

# SISTEMA INTELIGENTE PARA ANÁLISIS Y MONITOREO DE PROCESOS DE CONTRATACIÓN DE ENTIDADES PÚBLICAS EN COLOMBIA

María Barbarita Camargo García<sup>1</sup>, Yesica Paola Pérez Velazco<sup>2</sup>

Universidad De Los Llanos

Programa De Ingeniería De Sistemas

Villavicencio, Meta, Colombia

<sup>1</sup>, [yesica.perez@unillanos.edu.co](mailto:yesica.perez@unillanos.edu.co)<sup>2</sup>

## Resumen:

Como resultado, se logró obtener un modelo experimental de minería de datos aplicando la metodología espiral, donde se describe el proceso diseñado para el análisis, extracción, limpieza y modelado de la información obtenida del portal de contratación [ HYPERLINK \l "Col" 1 ]. Se usaron técnicas para la extracción de datos, se evaluaron modelos de reconocimiento, de identificación, caracterización y un mecanismo de detección de falsos positivos/negativos. Se creó un modelo de entrenamiento a partir de datos recolectados y proporcionados por el portal datos.gov 2]]; estos datos son procesos de años anteriores, los cuales cumplían con características similares a los obtenidos en la extracción de datos del portal.

Following the relevant regulations of the law 1266 and in compliance with article 7, which deals with handling and data integrity, the development of this project is based on the implementation of a system of data analysis, which allows the user easy access recruitment processes at national level and be aware of bids that may be of interest.

Making use of data mining, applying certain techniques and according to search criteria that the user must define the platform, we could identify which are the most important processes for him

It was a key factor in the development of this project, select a data mining technique to throw an accurate and consistent results with information; therefore different predictive and descriptive techniques were applied, to which was subjected to tests to ensure that throw obtain a coherent and consistent results with the aim of our system.

**Keywords:** *Processes, recruitment, portal, analysis, extraction, cleansing, modeling, techniques, integrity, criteria, search, bidding, data, prediction, mining tests.*

## 1. Introducción.

La minería de datos es una de las áreas donde encontramos técnicas que facilitan el manejo y uso de la información; es por ello que cuando se habla de generar modelos a partir de grandes cantidades de datos, la minería es nuestra mejor opción. En Colombia el proceso de contratación a nivel nacional, se desarrolla mediante el portal de contratación pública, en donde existen muchos procesos de contratación que a su vez tienen muchos estados, fases, clasificaciones; los cuales en la mayoría de los casos, resultan tener cierto grado de complejidad. Estos procesos se ofertan a diario en entidades del estado, quienes las publican en algunos portales o sitios web [ [HYPERLINK \l "Gob2" 1 2](#)] como el que ya se ha mencionado. En muchas ocasiones las personas o empresas no son notificados a tiempo y pierden la oportunidad de postularse a cualquiera de esos procesos que se encuentren en el momento abiertos.

## 2. Metodología.

A continuación, por medio de los cuatro pasos principales (Determinar Objetivos, Análisis de riesgo, Desarrollar, verificar y planificar) [ [HYPERLINK \l "RGa11" 4](#) ] de la metodología espiral, se describe todo el trabajo realizado.

### a) Determinar o fijar los objetivos

Como primera fase de desarrollo de la metodología, se plantearon los siguientes objetivos del proyecto:

#### Objetivo General

Diseñar y desarrollar un sistema inteligente para el monitoreo de procesos de contratación a través de la información pública proporcionada por el portal de [contratos.gov.co](http://contratos.gov.co), que permita identificar de manera automatizada procesos de interés dados por empresas, personas o entidades que estén interesados en concursar en los procesos de contratación.

