

EL CAMINO HACIA UNA EXPLOTACIÓN DE DATOS SOSTENIBLE

Gonzalo Mallo, PETROBRAS ARGENTINA S.A. E-mail gonzalo.mallo@petrobras.com

Pablo Durante, PETROBRAS ARGENTINA S.A. E-mail pablo.durante@petrobras.com

Alejandro Bascur, PETROBRAS ARGENTINA S.A. E-mail alejandro.bascur@petrobras.com

Eduardo Zenon García, PETROBRAS ARGENTINA S.A. E-mail eduardo.z.garcia@petrobras.com

PETROBRAS ARGENTINA S.A. Maipú 1, Piso 4, Buenos Aires Argentina

Keywords: **Explotación de Datos, Analytics, Fracturas, Reservorios,**

RESUMEN

El objetivo del trabajo es explicitar las etapas necesarias para alcanzar de manera sostenible la explotación y gobernanza de datos en la disciplina de E&P, determinándose como objetivo particular la selección de pozos a ser fracturados para maximizar la producción acumulada.

Como punto de partida se adoptó el estándar de datos PPDM como modelo para la base de datos integrada. En paralelo se especificaron los sistemas fuentes de datos y luego se desarrollaron las interfaces de carga (ETL). Se realizó a posteriori un proceso de caracterización y "data cleansing". Se seleccionaron los datos de un yacimiento para ser usado como piloto.

La plataforma usada para el análisis y predicción de datos fue *Spotfire Analyst & Miner* con la cual se logró un modelo estable que predice los pozos con potencial para maximizar su producción. Este resultado será estudiado en mayor detalle por el equipo de especialistas de Ingeniería de Reservorios.

Se destaca la importancia que para alcanzar el objetivo se debe disponer de la mayor cantidad de datos e información y contar con la ayuda de un equipo interdisciplinario.

Expondremos una aplicación práctica donde se muestra la metodología propuesta y las herramientas usadas para alcanzar este objetivo.

ABSTRACT

The goal of this paper is to achieve sustainable data mining and data governance in the discipline of E & P. We determined as particular target maximize the cumulative production of wells that will be fractured.

As a starting point the standard PPDM data model for the integrated database was adopted. At the same time, we specified the data sources systems and after we developed the load interfaces (ETL). Was performed a characterization process and "data cleansing". We selected the data of one oilfield to be used as a pilot.

The platform used for the analysis and prediction data was Analyst & Spotfire Miner with which a stable model was achieved to predict wells where could be maximize production.

This result will be studied in greater detail by the team of specialists of Reservoir Engineering.

We must emphasize the importance of having the largest amount of data and information and with the help of an interdisciplinary team.

We present a practical application in this paper where the proposed methodology and the tools used to achieve our goal is shown.

INTRODUCCIÓN

El crecimiento exponencial de los datos y la gran variedad de aplicativos de negocios enfrentan a los analistas de negocios al desafío de utilizar de manera eficaz y eficiente los datos producidos, en pos de optimizar el proceso de toma de decisiones basado en información precisa y conocimiento accionable que permita mejorar el retorno de inversión de los proyectos.

La industria del petróleo y gas no es una excepción a esta regla, contando con un gran volumen de datos, con una amplia variedad de aplicativos, procesos y necesidad de toma de decisiones que implican muchas veces la ejecución de presupuestos importantes.

Esta situación en su conjunto genera una complejidad en los procesos de decisión.

El objetivo del trabajo es definir un proceso y sus etapas necesarias para alcanzar una explotación y gobernanza de datos sostenible en la disciplina de E&P. De manera particular se trabajó sobre

el objetivo puntual de la generación de sugerencias para la selección de pozos a ser fracturados que logren maximizar la producción acumulada de hidrocarburos a modo de ejemplo de aplicación.

De manera general el proceso definido en este trabajo consta de las siguientes etapas:

- Definición de un equipo de trabajo multidisciplinario.
- Determinación de las necesidades del negocio.
- Selección de estándar de almacenamiento.
- Definición de la metodología para la explotación de datos.
- Selección de las plataformas tecnológicas que soporten el proceso.
- Implementación de la interfaces de integración y procesamiento
- Desarrollo de modelos analíticos
- Integración del producto desarrollado en el proceso de toma de decisiones.

Base integrada

La base integrada centraliza los datos de las diferentes fuentes de una manera estandarizada y consolidada de manera de cumplir con dos objetivos básicos: satisfacer las necesidades del negocio de E&P y ser un repositorio único de datos.

PPDM

PPDM (*Professional Petroleum Data Management*) es un estándar definido y creado por la PPDM Association para ayudar a las empresas de gas y petróleo a administrar sus datos de exploración y producción.

PPDM *Data Model* es un modelo de datos relacional robusto diseñado por expertos en distintas áreas de aplicación de la industria del petróleo, profesionales del *data management*, desarrolladores de soluciones, entes reguladores, etc. que atiende 53 temáticas diferentes de la industria, cubriendo un espectro muy amplio.

ETL

El término ETL (*Extract, Transform & Load*) se refiere al proceso de extracción de datos de sus fuentes originales, transformación de los mismos con el fin de adecuarlos al formato definitivo, y finalmente cargarlos en la base de datos destino.



Fig. 1: Diagramado del proceso de ETL.

Inteligencia analítica

Esta especialidad es el resultado de la confluencia orgánica de diferentes disciplinas tales como la estadística, aprendizaje automático, visualización de información, optimización matemática y bases de datos.

Alternativamente se define como el proceso que tiene como fin “explotar” los datos almacenados en los sistemas de la empresa, “extraer” patrones estables allí ocultos, “expresarlos” en modelos legibles, operativos y accionables para luego “transformarlos” en conocimientos del negocio que mejoren la rentabilidad del mismo.

Las técnicas empleadas en esta especialidad son comúnmente técnicas de predicción y descripción de datos, tales como regresiones lineales y logísticas, arboles de clasificación y

regresión, redes neuronales, algoritmos bayesianos, técnicas de segmentación y/o de reducción de dimensionalidad, asociación, pronósticos, etc.

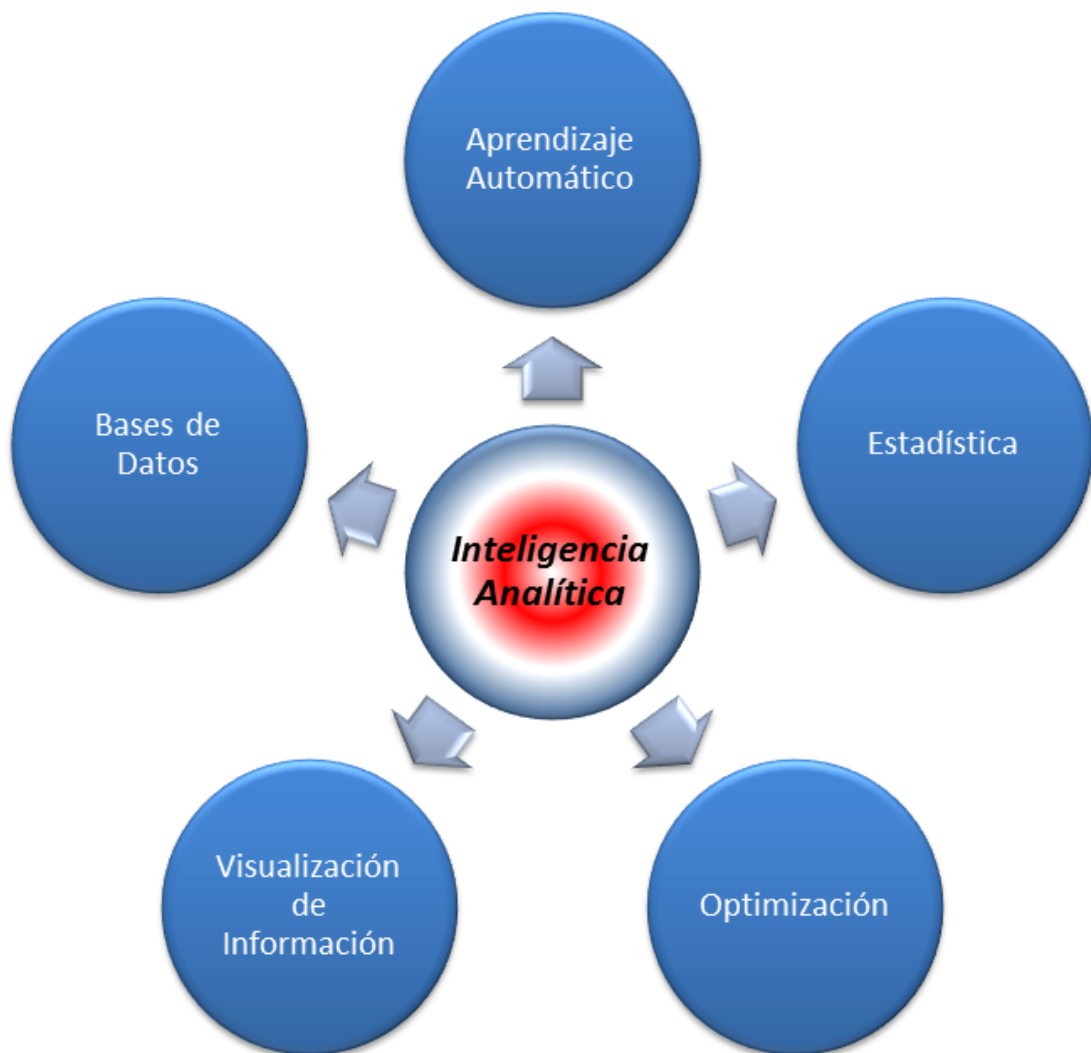


Fig. 2a: Componentes de la Inteligencia Analítica

Estas técnicas son empleadas para resolver problemas enmarcables en las siguientes tareas:

- Segmentación.
- Clasificación.
- Estimación.
- Asociación y secuenciación.
- Pronóstico.
- Visualización.

