

# **“Estimación de los rangos de los Oficiales Navales en Administración de Recursos Humanos para el Personal Naval basado en la implementación de modelos de Predicción y Clasificación”**

Fabricio Echeverría<sup>1</sup>, Miguel Aguirre Pazmiño<sup>2</sup>, Paulo Bedoya Piloza<sup>3</sup>, Jorge Gallegos Grunauer<sup>4</sup>

<sup>1</sup> Director de Tópico, Magíster en Sistema de información Gerencial, 2006, Profesor de ESPOL desde 2000; e-mail: pechever@uniplex.com.ec

<sup>2</sup> Ingeniero en Computación Sistemas Tecnológicos 2006; email:msaguirr@espol.edu.ec

<sup>3</sup> Ingeniero en Computación Sistemas Tecnológicos 2006; email:pbedoya@espol.edu.ec

<sup>4</sup> Ingeniero en Computación Sistemas Tecnológicos 2006; email:jgalleg@espol.edu.ec

## **Resumen**

*El presente proyecto se lo realizó para la MARINA DEL ECUADOR y está enfocado en la estimación y predicción del personal Naval basándose en la minería de datos.*

*La MARINA DEL ECUADOR como parte de sus procesos cotidianos hace cambios en su estructura organizacional (políticas internas de la institución) lo que puede decaer en problemas generales como: la ausencia de personal para rangos superiores o conglomeración de personal en rangos inferiores, lo que implica que estos problemas no aparezcan inmediatamente sino a largo plazo. Ante esto se ha desarrollado el sistema denominado “Predictor Básico Naval” que posee como objetivos establecer tendencias que ayuden a predecir la cantidad de personal que se encontraran en un determinado rango y establecer si una persona será o no capaz de ascender haciendo un análisis de las diferentes calificaciones obtenidas a lo largo de su vida militar.*

*El proceso de minería de datos del “PREDICTOR BÁSICO NAVAL” tiene como base el algoritmo de Naive Bayes que es plasmado en una red bayesiana y el proceso de modelación estadística paramétrica por medio de regresión polinomial. Se ha usado la base de datos del personal de la MARINA DEL ECUADOR para construir los modelos predictivos.*

## **Abstract**

*The present project was carried out it for the MARINA DEL ECUADOR and it is focused in the estimate and the naval personnel's prediction being based on the data mining.*

*The MARINA DEL ECUADOR like part of its daily processes makes changes in its organizational structure (internal politicians of the institution) what can decay in general problems as: personnel's absence for superior ranges or conglomeration of personal in inferior ranges, what implies that these problems don't appear immediately but long term. Before this, the denominated system “PREDICTOR BÁSICO NAVAL” has been developed that it possesses as objectives tendencies that help to predict the quantity of personnel that they were in a certain range and to settle down to settle down if a person will be or not able to ascend making an analysis of the different qualifications obtained along her military life.*

*The process of data mining of the “Naval Basic Predictor” has like base Naive Bayes's algorithm that is captured in a bayesian network and the process of parametric statistical carried out by means of regression polinomial. The database of the personnel of the MARINA DEL ECUADOR has been used to build the model predictive.*

## **Introducción**

Uno de los procesos mas importantes y claves dentro de una institución militar como La Marina del Ecuador son la mecánica de ascensos y bajas del personal, pues de esto depende en gran parte la planificación operativa de la institución a nivel de apertura de nuevas plazas que generen un número adecuado de vacantes orgánicas que permitan el desarrollo eficiente de la Marina.

Nosotros tenemos como objetivo desarrollar una herramienta, que basada en métodos de minería de datos, permita predecir desde diferentes aspectos la manera en que se van a distribuir los diferentes rangos en el corto y mediano plazo basados en el análisis de las tendencias examinadas en los años anteriores. Esto proveerá a la institución valiosa información sobre la cual podrá basarse al momento de tomar decisiones sobre las futuras planificaciones.

### **1. Análisis del Problema**

La Marina del Ecuador amparada en sus reglamentos realiza periódicamente procesos dentro de su personal que se manifiestan por medio de ascensos y descensos (o bajas). Estos procesos se encuentran fundamentados por medio de reglamentos internos que se aplican a los diferentes rangos y establecen en base a un conjunto de parámetros al personal idóneo que subirá de rango en un determinado momento.

El ascenso de rango le trae consigo a la Marina del Ecuador problemas Logísticos al no disponer de una herramienta de predicción que indique que cantidad de miembros subirán de un rango a otro en un lapso de tiempo dado.

La herramienta desarrollada permitirá realizar una predicción a corto y mediano plazo del personal naval para lo cual es necesario:

- 1.- Disponer de reglamentos internos y de todas las disposiciones generales hechas para extraer el conocimiento que debe ser analizado a lo largo del proceso.
- 2.- Conocer las variables que son necesarias tomar en cuenta para lograr crear un modelo determinístico que nos ayude a obtener un conocimiento que debe ser representativo del mismo. A estas variables fue necesario parametrizarlas a través del análisis histórico de datos contenidos en la base de datos de la Marina del Ecuador y luego tratarlas por medio de las herramientas de minería de datos.
- 3.- Analizar los diferentes casos o conductas que pueda presentar la información de los individuos que se encuentran en el repositorio de datos diferenciándolos por rango para lograr establecer tendencias propias de un grado determinado, conociendo que no será igual el comportamiento promedio de un oficial que el de un marino de tropa, ya que son dos grandes divisiones dentro de la institución.

- 4.- Identificar problemas propios de cada uno de los grupos a analizar y que tenga algún indicio de que ocurra o quiera conocer para aplicar métodos de análisis discriminante y obtener resultados que discernan esta inquietud.