#### Objetivos Específicos

- Establecer las técnicas de recolección de datos proporcionados por el portal [contratos.gov.co](http://contratos.gov.co), de tal forma que a través de ellos se permita la identificación de procesos de contratación a tiempo.
- Diseñar un modelo que permita el ingreso adecuado de acuerdo a la estructura de una base de datos propia, que contenga la información acerca de los procesos de contratación recolectados, para llevar a cabo un proceso de minería de datos sobre los mismos.
- Desarrollar una aplicación móvil que le permita a una persona o entidad a partir de unos criterios, obtener alarmas inteligentes acerca de procesos de interés.
- Definir un mecanismo de pruebas que permita establecer las tasas de efectividad respecto a falsos positivos / negativos.
- Identificar errores en la aplicación, a través de un sistema de pruebas para determinar el nivel de confianza que tendrán los distintos componentes de la aplicación, para saber si juntos funcionan de la manera correcta y si el sistema cumple con las necesidades de funcionamiento o si el contenido

del sistema es coherente y consistente, etc.

#### b) Análisis de riesgo.

Se evaluaron los riesgos que podía tener el proyecto y se presentó como riesgo principal no poder recolectar todos los datos que pudieran cumplir con lo requerido. Durante el desarrollo no sucedió esto, realmente se logró recolectar los datos necesarios y almacenarlos en nuestra base de datos. Se determinaron las metas de la minería de datos, que tenían como objetivo la elaboración de un sistema mediante las técnicas de minería, que lograra identificar los procesos con mayor probabilidad de terminarse obteniendo resultados de falsos positivos/negativos, permitiendo la búsqueda de procesos que sean sólo del interés del usuario.

El levantamiento de requerimientos se realizó con base a los objetivos del proyecto y a la facilidad que debía tener para que el usuario ingresara al sistema sin ningún inconveniente. De esta forma, se logró obtener una interfaz sencilla y de fácil manejo.

#### c) Desarrollar, verificar y validar

Para el desarrollo del proyecto se escogieron como herramientas:

- Python 2.7

Python es un lenguaje ideal para el desarrollo rápido de aplicaciones en diversas áreas y sobre la mayoría de las plataformas. En este caso, los scripts que se usaron para el funcionamiento de la aplicación fueron elaborados en este lenguaje5}}.

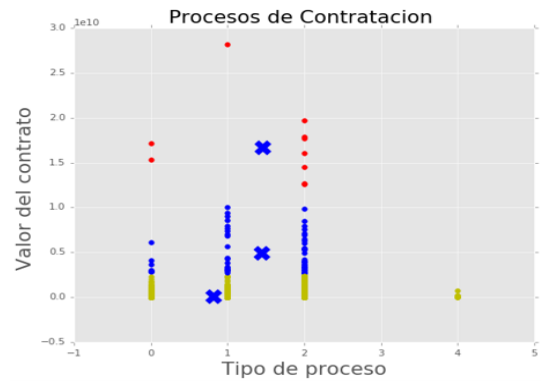


Figura 1. Agrupamiento de procesos de contratación de acuerdo al valor del contrato y tipo de proceso al que pertenece, organizado en 3 grupos.

Por la linealidad y desfase de los centroides respecto a los contratos, se descartó esta técnica, porque no se evidenciaba de manera clara, a qué grupos pertenecían los contratos.

- **Naive Bayes:** Se ubica entre las técnicas de aprendizaje supervisado y predictivo; generalmente se aplica a modelos en los cuales se quiere predecir una probabilidad como resultado [ [HYPERLINK \l "Sha16" 12](#) ]. En Naive Bayes, tuvimos que usar dos conjuntos de datos, uno para entrenamiento y uno para pruebas. Los datos del 2012 nos sirvieron como datos de entrenamiento y los datos del scraping como prueba para la clasificación; adicionalmente se hizo otra prueba utilizando contratos del 2015. En esta parte solo usamos tres variables, cuantía, proceso y la clase que es donde está el grupo al que pertenece. Los datos de entrenamiento fueron 198.872 del año 2012 y 6000 datos de prueba del año 2016. El resultado que nos arrojó fue el siguiente.

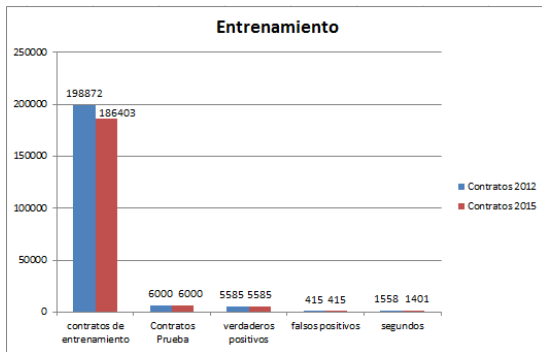


Figura 2. Comparación de entrenamiento en Naive Bayes, utilizando procesos de contratación de 2012 y 2015, utilizando como contratos de prueba, procesos del 2016, obteniendo falsos positivos y negativos.

