

UNAM

FACULTAD DE INGENIERÍA

“Examen Extraordinario”

Semestre : 2021-1

Profesor: Ing. Berenice Anell Martínez Cabañas

Luis Enrique Navarro Morales

GRUPO

#1

MATERIA

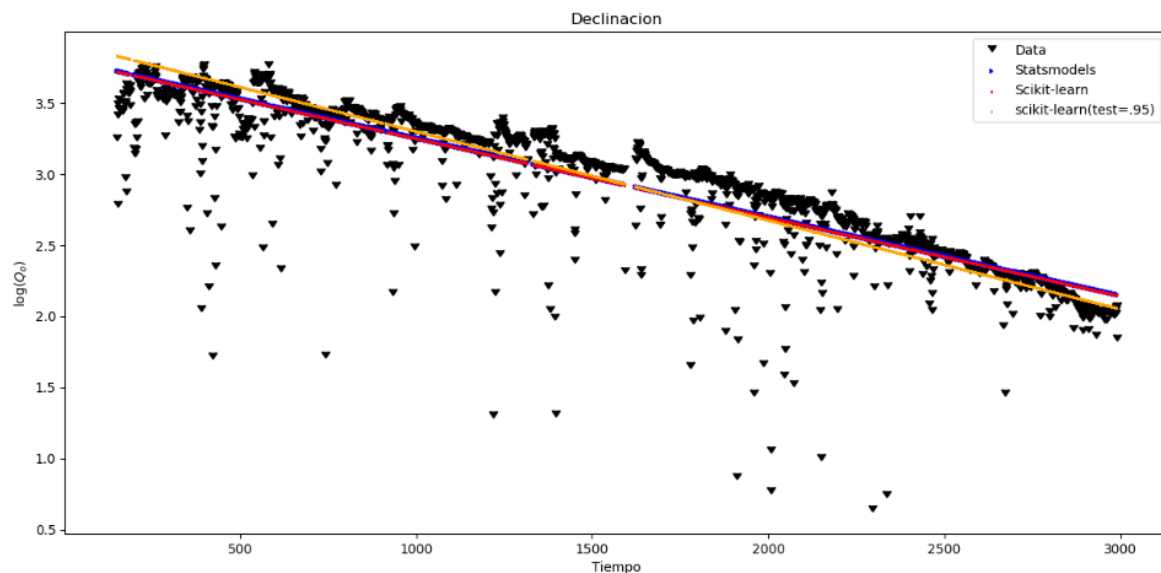
Temas Selectos
de Ingeniería
Petrolera

Resumen

Un factor esencial de análisis para el triunfo de cualquier proyecto es el económico, este análisis deberá apegarse a simular el comportamiento de producción de un campo, o un yacimiento en especial, este proceso se ejemplifica en este trabajo empleando metodologías de Machine Learning y ciencia de datos, así mismo se recalca la importancia del ingeniero petrolero por poseer estas habilidades para una satisfactoria práctica profesional si como, de una toma de decisiones más apegadas a las evidencias ofrecidas por los datos.

Se analizaron datos de producción de un pozo del mar del norte, para la obtención de la tasa de declinación, tiempo de abandono y recobro de reservas a determinado gasto de abandono, mostrando la simbiosis entre la ingeniería petrolera y la ciencia de datos.

Se obtuvo la regresión lineal para determinar el tiempo de abandono, de dicho pozo, mediante la regresión lineal ofrecida por la librería scikit learn, statsmodels, así como una variación del modelo de regresión de scikit learn, empleando el método de entrenamiento de datos, así mismo se muestra la importancia de un correcto modelado matemático y posterior programación, y mostrar el porque, de los clusters de procesamiento, para realizar complejos algoritmos en una base de datos extensa.



Se muestran los resultados de las aproximaciones a través de los métodos antes mencionados.

Contenido

Lista de Figuras

Lista de Tablas

Capitulo 1: Introduccion

Objetivo General

Objetivo Especifico

Capitulo 2 : Fundamentos de la IA y Ciencia de Datos

Inteligencia Artificial

Ciencia de Datos

Python

Capitulo 3 : Estado del Arte

Capitulo 4 : Aprendizaje Supervisado

Clasificacion

Regresion Lineal

Gradient Descent

Capitulo 5. Declinacion

Capitulo 6 : Estudio de caso

Campo Volve

Pre-procesamiento de datos

Modelado

Graficado

Capitulo 7 : Resultados

Capitulo 8 : Conclusiones

Referencias

Anexo A

Lista de Figuras

Figura1. Inteligencia Artificial

Figura 2. Aprendizaje automatico

Figura 3. Razonamiento

Figura 4. Prediccion ROP

Figura 5. Precio Crudo

Figura 6. Proceso

Figura 7. Regresion lineal

Figura 8. Gradient Descent

Figura 9. Learning rate

Figura 10. Curvas de declinación

Figura 11. Campo Volve

Figura 12. Columnas

Figura 13. Produccion Diaria por pozo

Figura 14. Produccion pozo

Figura 15. Graficos de declinación

Figura 16. Boxplot

Lista de Tablas

Tabla 1. DataFrame inicial

Tabla 2. Declinacion

Tabla 3. Modelos

Tabla 4. Resultados

Capitulo 1: Introduccion

*La posesión y el beneficio matan, mientras que la búsqueda,
desligada de cualquier utilitarismo pueden hacer a la humanidad
más libre, más tolerante y más humana.*
Nuccio Ordine

Objetivo general

Mostrar un ejemplo de aplicación en el ámbito de ingeniería petrolera utilizando algoritmos de inteligencia artificial y de metodologías de manejo de datos proporcionadas por el “Data Science”, esto como una forma de mostrar los aportes de estas disciplinas para la realización de estudios y practicas mas eficientes.

Objetivo especifico

Aplicar la regresión lineal mas precisa para la determinación de la declinación, y tiempo de abandono de un campo.

Capitulo 2 : Fundamentos de la IA y Ciencia de Datos

Inteligencia Artificial

La inteligencia artificial es una disciplina de la ciencia de la computación, que busca mediante el empleo de determinados algoritmos replicar los procesos que se llevan a cabo en el cerebro humano, mientras que el “Machine Learning” o aprendizaje automatico, es una rama de dicha disciplina que busca otorgarle inteligencia a una maquina sin estar explícitamente programado.

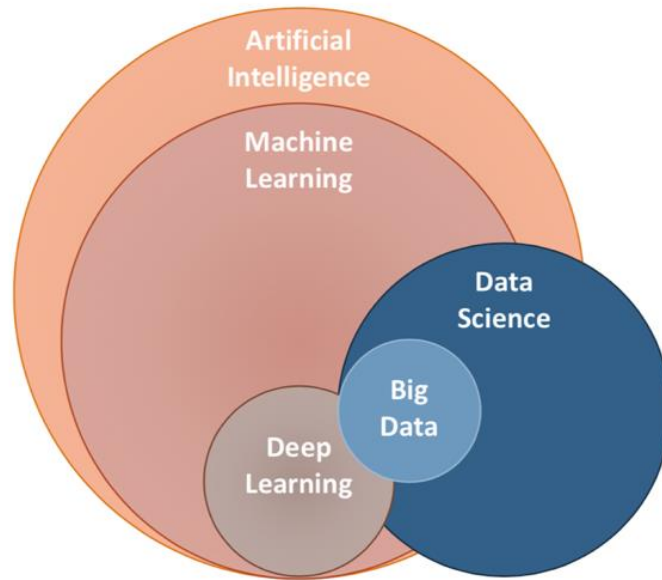


Figura1. Inteligencia Artificial

“Se dice que un programa aprende de una experiencia E con respecto a una tarea T, midiendo su desempeño con P, cuando el desempeño en la tarea T, medido por P mejora con la experiencia E.”