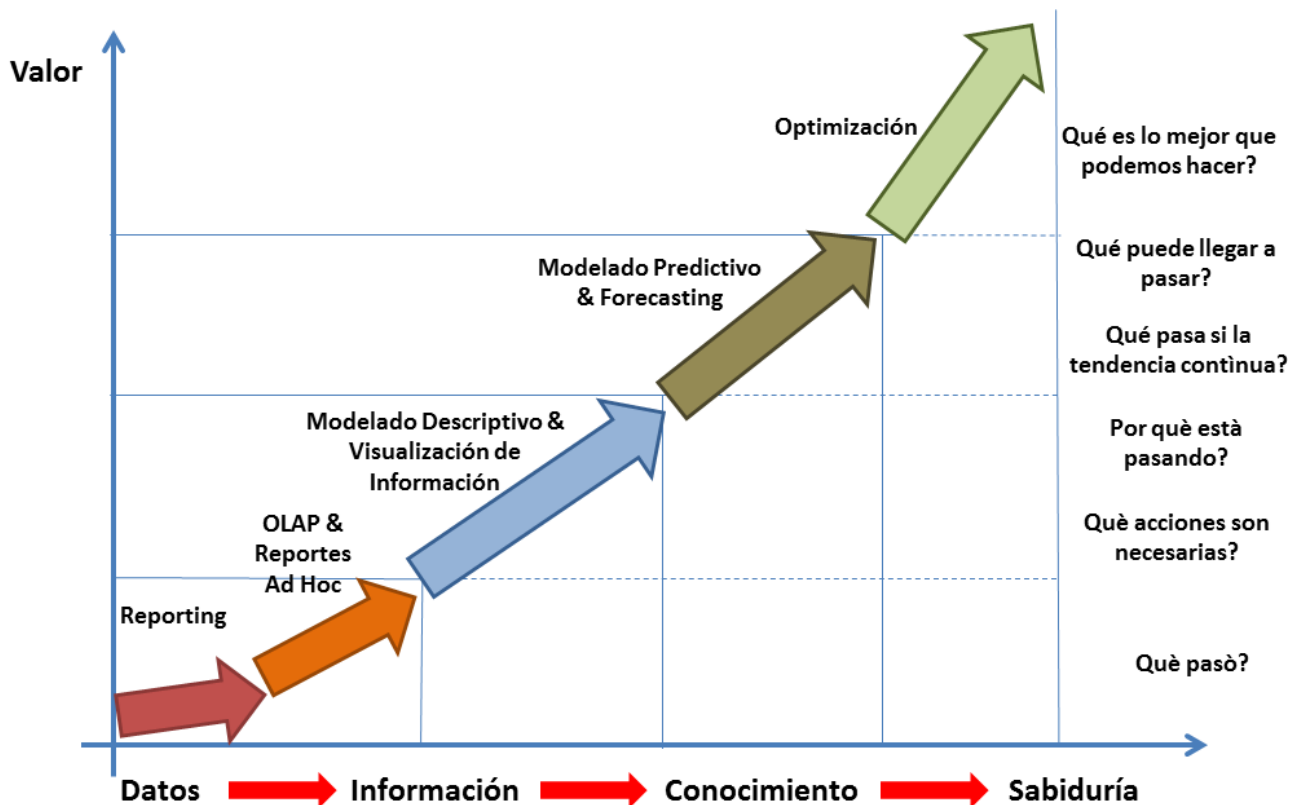


Fig. 2b: Evolución del Valor Aportado por el análisis de datos.

CRISP-DM

Acrónimo de *Cross Industry Standard Process for Data Mining*. Se trata de un modelo de proceso de minería de datos que describe los enfoques comunes que utilizan los expertos en esta materia. La metodología se compone de las siguientes instancias a saber:

- Comprensión del Negocio: Entender el objetivo y problema a resolver.
- Comprensión de los Datos: Conformer un equipo de interdisciplinario que nos permita entender las variable que intervienen en la solución del problema.
- Preparación de Datos: El arte y la clave del éxito en el desarrollo de modelos analíticos.
- Modelado: Es una representación simplificada de la realidad que nos permite comprender algunos aspectos de un determinado problema. (Ej.: Ecuaciones, Reglas generales o Puntuales, Diagramas, etc.)
- Evaluación: La validación de que el modelo funciona
- Implementación: La puesta en producción – El punto donde se genera valor.
- Monitoreo: El proceso de control (El modelo, funciona aún?).

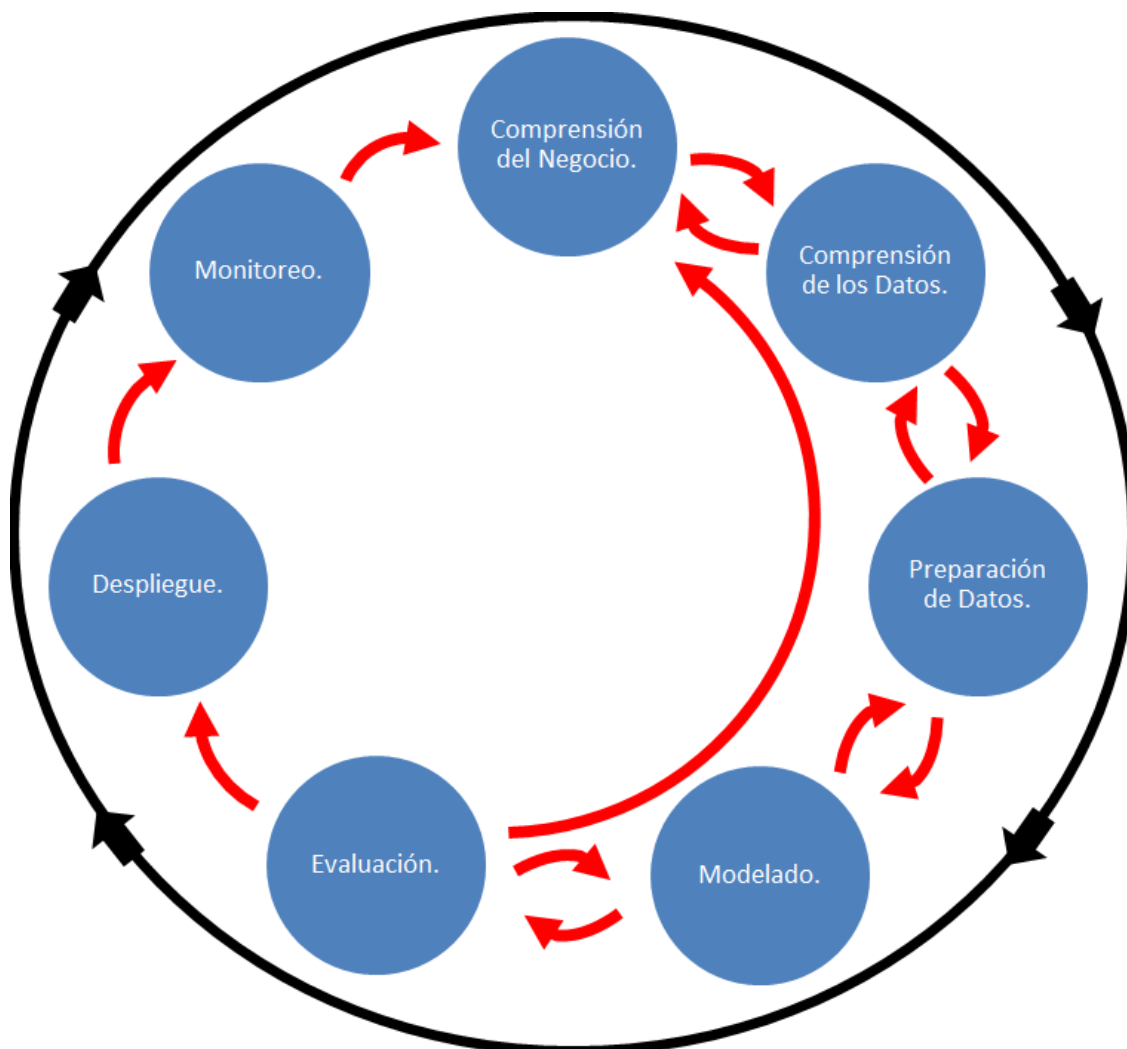


Fig. 3: Metodología CRISP-DM adaptada.

DESARROLLO

Definición de un equipo de trabajo multidisciplinario.

En Petrobras conformamos un equipo multidisciplinario integrado por personas con distintos perfiles, entre ellos, ingenieros de reservorios y producción, gestores de datos, consultores funcionales, ingenieros en sistemas y especialistas en inteligencia analítica.

Determinación de las necesidades del negocio.

El punto de partida de este estudio, definido por el equipo de trabajo, fue la búsqueda de una solución que se adecue a la satisfacción de las necesidades relevadas del negocio, entre ellas:

- Contar con una metodología de explotación de datos.
- Establecer la gobernanza de datos.
- Disponer de una solución técnica y funcional para el problema de calidad de datos.
- Contar con un repositorio de datos único y centralizado.
- Unificar el lenguaje con una nomenclatura estándar.
- Dar soporte al proceso de toma de decisiones.
- Disponer de los *Key Performance Indicators* (KPIs) en tiempo y forma.