La probabilidad aposteriori era la que nos decía si el proceso era de clase 0 o 1 y en este caso arrojó que todos eran 0, incluyendo los que eran 1; por ello, evidenciándose claramente la inexactitud de los resultados, se descartó, debido a que no acertaba a la clasificación de que un contrato se había terminado.

- **Random Forest:** Es una técnica de aprendizaje automático que está basado en los conocidos Árboles de decisión. Pertenece al conjunto de los algoritmos supervisados y predictivos. Una de sus principales características es que tiene un conjunto de clasificadores muy certeros. Random Forest Puede manejar múltiples variables y es conocido por tener un excelente rendimiento con bases de datos grandes [13]. Esta técnica fue la que se escogió debido a que arrojó mejores resultados en comparación con las anteriores. Este algoritmo toma muestras aleatorias del conjunto de datos de entrenamiento, lo que genera cientos de árboles en cada muestra, y se escoge entre las variables dadas cual es la más acertada y así nos da la predicción

de un resultado más certero [HYPERLINK \l "Dis" 13].

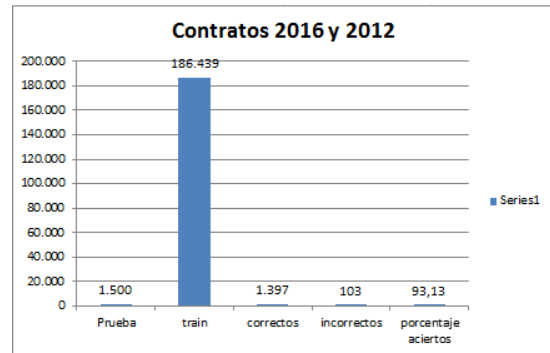


Figura 3. Entrenamiento de clasificación utilizando procesos de 2012 como entrenamiento y procesos de 2016 como prueba, evidenciándose clasificaciones correctas e incorrectas, como también el porcentaje de aciertos del resultado de la clasificación.

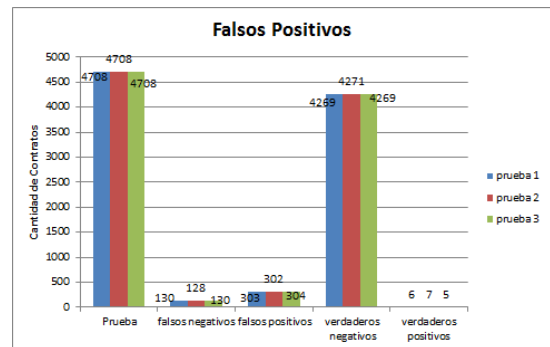


Figura 4. Pruebas de clasificación de procesos de contratación del año 2016, obteniendo falsos positivos y negativos.

## Diseño de la aplicación

Luego de haber modelado los datos, seleccionado la técnica, aplicarla a los datos de prueba y a los datos traídos del Scraping, verificar que efectivamente era más certera y más adecuada para el desarrollo; se empezó con la etapa de elaboración de la aplicación donde se crearon todas las interfaces y donde se logra que el script se ejecute en determinados tiempos, para que haga el proceso de clasificación internamente y logre notificar por medio de correo a los usuarios de la aplicación, que hay procesos nuevos, los cuales pueden ser

de su interés y que deben revisar su cuenta en la sección de preferencias.