Tom Mitchell

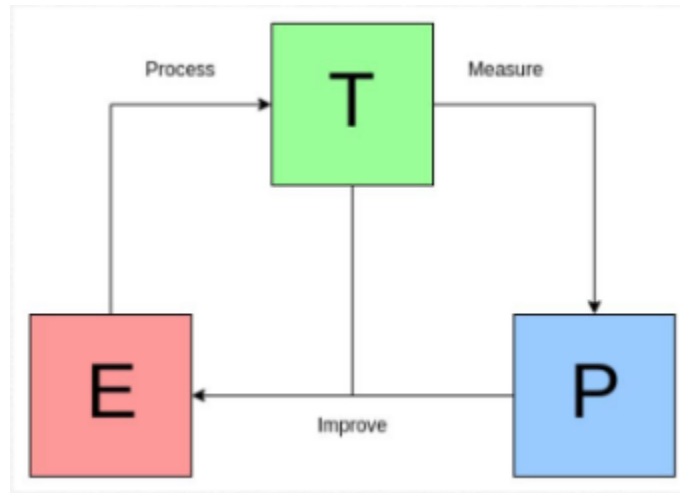


Figura 2. Aprendizaje automatico

Como se aprecia en la figura anterior, el desempeño es el que deberá de mejorar con el tiempo, este aprendizaje se podrá obtener con algoritmos y métodos estadísticos.

El aprendizaje automatico podrá entonces clasificarse en:

- Aprendizaje supervisado: algoritmos que tienen la capacidad de aprender mediante datos y etiquetas
- Aprendizaje no supervisado: Algoritmos que aprenden sin necesidad de etiquetas para dichos datos.

Mas adelante se explicaran con mayor profundidad.

Ciencia de Datos

El "Data Science" o ciencia de datos, es la disciplina de extraer, limpiar, transformar, visualizar, minar los datos obtenidos a través de determinados sistemas, en el ámbito petrolero estos sistemas están conformados y son variados dependiendo de la etapa de la cadena de valor que se este tratando o se este estudiando; el data Handling es una técnica utilizada para preprocesamiento de datos, limpiar y ordenar una gran cantidad de datos para su mejor acceso y visualización, mas adelante se mostrara la importancia.

Ambas disciplinas se apoyan para la realización de varias tareas, tales como el "PDA" o Petroleum Data Analytics, cuya principal tarea es el desarrollar modelos a partir de los datos obtenidos en campo, discrepando con los modelos deductivos que usualmente se aplican para la toma de decisiones.

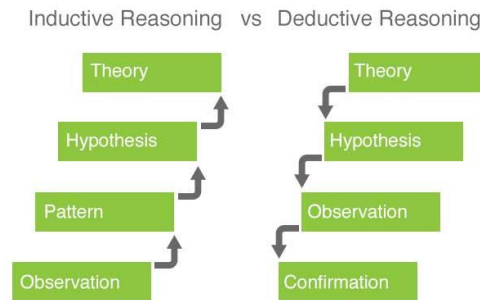


Figura 3. Razonamiento

Python

Python es un lenguaje informático denominado de alto nivel, esto es aquellos lenguajes que expresan sus algoritmos y ordenes teniendo en cuenta las capacidades cognitivas de los seres humanos, esto aligerando las estructuras de programación para que el programador realice mas complejas instrucciones para luego se ejecutadas mediante compiladores, asi mismo, las librerías son un aditamento de importancia para el funcionamiento de ciertos algoritmos desarrollados para este lenguaje.

Anaconda es una paquetería “IDE” enfocada aportar un interfaz de usuario para la programación en Python, esta posee herramientas tales como

Spyder, Software enfocado en la elaboración de scripts, con un interfaz semejante a una terminal de comandos.

Jupyter, software enfocado en el data science, en cual los bloques de código lo caracterizan, ejecuciones de dichos bloques facilitan la visualización y manejo de gran cantidad de datos.

Capitulo 3 : Estado del Arte

En los últimos años ha habido un aumento en el numero de publicaciones referentes al empleo de machine learning, una rama de la inteligencia artificial para la solución de problemas de la ingeniería petrolera, esto gracias al desarrollo de algoritmos aplicables y disponibles de forma gratuita con el lenguaje de programación Python.

Para el año de 1959, el artículo titulado “Preventing Technical Obsolence in Petroleum Engineers and Scientist” dilucida, “Los recién egresados no poseen una ventaja (tecnológica), ya que la tecnología enfocada en el petrolero cambia rotundamente y rápidamente que ellos junto con los veteranos pasaran a no poseer conocimientos fundamentales, a menos que tomen precauciones” (Guerrero Y Martinez,1959), ya para entonces se hacia una necesidad el actualizar nuestros repertorios de conocimientos y capacidades como próximos ingenieros petroleros.

En 1988, al publicarse el libro titulado “Nonlinear Neural Network” de Grossber, Zhang & Song aplicaban dichas metodologías del Deep Learning, para determinar propiedades mecánicas de las rocas, “probablemente las redes neuronales es el puente que unira la geología descriptiva con las expresiones numéricas” (Zhang & Song, 1988), avanzando mas rápido en el tiempo, el empleo de dichos algoritmos ha permeado en la selección de el método adecuado de Sistemas Artificiales de Produccion (Ounsakul,et.al 2019) aplicando arboles de decisión; asi mismo, para la predicción de ROP (Noshi et.al, 2019) mediante la aplicación de Random Forest , algoritmo supervisado obteniendo resultados de importancia económica.

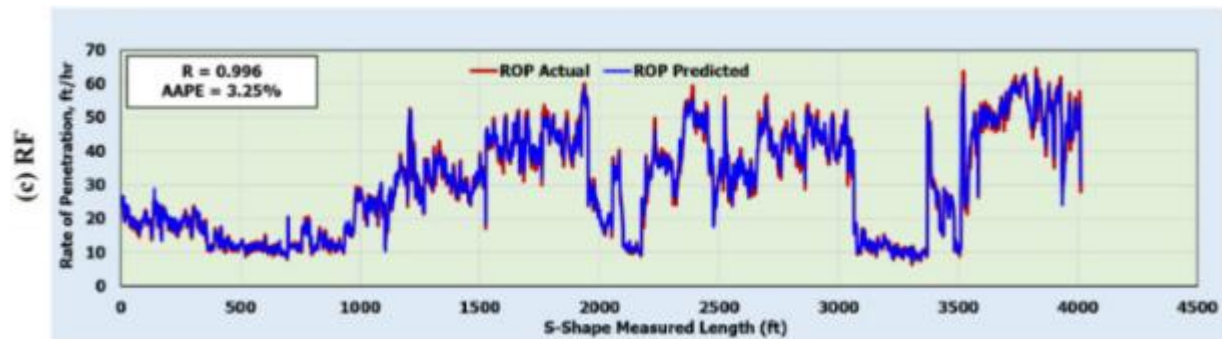


Figura 4. Predicción ROP

También en el ámbito económico se ha utilizado Deep Learning (Bristone,2019), obteniendo resultados apegados a un comportamiento actual.