De manera puntual y con el fin de evaluar tanto herramientas como metodología se eligió para el proyecto piloto un set de datos de casos de fractura pertenecientes al yacimiento Medanito de la cuenca Neuquina. Los datos pertenecen a las formaciones de Choiyoi, Petrolífera y Quintuco. Nuestros especialistas de Ingeniería de Reservorios proveyeron de las características de esta zona:

- Muy buenas propiedades petrofísicas en dolomías gruesas y finas de Quintuco Superior que aseguran una buena conectividad entre pozos vecinos.
- Las capas Complejo Superior y Capa 2 (Qco. Sup.) contienen el 77% del POIS del proyecto y ahí centramos la selección.
- Una secundaria bien desarrollada asegura tener medianamente presurizados los reservorios.
- Fracturas con una edad promedio de 14 años que pueden mejorarse con una nueva estimulación.

Selección de estándar de almacenamiento.

El equipo de trabajo estudió el estándar definido por PPDM concluyendo que este modelo seleccionado corporativamente por Petrobras Brasil cubría mayormente las necesidades del negocio local de E&P.

El proceso de análisis detallado de este estándar determinó que el modelo PPDM representa adecuadamente nuestro modelo de producción, en tanto que debió extenderse para satisfacer los requerimientos de datos del modelo de reservorios de Petrobras.

Del modelo estándar de PPDM se utilizaron los siguientes componentes:

- Áreas
- Campos
- Pozos (estados y clases)
- Punzados
- Completaciones
- Estratigrafías
- Producción
 - Oil
 - Gas
 - Por formación
 - Mensual
- Well Tests
- Fracturas y tratamientos
- Instalaciones

La extensión del modelo, propuesta por el equipo de trabajo sigue los lineamientos del estándar PPDM en cuanto al diseño de tablas y relaciones entre ellas.

Las adaptaciones se hicieron en torno a la información de:

- Propiedades petrofísicas
- Permeabilidades relativas
- PVT
- PLT
- Diseño y resultado de estimulaciones
- Pre-tratamientos en fracturas

Definición de la metodología para la explotación de datos.

Tras un análisis detallado de la información disponible el equipo concluyó que la metodología a seguir más se adecuaba a los procesos de nuestra empresa es la denominada CRISP-DM. Esta metodología fue ampliada agregándose de manera explícita una etapa de monitoreo de resultados considerada como imprescindible en el seguimiento permanente de los modelos analíticos

desarrollados. Esta instancia se incorporó para evitar problemas de pérdida de performance por envejecimiento de los modelos desarrollados puestos en producción.

Selección de las plataformas tecnológicas que soporten el proceso.

La base de datos de PPDM se instaló sobre un servidor Oracle 11g. Cada una de las interfaces de transformación de datos se implementó en *Sql Server Integration Service* 2008 (SSIS), la solución de Microsoft para la integración y transformación de datos.

Este conjunto de herramientas aportó las capacidades de almacenamiento, automatización y *scheduling* de los datos e interfaces.

El desarrollo de cada una de estas interfaces fue hecho con Microsoft Visual Studio 2008.

SSIS 2008 provee una gran cantidad de conectores a diferentes tipos de bases de datos y archivos, esto nos permite poder tomar diferentes fuentes de datos.

Estas tecnologías se seleccionaron en función del *know-how* técnico del equipo y disponibilidad tecnológica de la compañía.

Como plataforma de *analytics* se seleccionó TIBCO *Spotfire* (*Spotfire Analyst* y *Spotfire Miner*) a la cual el equipo consideró como la más apropiada y completa suite de *analytics*.

Implementación de la interfaces de integración y procesamiento.

Establecidos los datos a incorporar al estándar de PPDM se determinaron las fuentes de origen de los datos para poblar el modelo PPDM implementado.

En el proceso de integración intervienen personas con diferentes roles y perfiles, ya que es un proceso continuo e iterativo en el que se realizan las diferentes acciones para sanear los datos y asegurar su calidad.

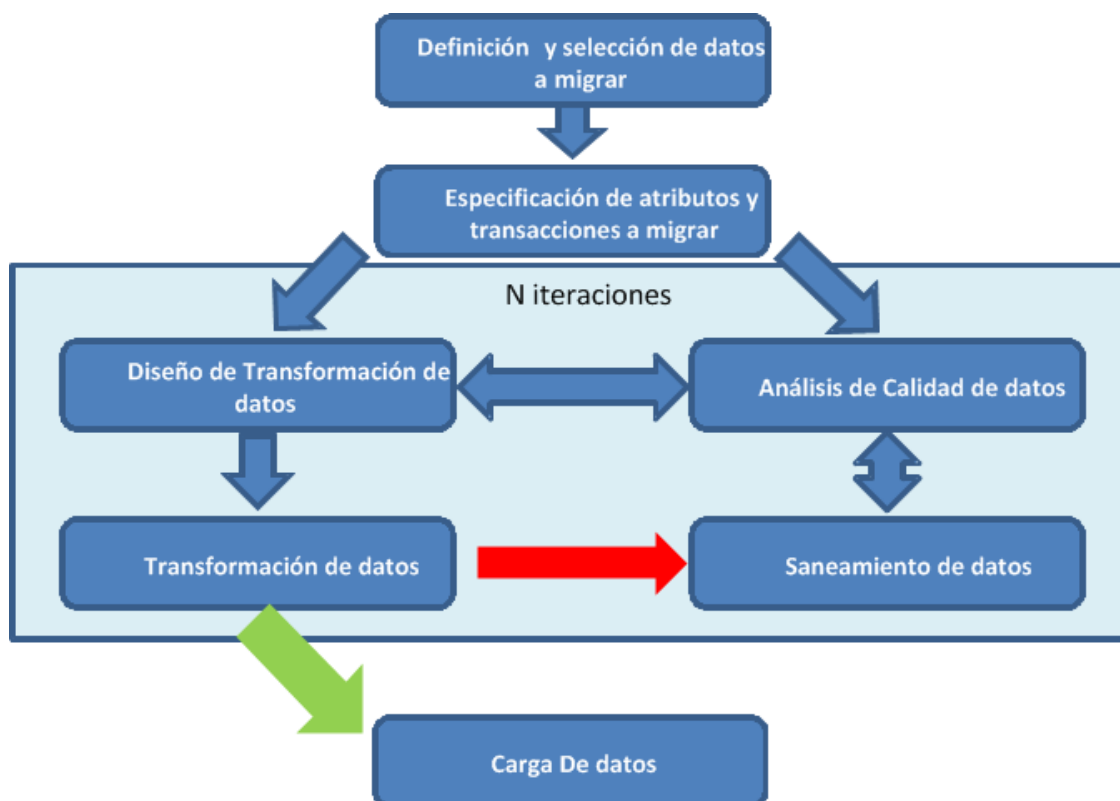


Fig. 4a: Proceso de saneamiento y carga de datos

La siguiente imagen ilustra cada uno de los componentes y su integración.

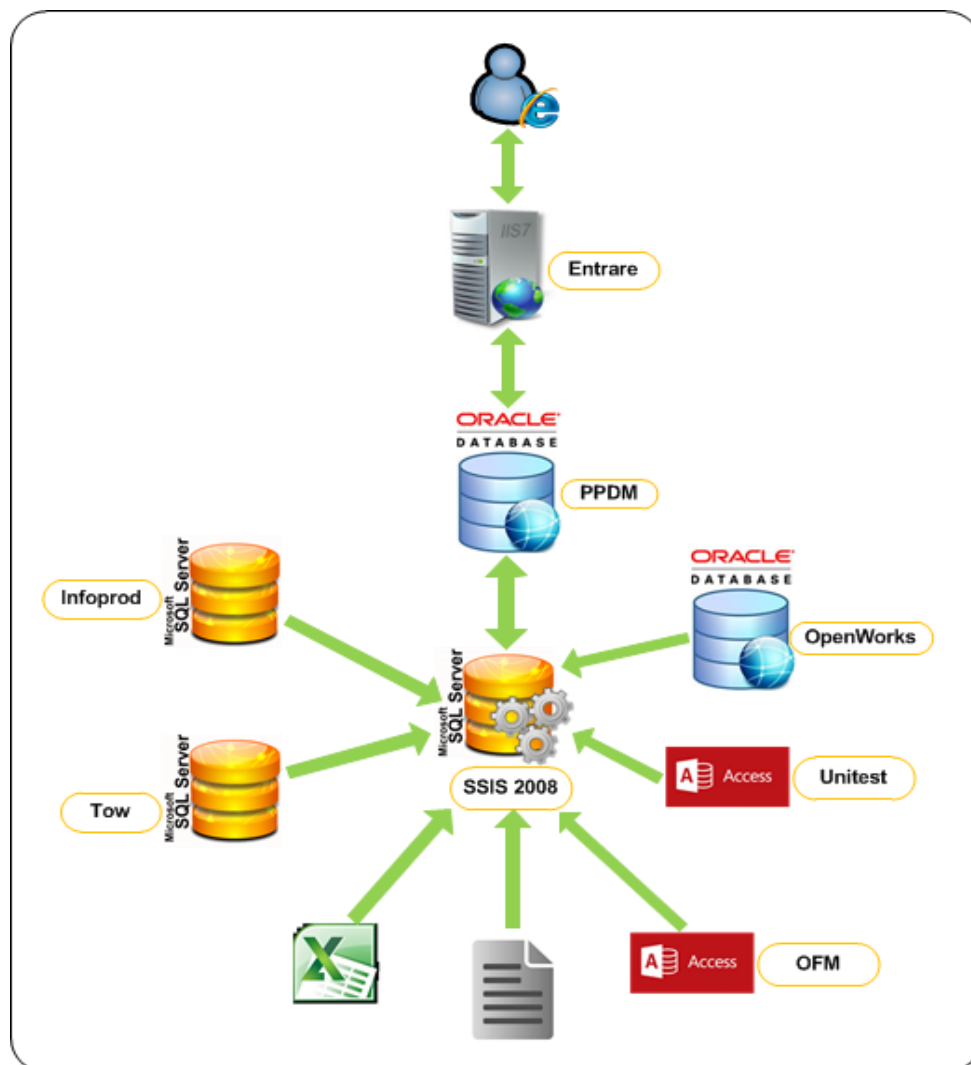


Fig. 4b: Esquema funcional de base de datos integrada y sus fuentes de datos

Con SSIS se implementaron las interfaces de migración de datos a PPDM desde los siguientes orígenes:

- SQL Server: TOW e Infoprod
- Oracle: Open Works
- Access: Unitest, OFM
- Excel: información de fracturas, PVT, Propiedades petrofísicas, Permeabilidades relativas, etc.
- Archivos de texto

Se desarrolló Entrare (Ensayos y Tratamientos de Reservorios) como la aplicación a ser utilizada como *front-end* de la base integrada. Esta aplicación tiene como objetivo principal ayudar a los usuarios a registrar de una manera sencilla y consistente los datos que originalmente eran registrados en planillas.

Entre las funcionalidades más destacadas de la aplicación se encuentran:

- Registro de fracturas y tratamientos: la información de las fracturas se registra en diferentes grupos de datos; se registran las relaciones entre fracturas y punzados, las propiedades y propantes definidos en el diseño de la fractura, como así también los del resultado. De esta manera se puede comparar lo planificado de lo realizado en campo. Otro grupo de datos se refiere a las propiedades petrofísicas que se encuentran en cada capa del pozo donde se realiza la fractura. También se registran las diferentes presiones y aditivos usados en la fractura. Por último se registran los fluidos inyectados en los pre-tratamientos.

- PLTs: se registra para cada punzado asociado el aporte en porcentaje de agua, oil y gas, utilizados luego para calcular la producción de gas.
- Propiedades petrofísicas: la aplicación permite registrar las propiedades petrofísicas de las formaciones y vincular estos datos con pozos, áreas y campos.
- *Well tests*: es el registro de los diferentes ensayos realizados sobre un pozo para poder determinar su potencial.

Técnicamente Entrare está desarrollada como una aplicación web utilizando el *Framework .NET* 4.0 de Microsoft, implementada en un servidor IIS 7.5, y consumida por los usuarios dentro de la intranet con Microsoft Explorer.

Desarrollo de modelos analíticos

Comprensión del Negocio:

Del análisis del problema surge la recomendación de particionar el modelado en dos etapas. Una primera etapa en la que se seleccionara el par POZO-CAPA a ser fracturado y una segunda etapa en la que se aportara la información de los tratamientos realizados a fin de mejorar la calidad de la fractura obtenida y por ende su rendimiento.

Una primera etapa, a la que llamamos MODELO PRE, en la cual se tiene en cuenta solamente la información conocida con antelación a la ejecución de los tratamientos que llevaron a la realización de la fractura hidráulica.

Esta información consta básicamente de las variables petrofísicas medidas y/o extraídas de los perfiles de los pozos luego de perforados.

La segunda etapa, a la que denominamos MODELO POST, incorpora las variables relacionadas al tratamiento realizado sobre la capa para realizar la fractura hidráulica en la capa seleccionada. El objetivo de esta segunda etapa es esencialmente aportar al diseño de una configuración óptima de los estímulos de fractura, dadas las variables petrofísicas y la puntuación PRE del par POZO-CAPA a ser tratado.

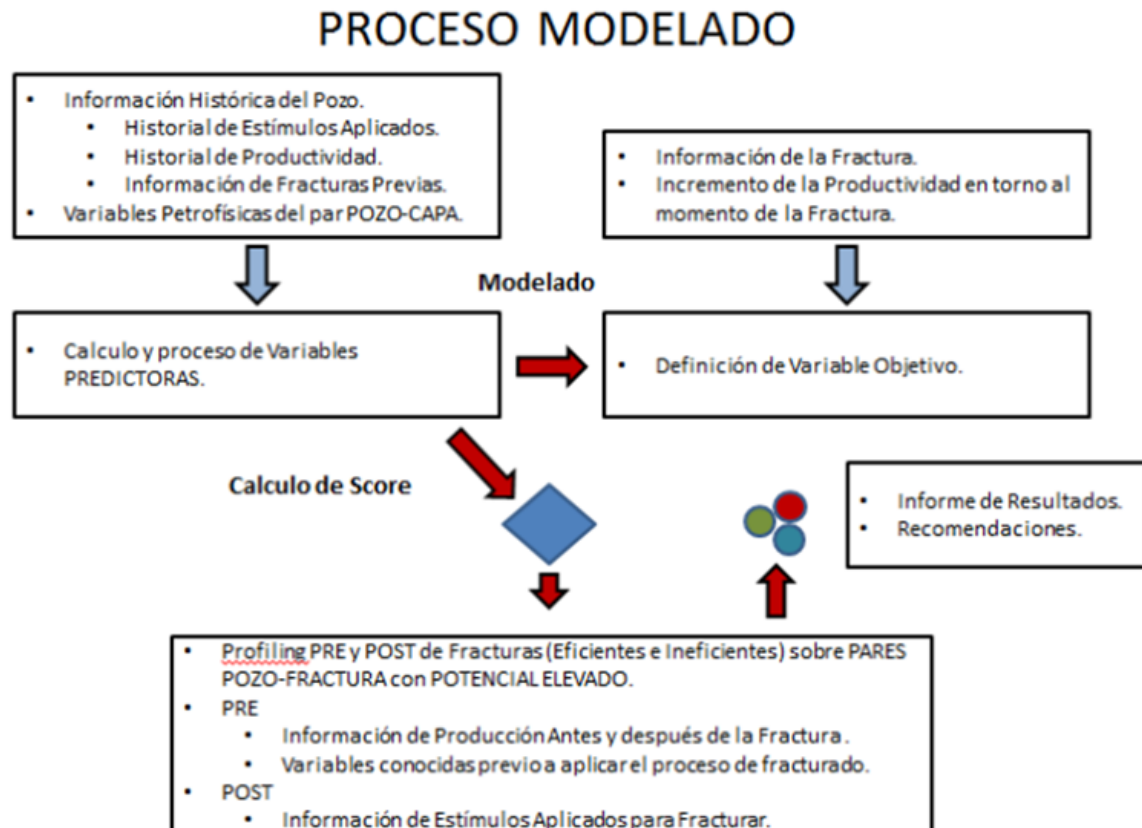


Fig. 5: Diagramado del proceso de Modelado, de los inputs, de sus etapas y resultados esperables.

En todos los casos el objetivo consiste en generar las mejores recomendaciones posibles de pares POZO-CAPA a ser fracturados que permitan maximizar el nivel de producción de la capa

(i.e. del pozo) luego de realizada la fractura, evitando y/o reduciendo las pérdidas económicas devenidas de una decisión errónea en la selección de la capa a ser fracturada en un pozo determinado.

Las fuentes de datos utilizadas contienen información de productividad antes y después del proceso de fracturado, variables de completación (datos de punzado), variables petrofísicas (de reservorios) e información de los estímulos aplicados (tipo de agentes, cantidad de bolsas, forma de la colocación en la formación, volúmenes inyectados, fluidos, presiones, etc.) para producir cada una de las fracturas en los pozos del reservorio Medanito.

En la figura 5 se muestra un esquema de cómo serán las etapas del proceso de modelado.

Comprensión de los Datos:

El proceso de comprensión de los datos se desarrolló con *Spotfire Analyst*. Con esta herramienta se caracterizaron las fuentes de datos en cuanto a su completitud, distribución, duplicidad de registros, concentración de valores y calidad de datos de manera general.

En la figura 6 se muestran capturas de pantalla de la herramienta en este proceso, conocido como *PROFILING* de la fuente de datos.