- El usuario debe registrarse y se le presenta una interfaz donde podrá realizar dicho procedimiento, donde deberá ingresar Nombre, Apellido, Dirección de email y contraseña.
- Si el usuario ya tiene una cuenta creada podrá iniciar sesión por medio del correo y la contraseña antes creada.
- Si el usuario olvida su contraseña el sistema muestra una interfaz donde podrá generar una nueva por medio del correo para que ingrese al sistema.
- Después de que el usuario ingresa al sistema, este le muestra 20 procesos aleatorios, si le interesa alguno tendrá la opción de escogerlo y guardarlo, para que el sistema haga la clasificación y poder predecir si el contrato tiene o no probabilidad de cerrarse. Para ver el resultado el usuario debe esperar un correo que se envía en segundos, avisándole que puede revisar la sección de preferencias donde se muestran todos los resultados.
- El sistema ofrece al usuario un filtro de contratos con parámetros que sean únicamente de su interés. Estos parámetros se guardarán y con base a estos, cuando el usuario no esté dentro del sistema, se ejecutará el script y se le dará aviso de los procesos nuevos que puedan ser de su interés.

**Figura 4. Interfaz donde se eligen los contratos de acuerdos a unos criterios específicos, como lo son tipo de proceso, valor del contrato y el objeto del contrato que se refiere a de que se trata el mismo.**

Luego de seleccionar los parámetros el usuario debe esperar unos segundos a que la página muestre los contratos que se obtuvieron.

Entidad	Tipo de Proceso	Cuántía	Estado del Proceso	Objeto del Proceso
EJÉRCITO NACIONAL	Contratación Mínima Cuantía	\$5500000	Convocado	MÍNIMA CUANTÍA NO. 067.- BASPC-2016 CUYO OBJETO ES ADQUISICIÓN DE MOBILIARIO Y ENSERES PARA EL GAULA GUARISE Y COMANDO EN CONJUNTO, UNIDAD CENTRALIZADA POR EL BATALLÓN DE ASPC No. 2 "CAJIQUE ALFONSO VELAZQUEZ" PARA LA VIGENCIA 2016

**Figura 5. Interfaz de filtro de contratos, en donde se muestra el resultado de la búsqueda de acuerdo a los criterios seleccionados.**

De esta manera si el usuario desea podrá guardar los contratos consultados, para que el sistema realice búsquedas con los parámetros establecidos cuando él no se encuentre dentro de la aplicación.

Por último, el usuario tiene la opción de cerrar sesión en el momento que lo desee.

#### d) Planificar

En este último paso, se revisó la aplicación terminada y se ejecutaron las pruebas.

#### Pruebas

Después de haber generado todo el código fuente, la aplicación debió ser probada para descubrir el máximo de errores posibles antes de la entrega final. De acuerdo a esto se diseñaron unos casos de prueba suponiendo distintos usuarios:

- Se crearon tres cuentas distintas y se realizaron búsquedas. Por cuestión de espacio se muestran sólo las pruebas con uno de los usuarios.
- Se verificó que no hubieran datos con criterios guardados, es decir, que la colección donde se guardan los contratos del usuario estuviera en 0.
- Se escogieron dos procesos y se guardaron; antes de dar clic se verificó que en la sección de preferencias no hubiera ningún resultado.
- Luego de que el script se ejecutara y realizara la clasificación, se verificó que llegara un correo al usuario que en ese momento estaba consultando:



Figura 6. Vista de la bandeja de entrada del correo electrónico jessikapull@hotmail.com

```
total: 4
correctos: 3
porcentaje de aciertos: 75.0%
falsos negativos: 1
falsos positivos: 0
verdaderos negativos: 0
verdaderos positivos: 3
sensibilidad: 0
especificidad: 0
aciertos: 0.75
precisión: 0.98
predicción de valor negativo: 0
predicción de valor positivo: 100.0
coeficiente de correlación de Matthews: 0
jessikapull@hotmail.com
To: jessikapull@hotmail.com
From: jessikapull
Subject: Procesos de contratación
Correo enviado!
```

Figura 7. Resultado arrojado desde la ventana de ejecución del servidor del sistema, luego de haber hecho la clasificación con Random Forest.

- Se revisó que en la sección de preferencia estuviera el resultado de la búsqueda, lo cual efectivamente se encontraba con el resultado si era o no recomendable.



Figura 8. Interfaz de notificaciones en donde presenta el resultado de la clasificación de los procesos elegidos por el usuario.

De esta manera se cumplen todas las fases de la metodología y se entrega una aplicación funcional desde un servidor localhost.