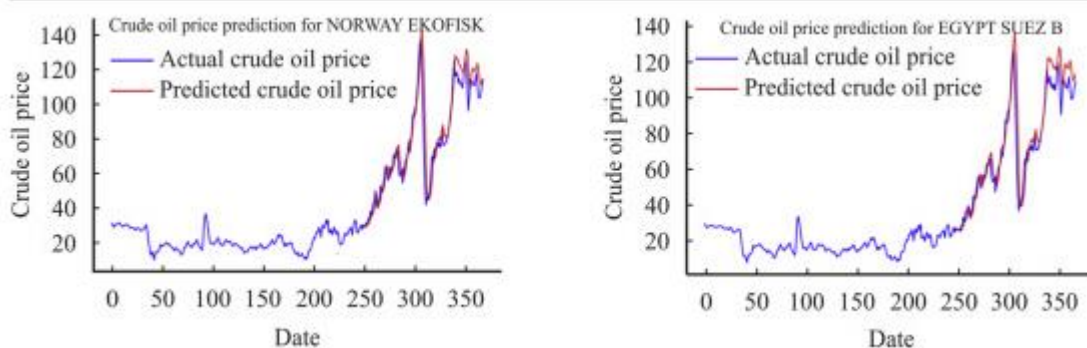


Figura 5. Precio Crudo

En el ámbito de la salud, se han empleado redes neuronales capaces de predecir Gas Kicks, con antelación para su correcto control y pudiendo evitar accidentes relacionados con el proceso de perforación (Yang, 2019).

Capitulo 4 : Aprendizaje supervisado

Como se menciona anteriormente, se le otorga al algoritmo las etiquetas de los datos, en otras palabras, las respuestas correctas, esta es la tarea esencial de estos, el aportar nuevas respuestas.

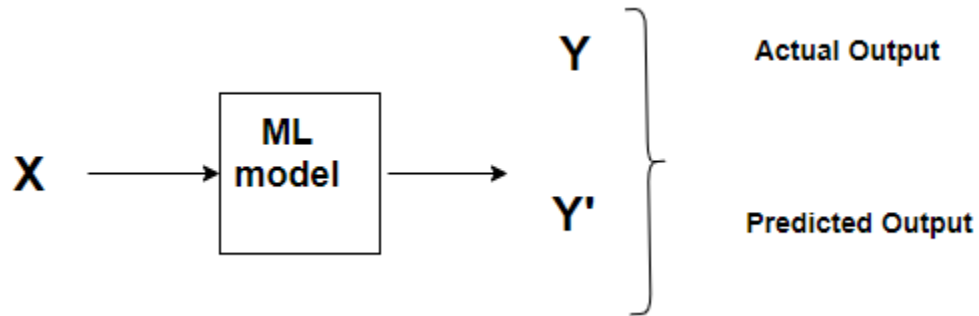


Figura 6. Proceso

De este tipo de aprendizaje se reconocen dos tipos principales, algoritmos de regresión, cuyos variables continuas son utilizadas para obtener un valor real; y algoritmos de clasificación enfocados en problemas de variables categóricas.

El desarrollo de los problemas de regresión son resueltos mediante los algoritmos de:

- Regresión lineal
- Regresión polinomial
- Regresión Ridge/Lasso

Mientras que los de clasificación:

- K-NN
- SVM
- Logistic Regression
- Arboles de Decision

Regresión Lineal

Algoritmo fundamentado en la aproximación de una variable dependiente "**X**" a partir de un número de variables independientes "**Y**", pudiendo ser una o varias para una regresión polinómica.

Esto lo hace mediante la formulación de la ecuación de una recta de la forma:

$$y = ax + b$$

Siendo la “a” la pendiente y la “b” la ordenada al origen.

Esta debe de aproximar los puntos minimizando los errores presentes dados como la distancia residual entre la recta y el valor real.

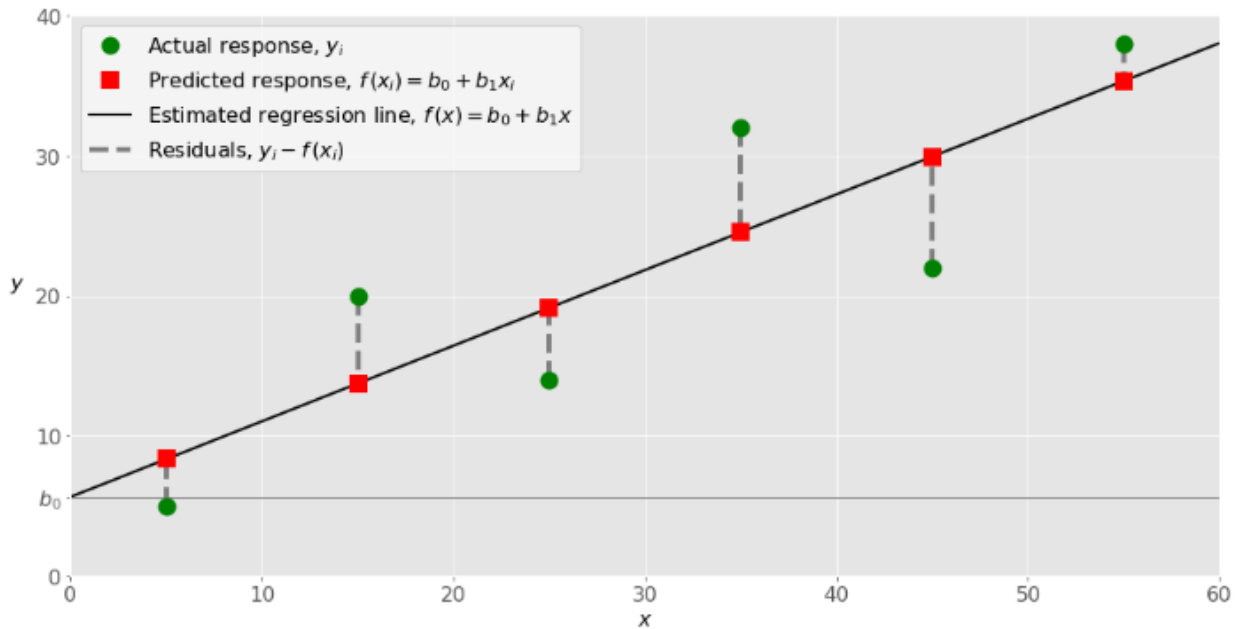


Figura 7. Regresion lineal

Para aplicar este algoritmo de aproximación, existen métodos variados en Python, ya compilados para ser empleados y obtener dichos valores:

- Statsmodels
- Scikit-Learn
- Scipy
- Numpy

Gradient Descent

Uno de los mas populares algoritmos de optimización, esta fundamentado en la iteración para la obtención de un minimo de determinada función, el calculo diferencial es constructor de este algoritmo.

