0.007306197	0.005372366
6	2018-04-15 16:10:01	B	0.2442059	0.74246981	0.007638369	0.005685944
7	2018-04-15 17:10:01	A	0.8078597	0.17450926	0.009762348	0.007868647
8	2018-04-15 23:10:01	B	0.1945181	0.79379346	0.006923103	0.004855321
9	2018-04-16 00:10:01	B	0.2897731	0.69714502	0.007690348	0.005391544
10	2018-04-16 01:10:01	A	0.9033554	0.08648433	0.006573103	0.003587122
11	2018-04-16 02:10:01	A	0.9030339	0.08678926	0.006585913	0.003590909
12	2018-04-16 03:10:01	B	0.2589724	0.72659503	0.007833337	0.006599268
13	2018-04-16 04:10:01	B	0.3837139	0.59995964	0.009162861	0.007163555
14	2018-04-16 05:10:01	B	0.3306397	0.65427511	0.008544532	0.006540700
15	2018-04-16 11:10:01	A	0.9070610	0.08162657	0.007425027	0.003887433
16	2018-04-16 12:10:01	A	0.9074903	0.08243996	0.006516329	0.003554399
17	2018-04-16 13:10:01	A	0.7579880	0.22795350	0.009644658	0.004415872
18	2018-04-16 14:10:01	A	0.8864479	0.10224388	0.007616401	0.003691791
19	2018-04-16 15:10:01	A	0.9073245	0.08263272	0.006489125	0.003553673
20	2018-04-16 16:10:01	A	0.9077006	0.08226774	0.006478092	0.003553553

Figura 9. Tabla de probabilidades de ocurrencia de cada clase.

Se realizó en primera instancia una limpieza de la curva utilizando filtros pasabajos y un filtro de mediana. Para la predicción de los valores de presión a futuro se utilizó el algoritmo de ARIMA y para el clasificador se empleó un ensamble de modelos.

En las pruebas realizadas se tiene una clasificación correcta del 78% en cuanto a la clase designada al segmento de curva en ventanas de 12 h. Se hace énfasis en indicar que para que se diagnostique la falla se debe mantener una tendencia. Este valor lo fijan los expertos en el tema en Loma Campana.

Adicionalmente, es de destacar que el modelo a pesar de que utiliza balanceo de clases y uno de sus algoritmos utiliza una penalización de clasificación por pesos hacia la clase mayoritaria, podría presentar sobre-muestreo, por lo cual su efectividad mejorará al entrenarlo con más casos de fallas. ■