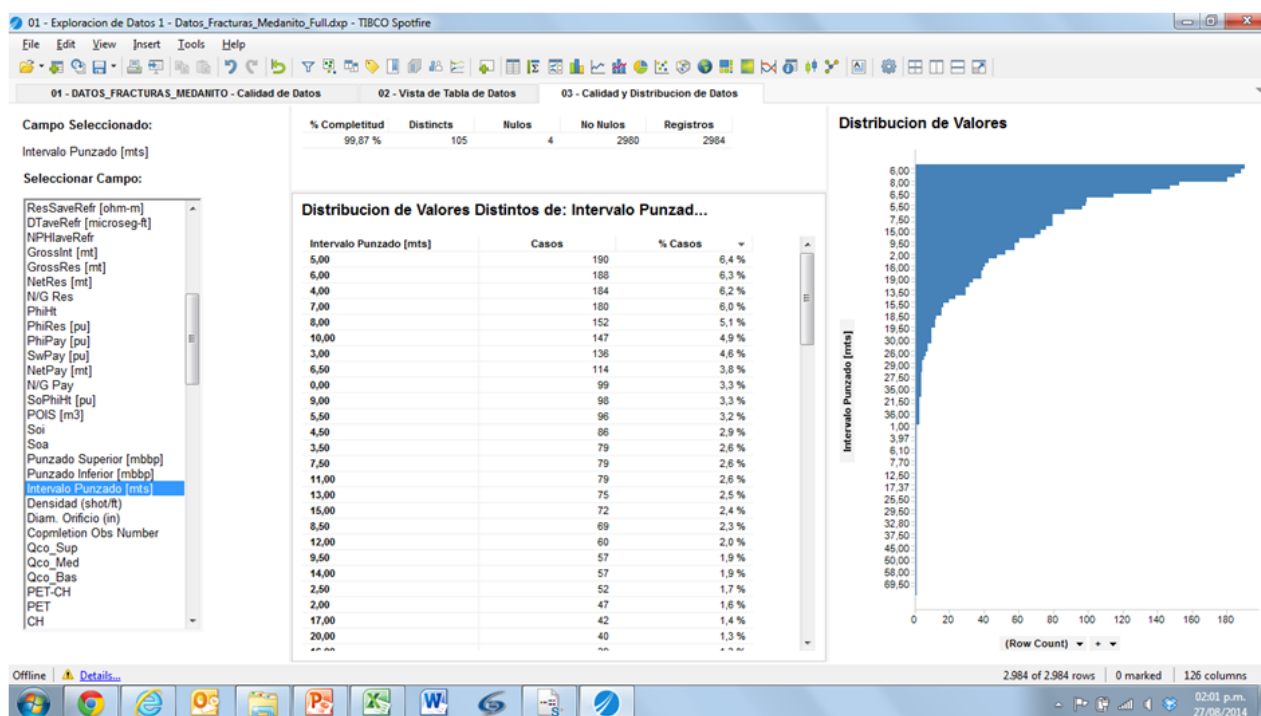


Fig. 6: *Spotfire Analyst* en el proceso de *PROFILING* (QA) de las fuentes de datos.

El resultado de esta primera instancia permite hacer una primera selección de variables, las que por su completitud y variabilidad puedan ser incorporadas en un análisis posterior para evaluar su poder predictivo.

Realizada esta pre selección de campos se procedió a la realización de un análisis exploratorio uní y multivariado a fin de caracterizar adecuadamente a cada una de las variables a ser empleadas.

A continuación, en la figura 7, se muestran capturas de pantalla de *Spotfire Analyst* utilizándose con este fin exploratorio.

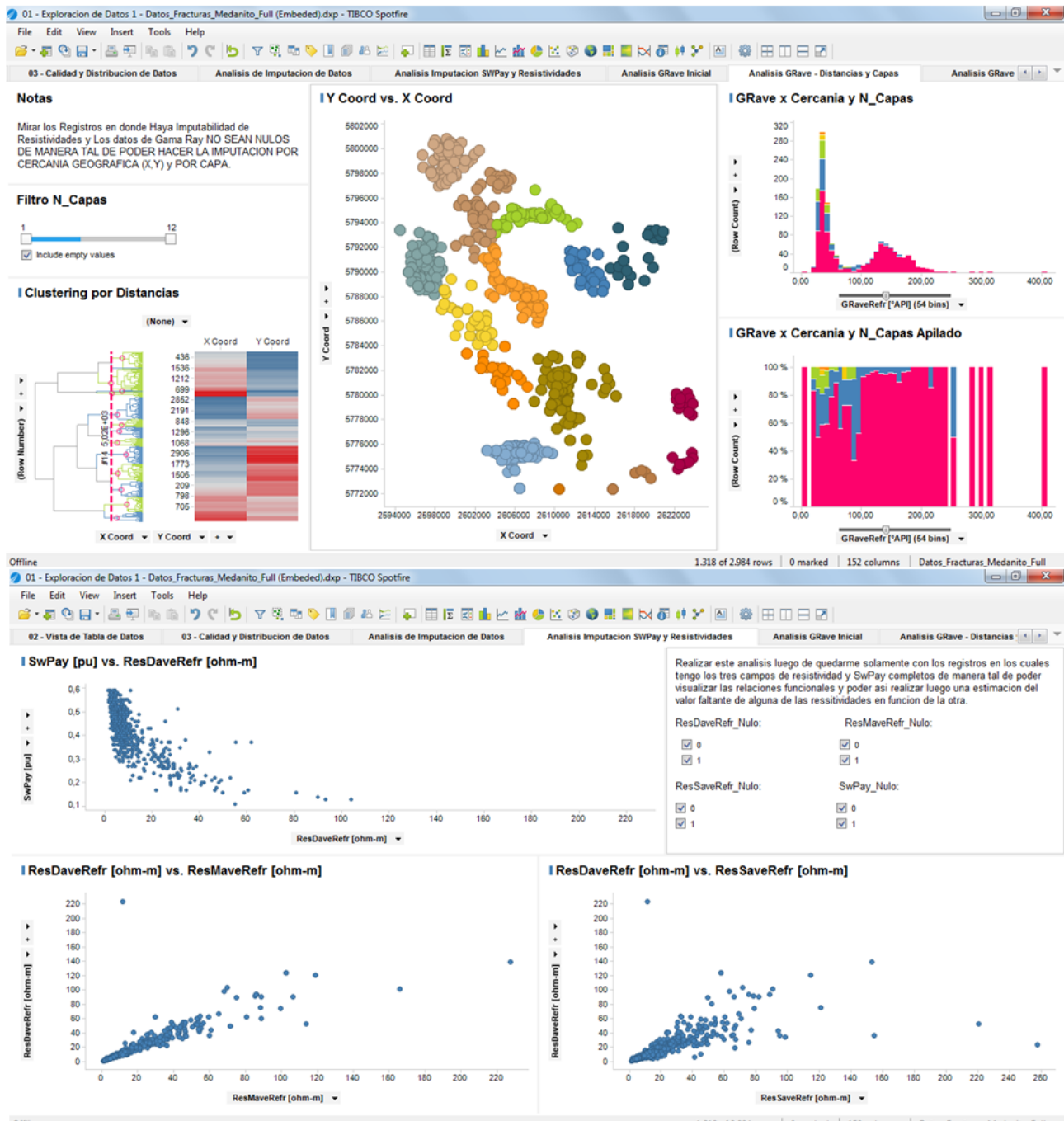


Fig. 7: Spotfire Analyst utilizado en el proceso de Exploración de Datos

En esta etapa exploratoria y de comprensión de los datos se definió también el OBJETIVO TÉCNICO del análisis. Este objetivo técnico consiste en la marca a ser utilizada como variable objetivo por los algoritmos de aprendizaje.

El mismo fue definido utilizando en *BOX Plot*, observando el cambio en la productividad diaria acumulada, calculada desde datos de la producción de la capa 6 meses antes y después del proceso de fractura.

En la figura 8 se muestra una captura de pantalla de *Spotfire Analyst*, en la cual se visualiza este proceso.

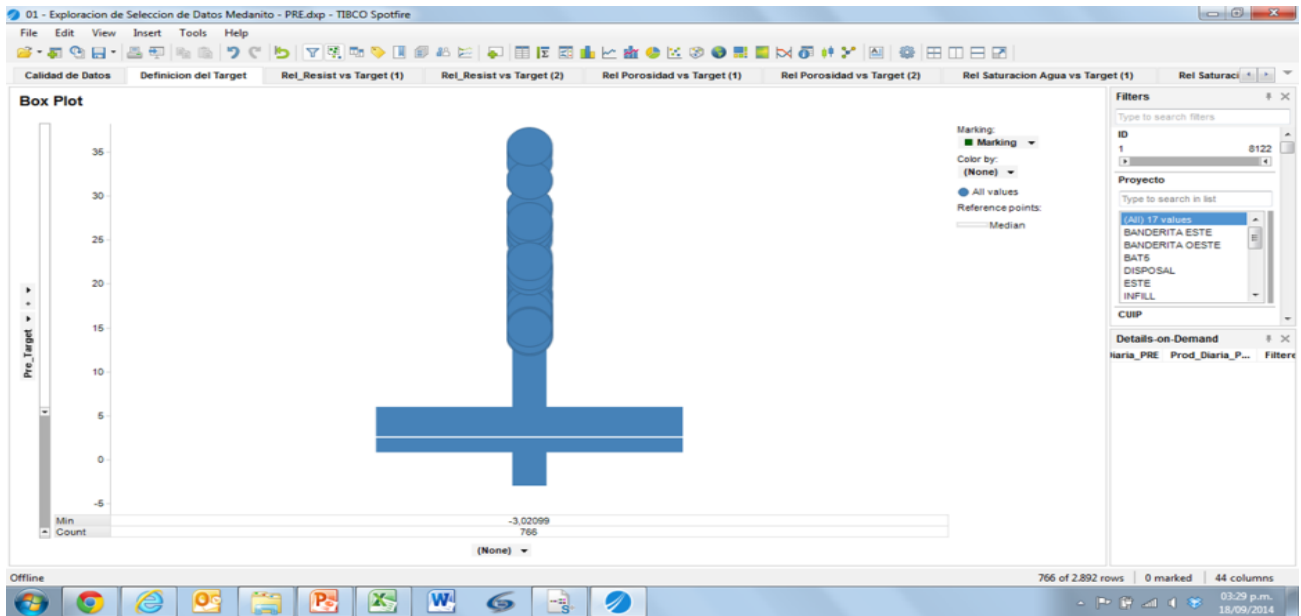


Fig. 8: Definición de Objetivo Técnico de análisis utilizando un Box Plot de Spotfire Analyst.