La función de costo, definida como

$$J(x_1, x_2) = \frac{1}{2n} \sum_{i=1}^n (y(x^i) - y^i)^2$$

Donde se aprecia que es promedio de la sumatoria de los errores al cuadrado, siendo esta la función a optimizar.

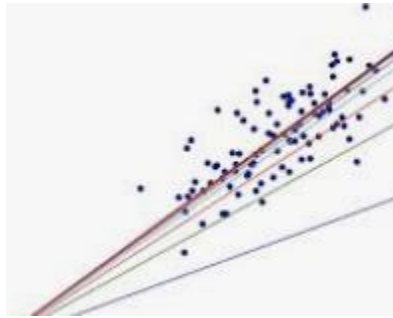


Figura 8. Gradient Descent

Intuitivamente se puede apreciar que el espaciamiento entre iteraciones esta dictado por una razón de aprendizaje definida por un “ α ”, así mismo, la pendiente, en cálculo diferencial esta definida como una pendiente, a una razón de cambio, incluyendo estos dos conceptos podemos llegar al valor a iterar:

$$x_j = x_j - \alpha \frac{\partial}{\partial x_j} (J(x_1, x_2))$$

En la figura se muestra la grafica de la función costo, y su respectiva derivada representada por la pendiente, optimizando hasta obtener un mínimo o un error mínimo, la letra “w” representando los inputs dentro de la función costo, también se muestra la razón de aprendizaje representada por la separación entre iteraciones.

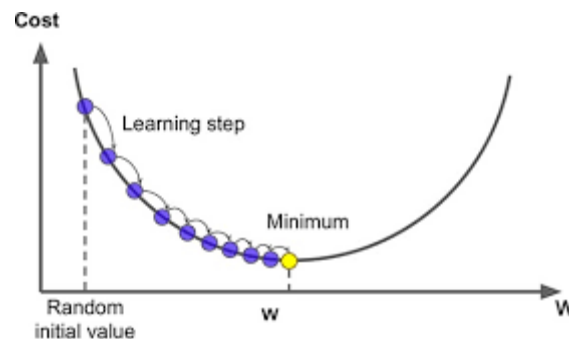


Figura 9. Learning rate

Capítulo 5 : Declinación

El análisis de curvas de declinación ha sido uno de los métodos mas utilizados para predecir la producción futura de campos petroleros (Rojas, 2003), siendo estas un metodo de calculo de reservas dinámico, estos análisis son fundamentales para el análisis económico.

El método de la extrapolación es el efecto de estimar el rendimiento futuro, dependiendo del gasto en algún momento en particular, así como del comportamiento de la producción, y el factor de declinación.

Arps en 1945, definio tres tipos de declinación, representadas en la siguiente figura.

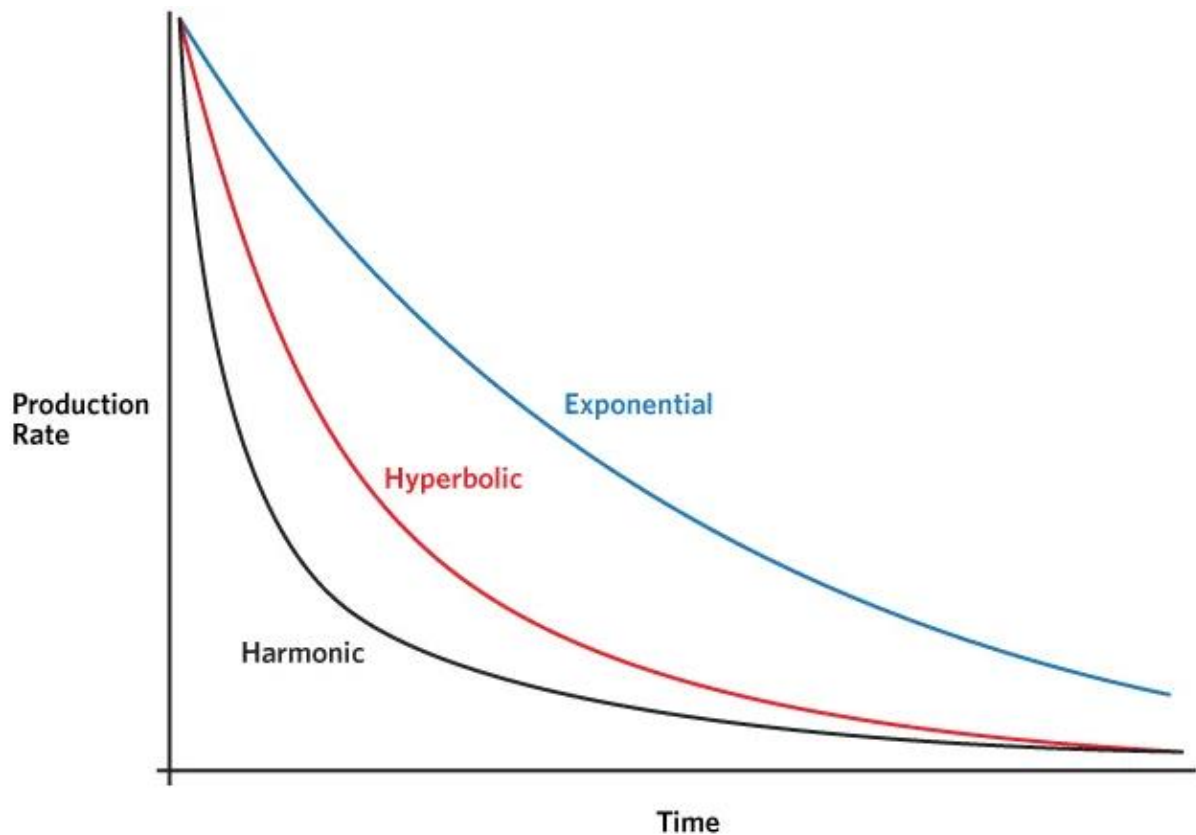


Figura 10. Curvas de declinación

Esta técnica es un análisis de datos que ajusta la historia de producción con un modelo matemático, siendo la declinación exponencial la mas utilizada ya que supone una tendencia conservadora.

Con estos datos es posible la obtención del tiempo de abandono de algún yacimiento en especial.

Capítulo 5: Estudio de caso

Campo Volve

En el año 2018, la operadora petrolera junto con los operarios del campo Volve, ubicado en el mar territorial de Noruega liberaron datos referentes al mismo.



Figura 11. Campo Volve

Campo producido desde el año 2008 al 2016, produciendo 56mil barriles por día, aportando una totalidad de 63millones de barriles a lo largo de su vida productiva.

Se otorgo al publico los datos de perforacion, petrofisica, produccion asi como de simulacion y sismica para el estudio por parte de estudiantes o instituciones de investigacion y contribuir al mejoramiento de practicas y procesos.

Para abordar el tema de la presente investigacion se utilizaron los datos proporcionados de produccion total diaria, para su posterior analisis en la obtencion de la tasa de declinacion, empleando algoritmos ofrecidos por las librerias de aprendizaje automatico supervisado.