### 3. Resultados

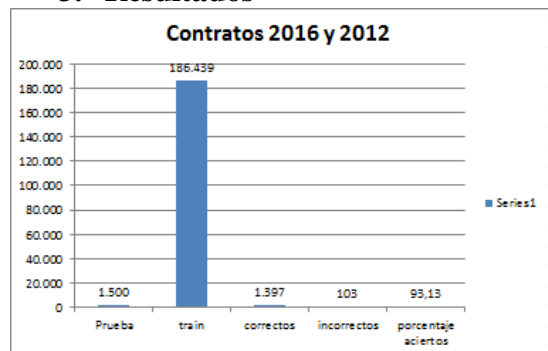
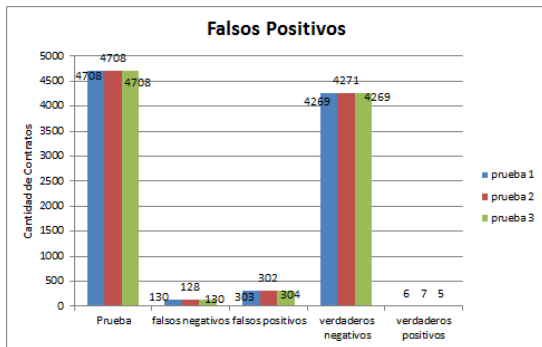
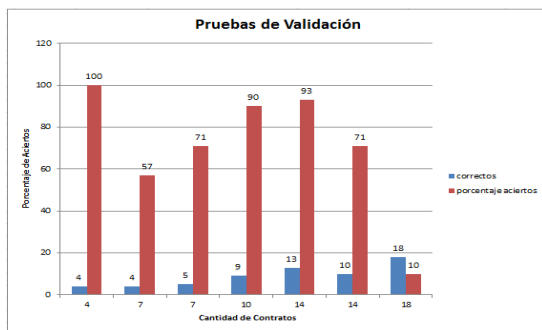


Figura 9. Entrenamiento de clasificación utilizando procesos de 2012 como entrenamiento y procesos de 2016 como prueba,

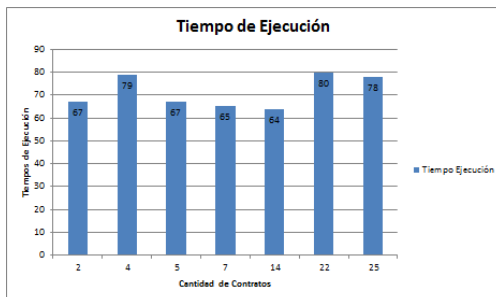
evidenciándose clasificaciones correctas e incorrectas, como también el porcentaje de aciertos del resultado de la clasificación, utilizando Random Forest.



**Figura 10. Pruebas de clasificación de procesos de contratación del año 2016, obteniendo falsos positivos y negativos, utilizando Random Forest.**



**Figura 11. Resultados de clasificación, poniendo a prueba el sistema.**



**Figura 12. Tiempos de ejecución, de acuerdo a una cantidad específica de contratos a clasificar.**

Random Forest, fue la técnica de clasificación más acertada. Se hicieron diferentes pruebas y aunque también había momentos en que se equivocaba, el porcentaje de error era muy bajo y acertaba en la clasificación de la mayoría de los procesos.

Se obtuvo la creación de un modelo y sistema de aprendizaje automático, a

partir de la aplicación de unas técnicas de minería de datos y el clasificador Random Forest.

Se desarrolló una aplicación de fácil uso y acceso para los usuarios, que sirve de apoyo en la toma de decisiones, referente a las empresas en donde aspiren adquirir determinados contratos.

En el transcurso del desarrollo de la aplicación, se creó un algoritmo a partir de técnicas predictivas y descriptivas para la clasificación de contratos, que pudieran ser de interés para el usuario a partir de unos criterios. Los resultados de las técnicas predictivas, se validaron de acuerdo a un conjunto de fórmulas utilizadas para darle certeza a los porcentajes obtenidos.

#### 4. Conclusiones

La minería de datos, es una herramienta estratégica para predecir o describir comportamientos en grandes cantidades de datos, ésta, fue clave en todo el proceso de la creación de la aplicación.