Como parte de este proceso exploratorio, se realizó en proceso de pre selección de variables teniendo en cuenta el aporte de poder de discriminación “APARENTE” de cada una de ellas en relación al objetivo técnico de análisis.

De este análisis se seleccionaron para el modelo PRE las siguientes variables:

- ResDaveRefr - Resistividad *Depth* o Profunda, promedio
- ResSaveRefr - Resistividad *Sallow*, superficial o poco profunda, promedio.
- Se calculó además, la diferencia entre estas dos resistividades.
- GraveRefr – Emisión de Rayos Gama promedio – Indicador de Litología.
- Porosidad Media
- SWPay – Saturación de Agua promedio.
- RHOBaveRefr – Densidad Promedio.
- Flg_Mas_de_Una_Capa –Indicador distingue entre los casos en que se fracturo una o más de una capa.

Y las siguientes se adicionaron a las anteriores para el Modelos POST:

- Intervalo_Punzado
- Tipo_Arena_Marca
- Arena_Bombeada
- Arena_en_Formacion
- Caudal_x_mt_Punzado
- Caudal
- Colchon
- HHP
- Vol_Iny_x_mt_Capa
- Vol_Iny
- Vol_TT_x_mt_Punzado
- Vol_Tratamiento
- Fluido_Fractura

En esta etapa se definieron además los procesos a realizarse sobre estas variables con la finalidad de lograr la expresividad y repetitividad posible existente en los patrones ocultos en los datos.

Preparación de Datos:

En esta instancia se realizó el procesamiento de las variables definido en la última etapa del paso anterior. Las variables se procesaron aplicándose procesos de *capping*, *ranking*, *centering* y *flagging* según resultado necesario de manera tal de lograr esta expresividad.

En la figura 9 se ven partes de este proceso, el cual fue implementado posteriormente e *Spotfire Miner*.

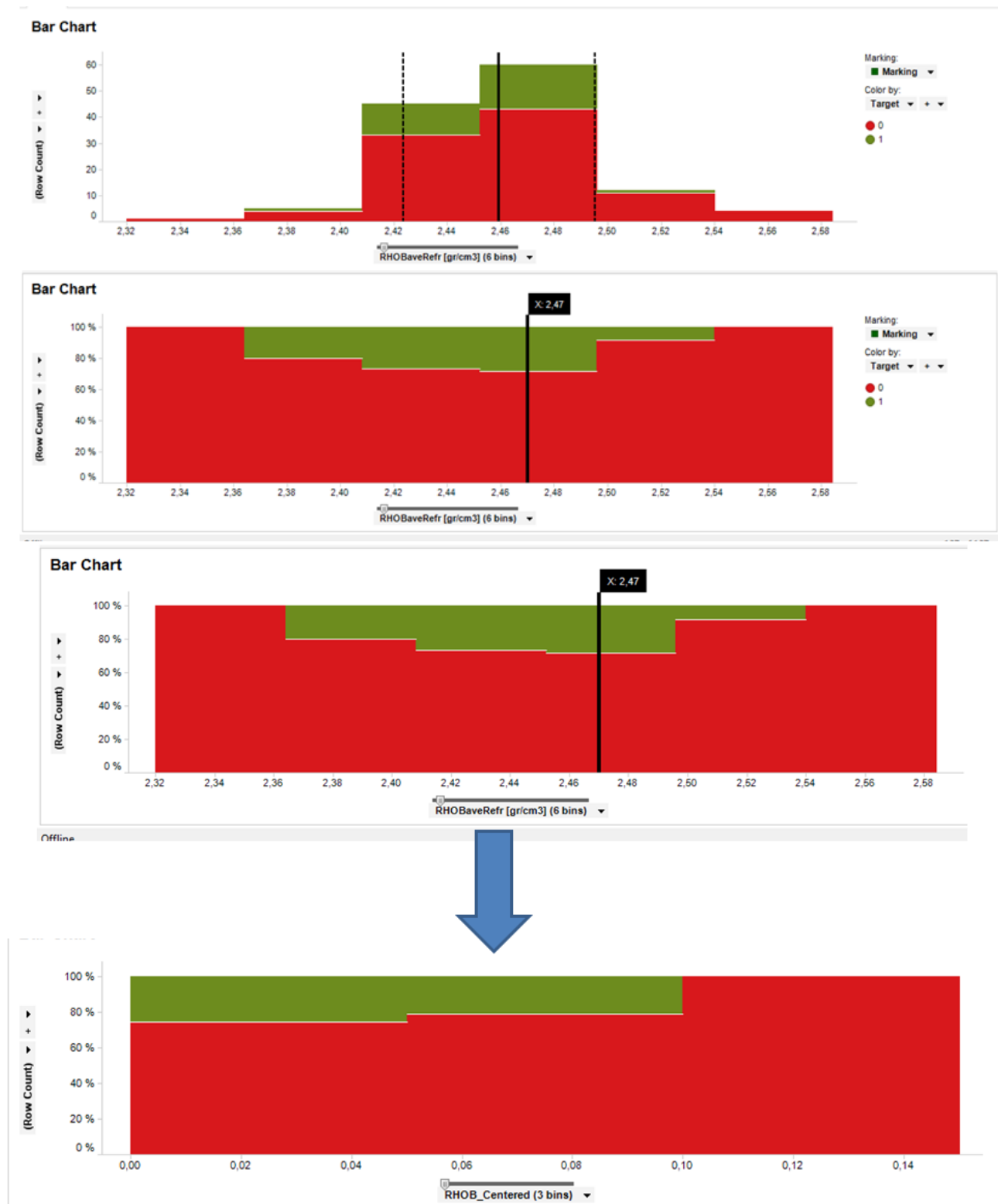


Fig.9: Procesamiento de variables

El paso posterior consistió en determinar las muestras sobre la que se desarrollaran los modelos y la muestra sobre la cual será testado el correcto aprendizaje de los modelos.

Las proporciones de las mismas fueron configuradas en 70% para entrenamiento y 30% para validación. Esta selección fue hecha a fin de contar con los registros suficientes en la muestra de entrenamiento para lograr el mejor ajuste posible de los modelos.

En la figura 10 se muestra una ruta de *Spotfire Miner* en la que se ajustan diversos modelos para luego comprarse y seleccionarse.

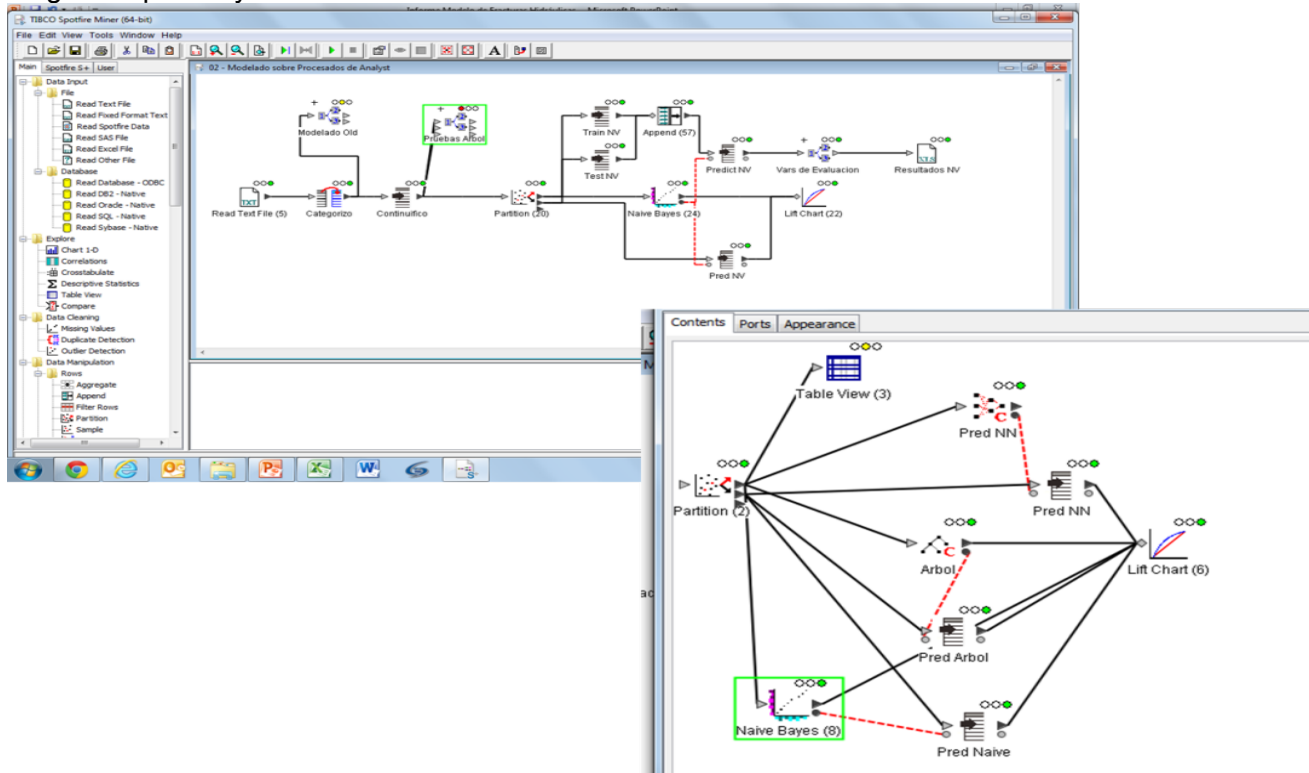


Fig. 10: Ejemplo de rutas de análisis en *Spotfire Miner* en la que se ven el procesamiento de datos, el ajuste de modelos y su posterior evaluación y comparación.