Pre-procesamiento de datos

Los datos de producción se muestran en la figura 8, también dentro de la base de datos se encuentran mediciones de presión, temperatura, radio de tubería de producción, corte de agua, por mencionar algunos, figura 7.

Daily Production

Columns:

```
Index(['DATEPRD', 'WELL_BORE_CODE', 'NPD_WELL_BORE_CODE', 'NPD_WELL_BORE_NAME',  
      'NPD_FIELD_CODE', 'NPD_FIELD_NAME', 'NPD_FACILITY_CODE',  
      'NPD_FACILITY_NAME', 'ON_STREAM_HRS', 'AVG_DOWNHOLE_PRESSURE',  
      'AVG_DOWNHOLE_TEMPERATURE', 'AVG_DP_TUBING', 'AVG_ANNULUS_PRESS',  
      'AVG_CHOKE_SIZE_P', 'AVG_CHOKE_UOM', 'AVG_WHP_P', 'AVG_WHT_P',  
      'DP_CHOKE_SIZE', 'BORE_OIL_VOL', 'BORE_GAS_VOL', 'BORE_WAT_VOL',  
      'BORE_WI_VOL', 'FLOW_KIND', 'WELL_TYPE'],  
      dtype='object')
```

Figura 12. Columnas

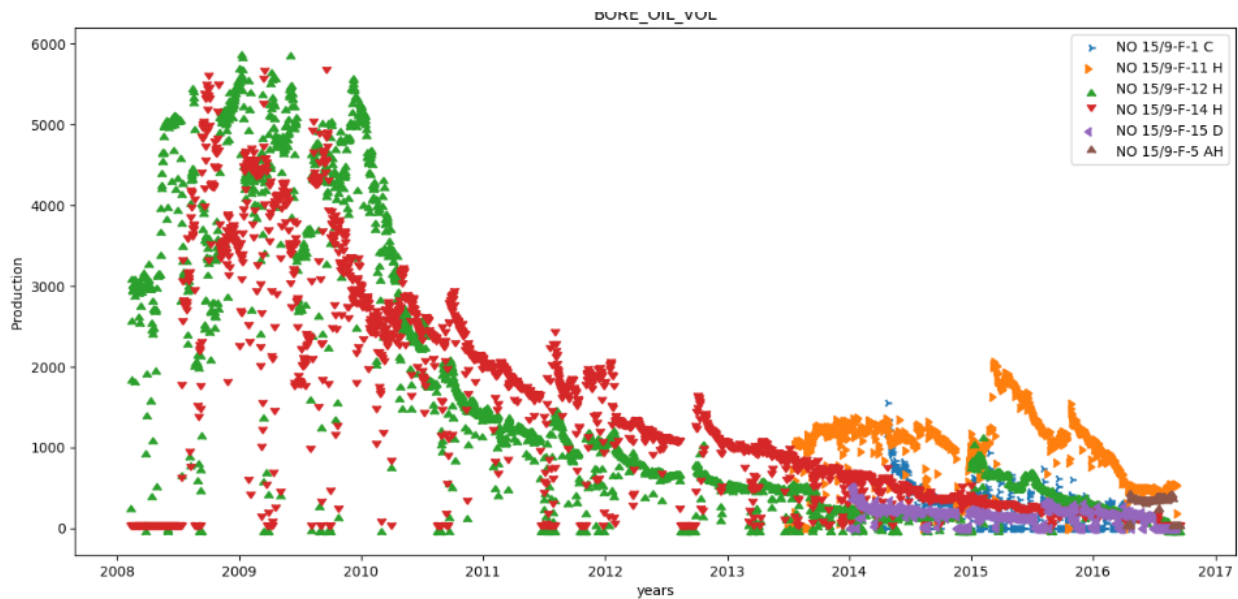


Figura 13. Produccion Diaria por pozo

Debido a la densa cantidad de datos ofrecidos por la compañía se decidió acotar el estudio a un solo pozo productor denominado "NO 15/9-F-14 H". para poder apreciar de mejor forma la declinación.

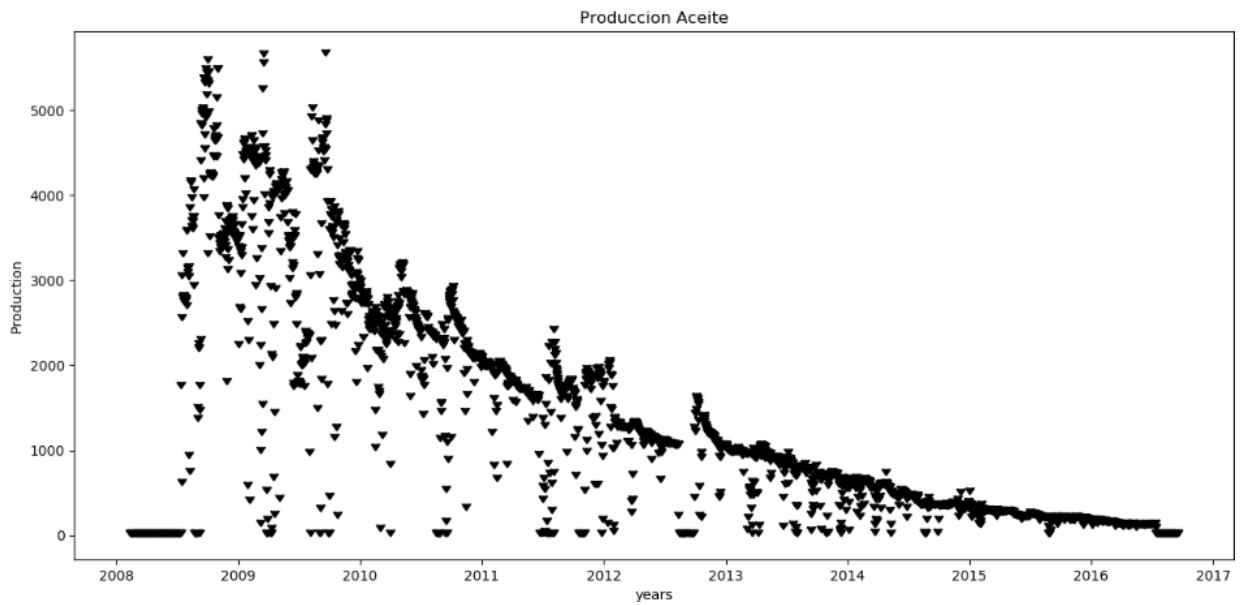


Figura 14. Produccion pozo

Un análisis mas minucioso y que muestra el potencial de la ciencia de datos, puede encontrarse en el anexo A.

Al emplear analisis de datos y limpieza de datos se pudo obtener el siguiente DataFrame, conteniendo los datos requeridos para la determinación de la tasa de declinación y luego obtener el tiempo de abandono.

Tabla 1. DataFrame inicial

	Dias	BORE_OIL_VOL
0	1	0.0
1	2	0.0
2	3	0.0
3	4	0.0
4	5	0.0
5	6	0.0
6	7	0.0
7	8	0.0
8	9	0.0
9	10	0.0

Para este método se necesito obtener la siguiente tabla.