Durante el proceso, se evidenció, cómo a partir de unos datos y aplicando las técnicas de minería correctas y elaborando un modelo de predicción, se puede llegar a resultados de gran utilidad, de tal manera que los usuarios puedan acceder a los procesos de contratación que se estén ofertando en el momento.

Durante el desarrollo, en muchos casos los resultados no eran 100% exactos, ya que el método o la técnica que se usó, tenía su porcentaje de error como todo mecanismo de predicción. Sin embargo, debido a la cantidad de procesos, se pudo obtener un porcentaje considerable de



confiabilidad, y estadísticas que nos permitieron evidenciar el nivel de confianza de los resultados, que arrojaba el sistema de clasificación. Fueron abordados y alcanzados cada uno de los objetivos planteados desde un inicio. Se evaluaron las distintas técnicas de minería de datos, para elaborar un modelo de predicción acertado y se entrega un sistema de clasificación para los distintos procesos de contratación a nivel nacional.

## Referencias

- [1] Gobierno. colombiacompra.gov.co. [Online].  
HYPERLINK  
"http://www.colombiacompra.gov.co/sites/default/files/capacitacion/I-inicial.pdf"  
<http://www.colombiacompra.gov.co/sites/default/files/capacitacion/I-inicial.pdf>
- [2] Gobierno de Colombia. (2015, Octubre) Catálogo de Datos Abiertos. [Online]. HYPERLINK  
"http://www.datos.gov.co/frm/Acerca/frmAcercaDe.aspx"  
<http://www.datos.gov.co/frm/Acerca/frmAcercaDe.aspx>
- [3] José Miguel Rodríguez Parrilla, *Cómo hacer inteligente su negocio: business intelligence a su alcance.*: Larousse- Grupo Editorial Patria, 2014.
- [4] Galo Fariño R. (2011) ojovisual.net. [Online].  
HYPERLINK  
"http://www.ojovisual.net/galofarino/modeloespiral.pdf"  
<http://www.ojovisual.net/galofarino/modeloespiral.pdf>
- [5] Python. python.org. [Online]. HYPERLINK  
"https://www.python.org/doc/"  
<https://www.python.org/doc/>
- [6] MongoDB. mongodb.com. [Online].  
HYPERLINK "https://www.mongodb.com/es"  
<https://www.mongodb.com/es>
- [7] flask.pocoo.org. flask.pocoo.org. [Online].  
HYPERLINK "http://flask.pocoo.org"  
<http://flask.pocoo.org>
- [8] Julián Terán Anciano, *Manual de Introducción al lenguaje HTML. Formación para el Empleo.*: Editorial CEP, S.L, 2010.
- [9] JQuery.com. jquery.com. [Online]. HYPERLINK  
"https://jquery.com/" <https://jquery.com/>
- [10] Ryan Mitchell. (2015) Web Scraping with Python : Collecting Data From the Modern Web.
- [11] Deib.polim. home.deib.polimi.it/. [Online].  
HYPERLINK  
"http://home.deib.polimi.it/matteucc/Clustering/tutorial\_html/kmeans.html"  
[http://home.deib.polimi.it/matteucc/Clustering/tutorial\\_html/kmeans.html](http://home.deib.polimi.it/matteucc/Clustering/tutorial_html/kmeans.html)
- [12] Shan Feng. (2016) Enterprise Innovation with Data Mining Method Based on Naive Bayes Model Algorithm.
- [13] Disqus. bigdata.com.ar. [Online]. HYPERLINK  
"http://bigdata.com.ar/2014/09/10/random-forest-que-es-como-funciona-en-criollo/"  
<http://bigdata.com.ar/2014/09/10/random-forest-que-es-como-funciona-en-criollo/>
- [14] Inaki Inza y Pedro Larrañaga Abdelmalik Moujahid. sc.ehu.es. [Online]. HYPERLINK  
"http://www.sc.ehu.es/ccwbayes/docencia/mmcc/docs/t6s-evaluacion"  
<http://www.sc.ehu.es/ccwbayes/docencia/mmcc/docs/t6s-evaluacion>