Modelado:

En esta etapa se desarrollaron los dos modelos antes mencionados (modelos PRE y POST respectivamente).

Se probaron distintas técnicas de modelado tales como Árboles de Clasificación, Naive Bayes y Redes Neuronales, dejándose de lado a las Regresiones Logísticas debido a que la cantidad de registros en las muestras no era suficiente como para garantizar un ajuste estable de este tipo de modelos.

En la figura 10 pueden verse distintas configuraciones de modelado.

Evaluación de Modelos:

Los distintos modelos fueron evaluados con los gráficos de Ganancias. En las figuras 11 y 12 se muestran estos gráficos para los modelos PRE y POST respectivamente.

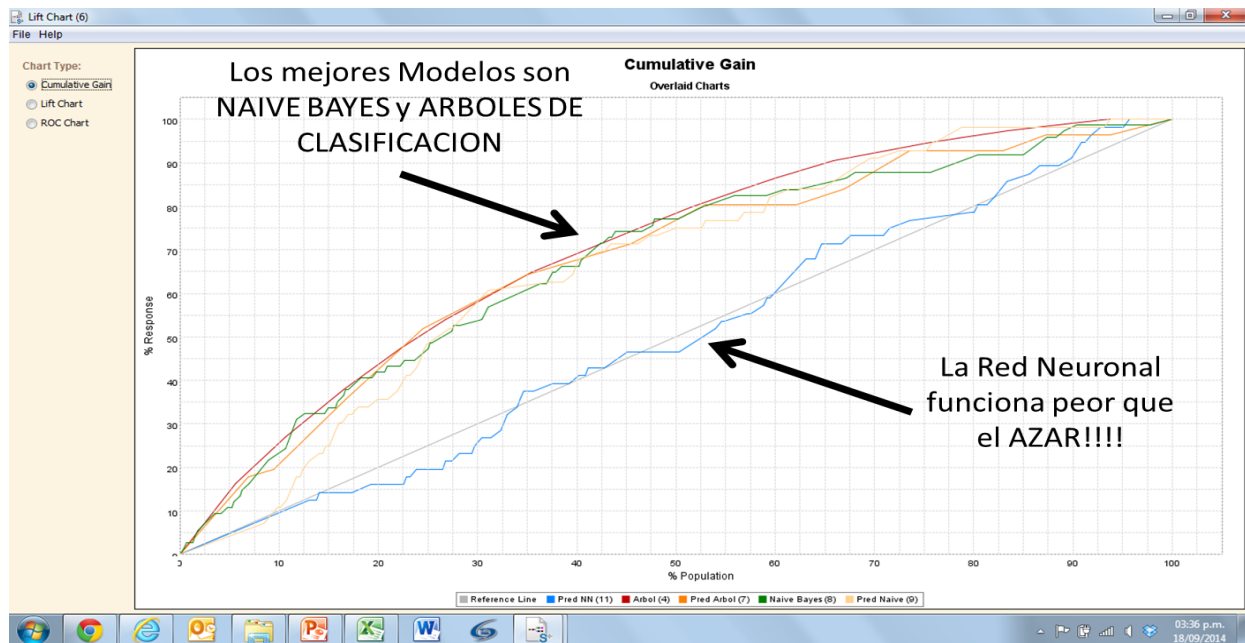


Fig. 11: Evaluación de Modelo Pre.

En la figura 11 se muestra por qué se descartaron a las redes neuronales. El modelo PRE finalmente seleccionado fue el NAIVE BAYES, debido a que por la cantidad de registros disponibles, los arboles de decisión no resultaban fuertemente estables.

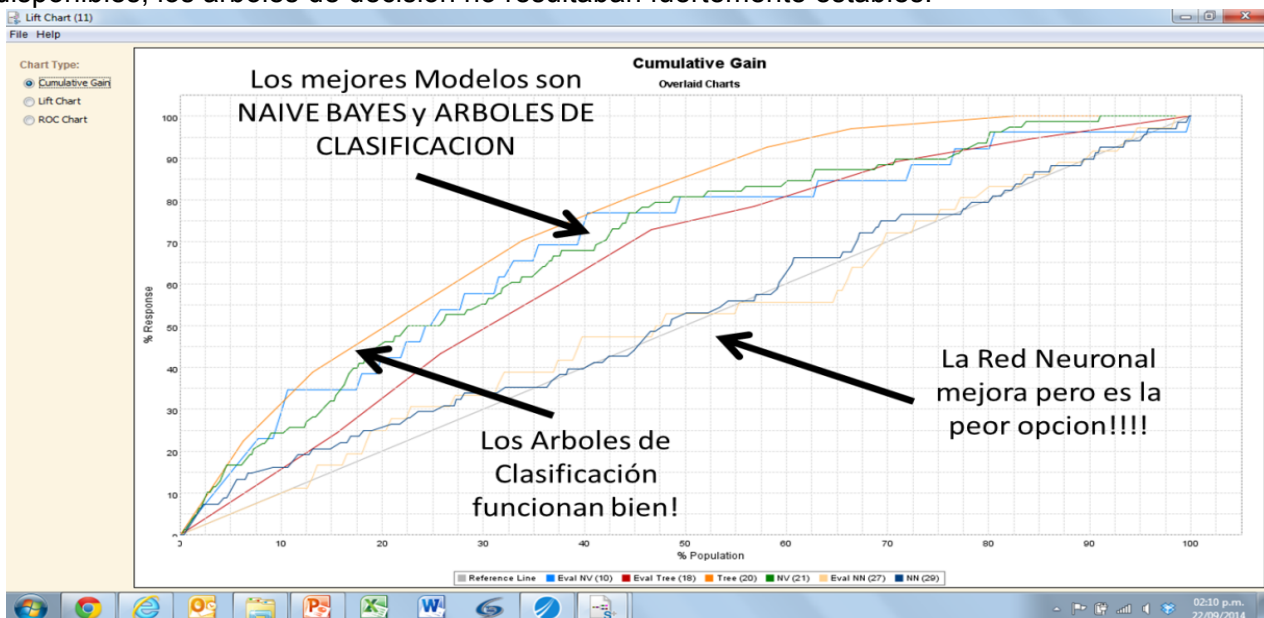


Fig.12: Evaluación de Modelos POST

En la figura 12, se visualiza la evaluación para los modelos POST. Nuevamente la cantidad de registros disponibles nos llevó a seleccionar a NAIVE Bayes como el modelo a ser implementado. En la figura 13 se muestra la curva de ganancias para el modelo POST con un par de conclusiones:

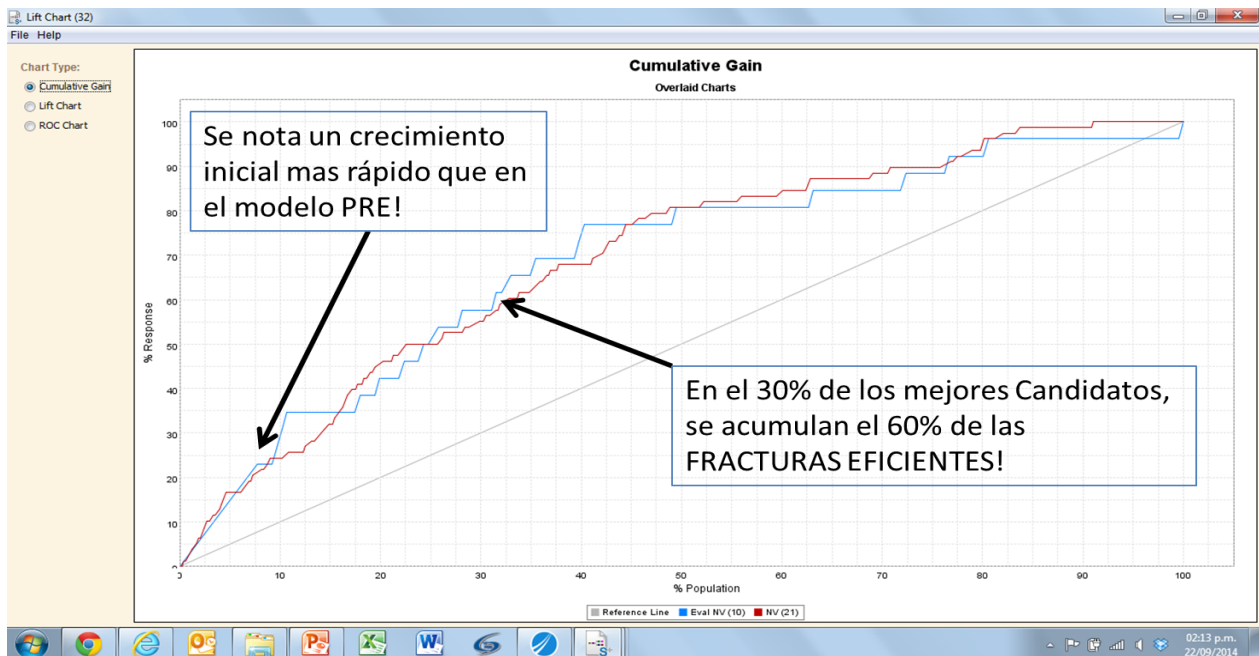


Fig. 13: Curva de Ganancias para Evaluación del Modelo POST seleccionado.