Tabla 2. Declinacion

	Tiempo	BORE_OIL_VOL	dQo	Di	Log(Qo)
151	152	1735.26	1735.26	1.000000	3.239385
152	153	582.36	-1142.90	-1.929401	2.772586
153	154	2532.21	1939.85	0.768070	3.403500
154	155	3030.71	498.50	0.164483	3.481544
155	156	3288.08	257.37	0.078274	3.516942

Para la obtención de esta, se utilizaron las siguientes ecuaciones:

Obtener el diferencia de presión

$$\Delta Q_o = Q_{o_i} - Q_{o_{i-1}}$$

La razón de la diferencia de producciones sobre la producción original

$$D_i = \frac{\Delta Q_o}{Q_o}$$

El la obtención del logaritmo para graficar y obtener la regresión lineal en ejes semilogaritmicos

$$\log Q_o$$

Y graficando lo obtenido:

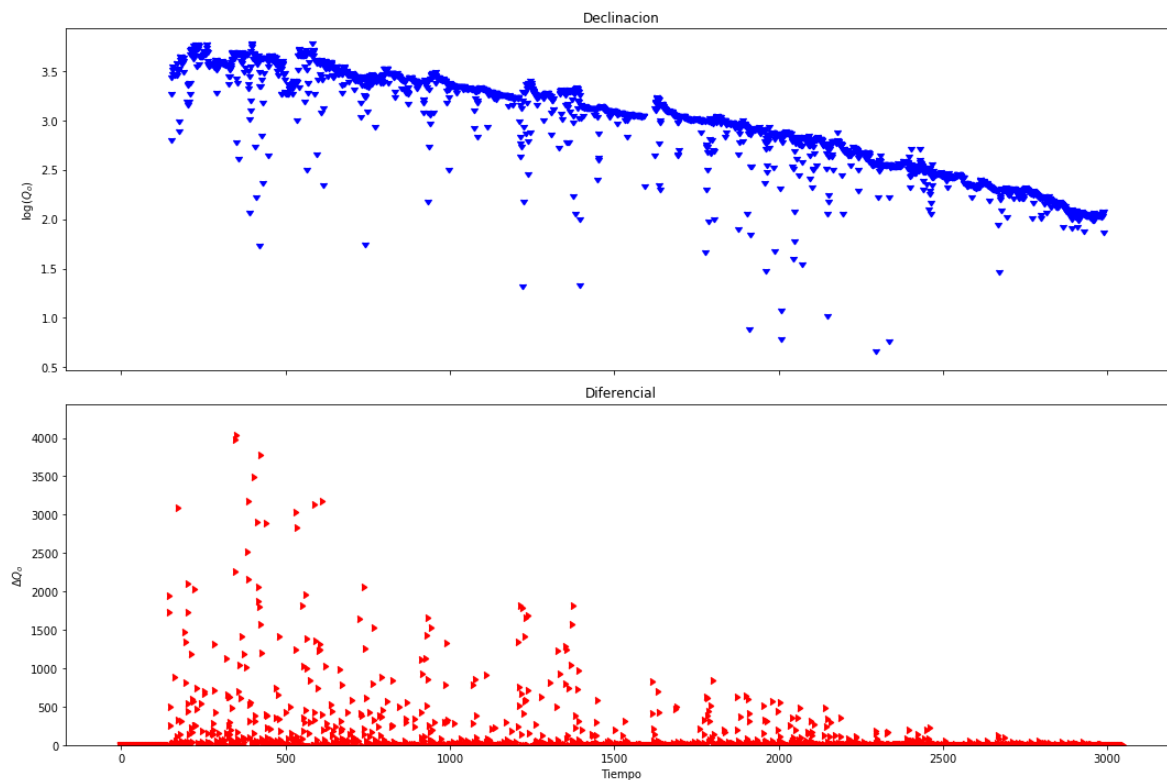


Figura 15. Graficos de declinación

Estas graficas muestran la declinación (superior) del campo en estudio, y la variación de gastos(inferior), esta variación del gasto se concentran mas en el eje inferior, por lo que prevalecen las variaciones leves, indicando una producción mas o menos estable, esto visualizado de una forma probabilística mediante el boxplot, instrumento explorado mas adelante.

Modelado

Teniendo la tabla de datos relevantes para nuestro estudio, se analizo preliminarmente para ver la incidencia de los datos mediante el boxplot, aporte de los métodos estadísticos para un analisis de los datos, en ellos se pueden apreciar los “outliers” o datos que nos ocasionan ruido e inexactitud en la aplicación de la regresión, en los gastos, esto pudiendo ser ocasionado por un desperfecto en el medidor, fluctuaciones del gasto a consiencia de los operarios o por alguna cuestión de infraestructura.

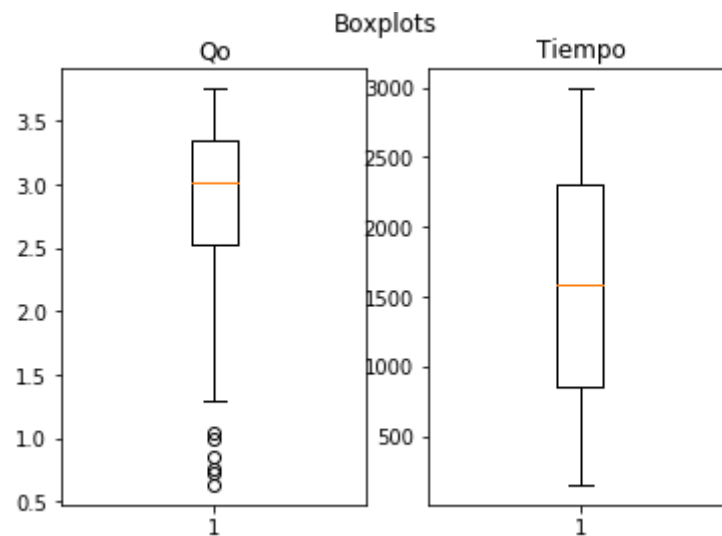


Figura 16. Boxplot

Se procedio al calculo con la librería statsmodel, en el siguiente código se aprecia la metodología para su uso y aprovechamiento.

Librerías

```
1 import statsmodels.formula.api as smf
```

Construccion del modelo

```
1 lm1 = smf.ols(formula="y~x", data = DFF).fit()
2 lm1.params
```

```
28]: Intercept      3.812261
x                -0.000554
dtype: float64
```

```
1 print("Modelo lineal:")
2 print("y = -0.000554 x + 3.812261\n")
3 print("Factor de correlacion:")
4 print("R^2 =", lm1.rsquared)
```

```
Modelo lineal:
y = -0.000554 x + 3.812261
```

```
Factor de correlacion:
R^2 = 0.8357514960741332
```

```
1 DF["StatsModels"] = -0.000554*DF["Tiempo"] + 3.812261
```

La importación de librerías, es un detalle de importancia de Python, por lo que cada vez que se utiliza alguna, esta deberá importarse manualmente, ya obtenido el DataFrame de datos, se crea un objeto denominado "lm1" donde será realizado el procedimiento de regresión aproximando los datos "y" a partir de "x" de nuestro DataFrame.

Los parámetros de "intercept" y los de x corresponden a la ordenada al origen y a la pendiente respectivamente, mientras que el factor de correlacion nos muestra, que tanto se acoplo nuestro modelo a los datos reales.

A continuación se importo la librería de Scikit_Learn.

Librerías

```
In [49]: 1 from sklearn.linear_model import LinearRegression
2 from sklearn import metrics
```

Construccion del modelo

```
In [50]: 1 X = DFF.iloc[:, :-1].values
2 y = DFF.iloc[:, 1].values
3 print("X:\n",X)
4 print("y:\n",y)

X:
[[ 152]
 [ 153]
 [ 154]
 ...
 [2908]
 [2909]
 [2990]]
y:
[3.23936456 2.77258572 3.40349972 ... 2.03678853 2.05629489 1.83550033]
```

```
In [51]: 1 lm2 = LinearRegression()
2 lm2.fit(X,y)
```

```
Out[51]: LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None, normalize=False)
```

```
In [52]: 1 print("Modelo lineal:")
2 print("y =",lm2.coef_,"x +",lm2.intercept_)
3 print("\nFactor de correlacion:")
4 print("R^2 =",lm2.score(X,y))
```

```
Modelo lineal:
y = [-0.00055423] x + 3.8122609307297797
```

```
Factor de correlacion:
R^2 = 0.8357514960741332
```

En esta se muestra la importancia del manejo de la ciencias básicas, tales como algebra lineal, esta librería hace uso de vectores o "arrays" para el procesamiento de estos datos, para tener un buen desempeño en la obtención de los parámetro de nuestra regresio, que se pueden consultar mediante los comandos de "coef_" para la pendiente y "intercept_" para la ordenada al origen, sin olvidar el respectivo factor de correlacion "score"

Asi mismo, se muestra otra herramienta útil y practica para la toma de decisiones, cuantos valores son necesarios para una regresión lineal, para ejemplificar esto, se utilizo el método de Training y Testing sets.

Training y Testing sets

```
In [59]: 1 from sklearn.model_selection import train_test_split
2 X = DFF.iloc[:, :-1].values
3 y = DFF.iloc[:, 1].values
4 X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=.99, random_state=0)
```

```
In [60]: 1 lm22 = LinearRegression()
2 lm22.fit(X_train, y_train)
```

```
Out[60]: LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None, normalize=False)
```

```
In [61]: 1 print("Modelo lineal:")
2 print("y =",lm22.coef_,"x +",lm22.intercept_)
3 print("\nFactor de correlacion:")
4 print("R^2 =",lm22.score(X,y))
```

```
Modelo lineal:
y = [-0.00062492] x + 3.93411093210168
```

```
Factor de correlacion:
R^2 = 0.8217661661127307
```

En este se muestra el método para definir un "Training set" donde el algoritmo encontrara los parámetros pertinentes de regresión para después aplicarlos en un "Testing set" esto para contrarestar el fenómeno denominado "overfitting" que consiste en la falsa percepción de que el modelo es correcto, teniendo un buen

factor de correlacion, y a su vez no reflejando el fenómeno correctamente, eso es mas visible en las regresiones multiples.

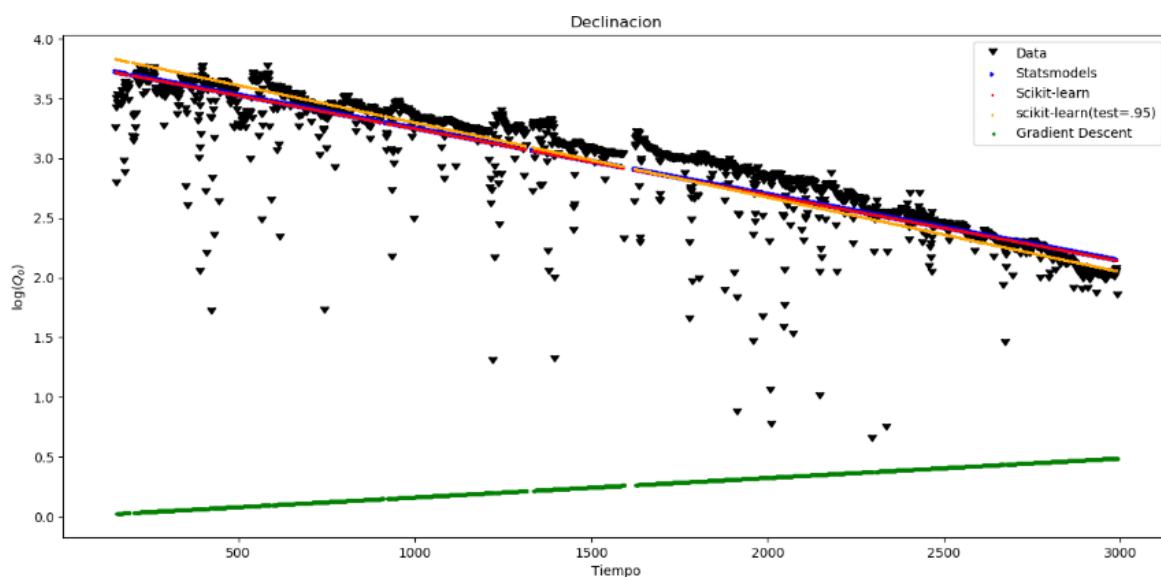
Graficado

Se obtuvieron los siguientes datos:

Tabla 3. Modelos

Tiempo	BORE_OIL_VOL	Log(Qo)	StatsModels	Scikit-Learn	Scikit-Learn2.1	GradientDescent	
151	152	1735.26	3.239365	3.728053	3.728018	3.839123	0.026154
152	153	592.36	2.772586	3.727499	3.727464	3.838498	0.026318
153	154	2532.21	3.403500	3.726945	3.726910	3.837873	0.026481
154	155	3030.71	3.481544	3.726391	3.726355	3.837248	0.026644
155	156	3288.08	3.516942	3.725837	3.725801	3.836623	0.026808
...
2985	2986	102.09	2.008983	2.158017	2.157330	2.068100	0.488830
2986	2987	113.38	2.054536	2.157463	2.156776	2.067475	0.488994
2987	2988	108.84	2.036789	2.156909	2.156222	2.066850	0.489157
2988	2989	113.84	2.056295	2.156355	2.155667	2.066225	0.489320
2989	2990	68.47	1.835500	2.155801	2.155113	2.065600	0.489483

2723 rows × 7 columns



Una vez determinada el modelo mas adecuado en función de su R^2 , se obtuvo el mas adecuado, siendo el de Scikit-Learn el mas veloz y mas preciso, a continuación, dichos parámetros se

utilizaron en la construcción de la recta que representa la declinación para luego obtener el tiempo de abandono, a partir de este, y teniendo un Gasto deseado definido.

Determinacion de tiempo de abandono

$$Q_o = 85,000[bpd]$$

Modelo lineal

$$\log(Q_o) = -0.00055423(T) + 3.8122609307297797$$

$$T = \frac{3.8122609307297797 - \log(Q_o)}{-0.00055423}$$

Y posteriormente, determinar las reservas producidas a partir de un gasto de abandono, y realizar el analisis económico correspondiente al fin de la vida productiva del pozo.