Implementación (Generación de Recomendaciones):

Una vez seleccionados los modelos analíticos a ser implementados, los mismos se incorporaron en una ruta de Spotfire Miner a fin de iniciar el proceso de generación de recomendaciones. En la figura 14 se muestra esta ruta impactando sobre un archivo Excel como salida.

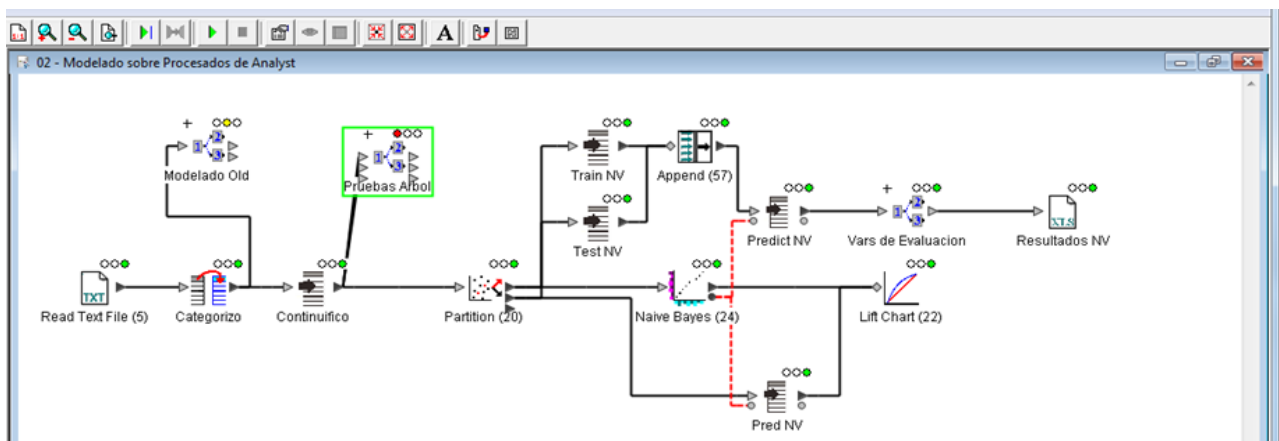


Fig. 14: Ruta en Spotfire Miner con la Implementación de los modelos para la generación de Recomendaciones.

Integración del producto desarrollado en el proceso de toma de decisiones.

La información proporcionada por el modelo será estudiada en detalle por los especialistas del negocio quienes evaluarán la factibilidad física y económica de continuar con las recomendaciones generadas en el campo.

Esta metodología cambia la forma de trabajo a la que están habituados los ingenieros del sector, agregando una nueva serie de indicadores desarrollada desde los datos y produciendo la necesidad de generar un cambio en el método de trabajo actual.

Es importante señalar que este modelo puede ser utilizado repetidamente por ingenieros sin conocimientos de *Data Mining* ya que solamente deben suministrar datos en la entrada y estudiar la salida aportando una manera sencilla y precisa para la toma de decisiones.

CONCLUSIONES

La primera conclusión en este trabajo es que el camino hacia la explotación de datos sostenible conlleva una gobernanza de datos mandatoria, así como un examen continuo de las soluciones que existen en el mercado tanto para predicción como para análisis.

El conocimiento detallado de los datos y una adecuada selección de variables acorde al problema a estudiar es otra condición necesaria para el éxito de los proyectos.

Alcanzar este objetivo implicó disponibilizar la infraestructura tecnológica necesaria, generar los procesos de trabajo y sobre todo trabajar en la capacitación y el cambio de la forma de trabajo de todos los intervinientes. Este trabajo es continuo y requiere de gran compromiso de la alta gerencia y sobre todo de equipos multidisciplinarios integrados en cada una de las etapas. Una de las claves en la conformación del equipo fue la inclusión de un especialista de analytics perteneciente a una empresa de servicios que permitió acelerar la curva de aprendizaje y el proceso de desarrollo en sí.

Establecer el objetivo del trabajo es fundamental para entender el problema y el tipo de herramienta a utilizar. Entender qué necesitamos resolver posibilita la solución.

De manera general las conclusiones obtenidas en el desarrollo del presente trabajo se encuadran en el hecho de que la implementación de técnicas de *analytics*, no solo son posibles en la industria del Petróleo y gas, sino que además aportan un gran valor al negocio, optimizando el proceso de toma de decisiones en la selección de pares POZO-CAPA a ser fracturados con el fin de maximizar la capacidad de producción acumulada de un pozo.

La calidad de dicha selección, basada en un *SCORE* obtenido por técnicas de aprendizaje automático, produce un gran valor económico para la empresa debido a que posibilita la reducción de costos de fracturas que no han producido los efectos deseados en la productividad del pozo.

De manera puntual podemos concluir que los mejores modelos han sido obtenidos con los algoritmos NAIVE Bayes y árboles de clasificación. Dejando finalmente fuera a estos últimos, y a las regresiones logísticas debido a que la cantidad de registros no fue la suficiente para obtener resultados estables en los modelos.

Por otro lado, pudimos ver como los modelos basados en redes neuronales para clasificación no funcionaron de manera adecuada, pudiendo inferir que quizá, este tipo de problemas no sean del tipo adecuado para ser tratados con dichas técnicas.

En cuanto a la evaluación de la plataforma analítica *Tibco Spotfire*, podemos concluir que ha resultado buena en cuanto a su flexibilidad ya que contiene todas las funcionalidades y algoritmos necesarios para el desarrollo y evaluación de modelos analíticos, alcanzando el grado de óptima al conjugar simpleza y velocidad en el desarrollo además de aportar información y conocimiento instantáneo en el proceso.

Podemos destacar que la preparación de datos es crítica para asegurar el éxito del proyecto y al que más tiempo recomendamos dar en la planificación de un proyecto de estas características.

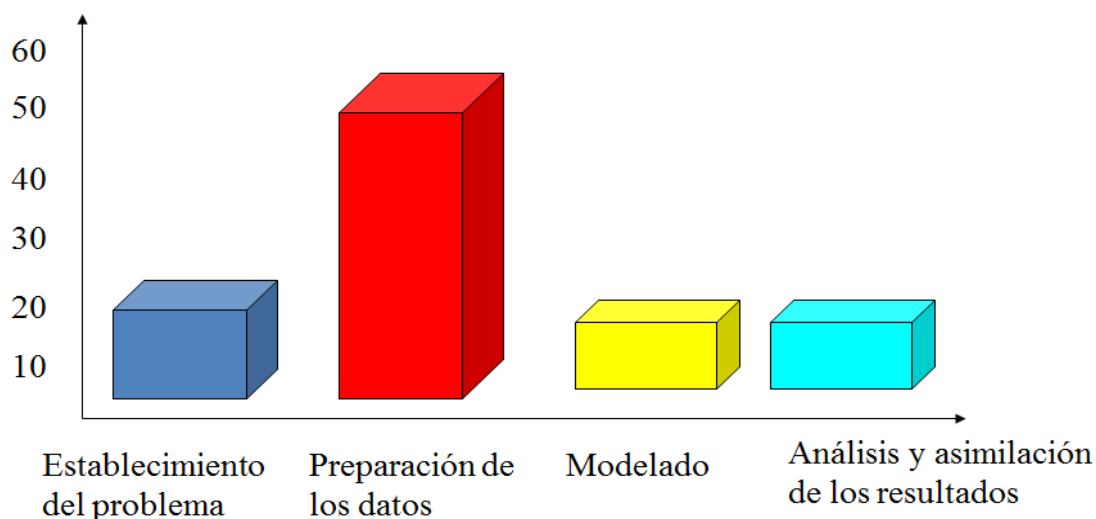


Fig.15. Distribución de tiempos de proyecto.

Como corolario de este trabajo se ofrece un conjunto de recomendaciones clave para lograr una implementación exitosa:

- Comenzar con un fin en mente (enfocar el problema).
- Seguir una metodología probada. Entender que Inteligencia Analítica es un proceso iterativo e interactivo.
- Adecuar las expectativas a las posibilidades reales de la Inteligencia Analítica.
- Delimitar claramente el problema y las métricas de evaluación.
- Participación de gente con conocimiento del negocio y el suficiente nivel de autoridad en cada etapa del proceso.
- No es posible hacer inteligencia analítica solo con consultores externos.
- Disponer de los datos en tiempo y forma, con soporte de todas las áreas involucradas en su registro, almacenamiento, extracción y manejo.
- Que el usuario pueda interpretar los resultados (conocimiento, indicador, pronóstico, etc.)

Bibliografía:

Estanislao Irigoyen, 2014, Introducción a la visualización de Información y *Analytics*, Jornadas de Inteligencia de Negocios, IAE, Instituto Argentino de la Empresa, Ciudad Autónoma de Buenos Aires.

PPDM *Standarts*, 2013 , PPDM 3.8 Model , *The Professional Petroleum Data Management* (PPDM) Association , <http://www.ppdm.org>

Víctor Paci, Eduardo Zenón García, 2013, Selección de Candidatos con Inteligencia Artificial, Petrobras Argentina.

Tim Crocker, 2014, *Spotfire Oil & Gas Production Optimization*, *TIBCO Spotfire Conference* , USA.