Determinacion de Reservas producidas

$$NOAB = \frac{Q_{in} - Q_p}{D_i}$$

Gasto de abandono: $Q_p = 100[bpd]$

Capitulo 6: Resultados

Tabla 4. Resultados

	StatsModels	Scikit_Learn	Scikit_Learn2.1	GradientDescent
Tiempo	35 ms	3 ms	2.82 ms	12min 6s
R^2	0.83575	0.83575	0.82176	-

Estos resultados muestran la utilidad y rapidez de estos algoritmos cuyo fundamento establecen el principio de la inteligencia artificial en el modelado numérico.

Días para el abandono: 2015.6938364659316
 Meses para el abandono: 65.02238182148166
 Años para el abandono: 167.97448637216095

Los resultados obtenidos a partir de los parámetros de regresión fueron utilizados para la determinación del tiempo de abandono, los cuales muestran una discrepancia con la realidad, ya que el campo ya no se produce, sin embargo como se menciona, este método es de altas expectativas, así mismo, las cualidades y propiedades petrofísicas y de flujo de fluidos, dictaminan en la realidad la explotación de un campo, hecho que resulta evidente con estos resultados.

Se concluye también que, los algoritmos ofrecidos por el aprendizaje automatico, así como los métodos de análisis ofrecidos por la ciencia de datos son una herramienta que cada vez adquiere una plusvalía en el sector petrolero gracias a su utilidad, practicidad y rapidez.

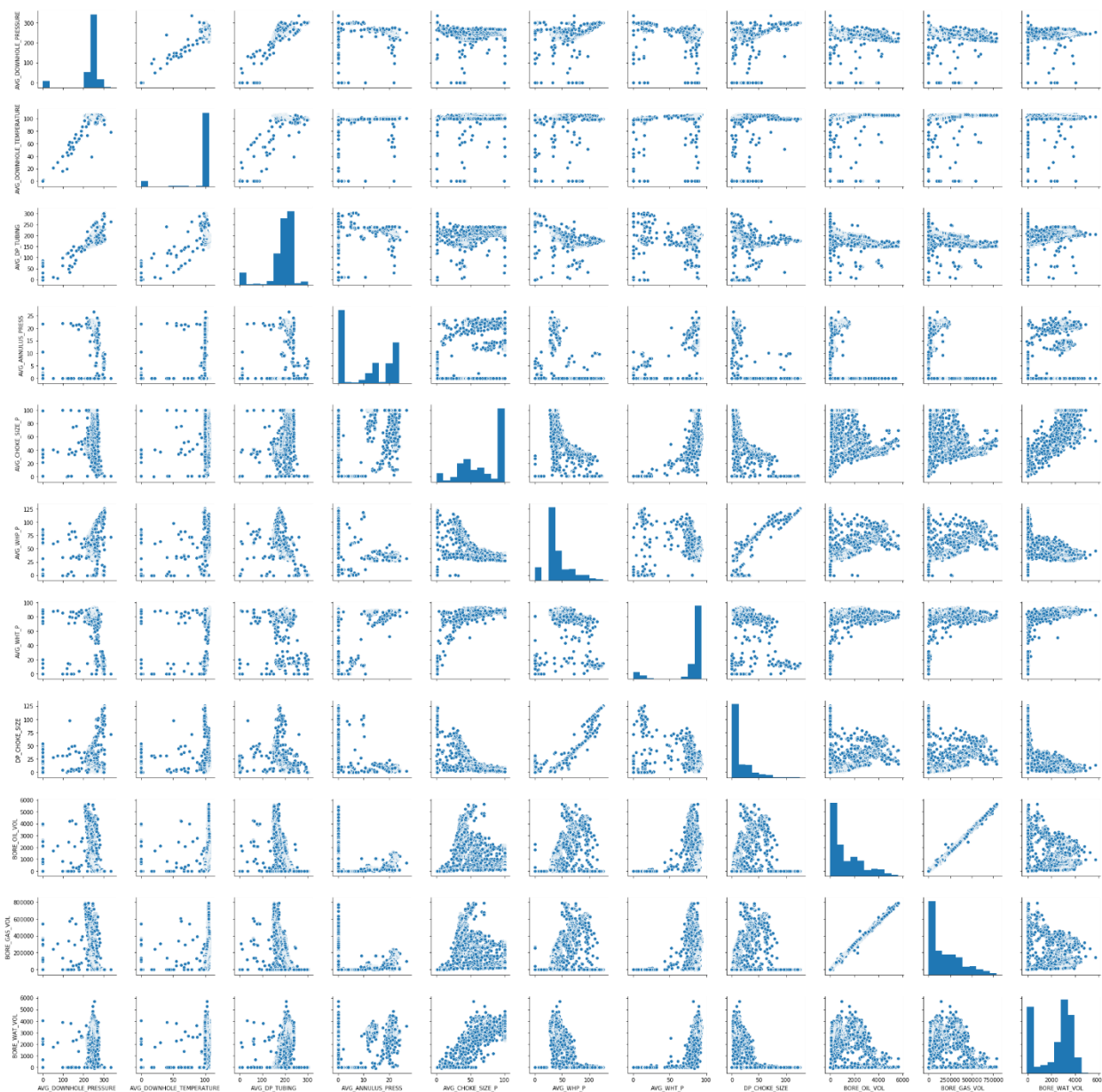
Estos datos deberán de ser analizados de una forma concreta y adecuada dada su cantidad y sus posibles aportes a la toma de decisiones.

Referencias

- Rivera,L.E.. (2015). Apuntes de Recuperacion Secundaria y Mejorada. Mexico: Tesis, UNAM.
- Brownlee, J.. (2016). Master Machine Learning Algorithms. USA: SN.
- Bhasin, H.. (2019). Python Basics. USA: New Age International.
- Johansson, R.. (2019). Numerical Python . Japan: Apress.
- Elichev, V., Bilogan, A. & Livinenko,K. (2019).SPE-196861-MS Understanding Well Events with Machine Learning, USA, Society of Petroleum Engineers
- Noshi,C. & Schubert, J. (2018). SPE-191823-18ERM-MS The Role of Machine Learning in Drilling Operations; A Review, USA, Society of Petroleum Engineers.
- Yang, J., Sun, T. & Zhao, Y. (2019). ISOPE-I-19-02 Advanced Real-time Gas Kick Detection Using Machine Learning Technology, USA, International Society of Offshore and Polar Engineers.
-
- Curso : Recuperacion Secundaria y Mejorada, Alfredo Leon Garcia, UNAM, 2020
- Curso : Machine Learning, Andrew N.G., Stanford University, 2020

Anexo A

La ciencia de datos ofrece una herramienta invaluable para el análisis de muy densos datos, a continuación se muestra una librería de Python llamada “seaborn” que nos permite visualizaciones sobre las relaciones entre los datos ordenados de nuestra base de datos.



En ella nos muestra en la línea de identidad la distribución probabilística que obedece cada variable de nuestra base de datos, así como relaciones entre las mismas. Importante mencionar la relación casi lineal entre la relación de producción de aceite y la producción de gas, así como la relación que existe entre la presión del cabezal y el diámetro del estrangulador, relaciones que el ingeniero puede encontrar y analizar para un sinfín de inferencias y toma de decisiones.

También nos apoya en la elaboración de una matriz de correlación, matriz útil para encontrar correlaciones numéricas de primera mano.