### **2. Resultados Esperados**

El escenario actual contempla muchos compromisos a la hora de hacer una buena selección en la que es necesario contemplar primeramente el proceso y luego analizar las falencias que conlleva hacerlo de la manera tradicional.

Inicialmente los objetivos específicos están enfocados a estimar el personal ideal que tendrá la Marina del Ecuador para los años posteriores en forma general por cada división, el cual será una estimación oficial por oficial o tripulante por tripulante y demás grados, considerando la cantidad de personal pero en forma detallada, por los grados que presenta cada división.

En este proceso automatizado de selección, se necesita determinar la cantidad de personal existente en cada año por cada división. Los datos deben de obedecer a los diferentes años en la que cada candidato tuvo permanencia activa dentro de la institución por decir: las actividades de un soldado en el año 1960 y posteriores años para ser considerado para un rango superior.

Para llevar a cabo este proceso de análisis es necesario ir estudiando el proceso cotidiano e ir analizando las posibles consideraciones que salgan, pues se analizan datos históricos ricos en una tendencia oculta que no es posible saber a simple vista los resultados esperados.

### **3. Algoritmo Predictivo: Red Bayesiana**

Las tareas de clasificación se encuentran asociadas a los modelos predictivos, la tarea que se usa dentro del proyecto es la clasificación por medio de estimación de probabilidades, ya que contamos con un conjunto de características asociadas a una población y el conjunto de dichas características nos dan una respuesta a la pregunta de si un miembro de la Marina del Ecuador asciende o no asciende por lo tanto tenemos una clase bien diferenciada con valores de éxito o fracaso. Se realiza por tanto un proceso de entrenamiento donde quedan clarificados los patrones que conducen a un determinado valor de la clase dentro de la población de entrenamiento. Luego se procede a estimar a que clase pertenece un determinado individuo que posee ciertas características. El método empleado para resolver este problema y llegar a poseer resultados que realmente estimen y predigan si un marino tendrá posibilidades de ascender, si mantiene el mismo patrón en sus calificaciones, son las técnicas

bayesianas que se basan en estimar la probabilidad de pertenencia (a una clase o grupo), mediante la estimación de probabilidades condicionales utilizando el teorema de Bayes y llegando a implementar una Red Bayesiana.

Tenemos las siguientes variables aleatorias para la población de la Marina del Ecuador:

- 1.- Posee calificaciones en su nota de Rango por arriba del promedio general.
- 2.- Sanciones disciplinarias.
- 3.- Rendimiento físico.
- 4.- Puntaje de Antigüedad alto.
- 5.- Ascende de rango

Las relaciones causales y el conocimiento probabilístico asociado quedan reflejadas en una red bayesiana.

A partir de los resultados obtenidos al aplicar las fases de la minería de datos a la Base de datos de la Marina del Ecuador aplicamos el modelo predictivo (métodos bayesianos). Para esto nos valemos del teorema de Bayes para tabular las probabilidades de que un individuo dado que posee ciertas calificaciones tenga opción a ascender o no. Nuestra tarea en el proyecto fue obtener una red Bayesiana que nos indique la posibilidad de ascenso y por consiguiente la de no ascenso dado que las calificaciones de un individuo sigan el mismo patrón que el que tienen en el momento de realizar la predicción.

#### **1ra Fase.**

Durante la primera fase creamos la estructura de una red bayesiana realizando una etapa de entrenamiento en que se cargan los datos, en esta fase de entrenamiento se calculan las probabilidades de éxito o fracaso para las variables usando el teorema de Naive Bayes.

$$P(A_i|B) = \frac{P(A_i \cap B)}{P(B)} = \frac{P(B|A_i)P(A_i)}{\sum_{j=1}^n P(B|A_j)P(A_j)}$$

Aplicando la formula creamos un Árbol (red bayesiana) que posee en sus hojas las probabilidades estimadas por el teorema.

Nosotros utilizamos aproximadamente un 60% de los datos de calificaciones por rango para realizar la etapa de entrenamiento de la red bayesiana.

#### **2da Fase**

Durante la segunda fase se realiza el proceso de pruebas (testing) de la red Bayesiana mediante el proceso de alimentación de nuevos datos y de su inclusión como miembros del árbol, estos nuevos miembros pasaran a alimentar los datos de entrenamiento y aprendizaje de la red e influirán en

la próxima ocasión en que se intente predecir un ascenso.

En esta fase también se nos presenta el problema de contar en ciertos casos con datos desconocidos, estos casos los manejamos haciendo una propagación de probabilidades en la red bayesiana, introduciendo como evidencia observada los valores del caso que no son desconocidos y por supuesto la variable objetivo, en este caso la que especifica ascenso o no de un individuo.

#### **4. Algoritmo Predictivo: Regresión Polinomial**

Asociada a la tarea de Regresión usamos el método de “Regresión Polinomial” que se encuentra catalogado como una técnica algebraica y estadística que expresa modelos y patrones por medio de análisis de ecuaciones matemáticas, funciones lineales, expresiones algebraicas, medias y varianzas. Por medio de este modelo se llegan a tener expresiones algebraicas que nos permiten elaborar una función polinómica que al ser evaluada para los próximos 5 años nos devuelven las proyecciones de personal naval por grados.

El sistema construye un modelo matemático para estimar los parámetros “b<sub>i</sub>” desconocidos y establecer la curva representativa del rango de acuerdo al siguiente modelo matemático:

$$y = b_0 + b_1x + b_2x^2 + \dots + b_nx^n$$

Como en este caso tenemos una sola variable dependiente entonces realizamos una regresión polinómica simple. Computacionalmente nos valemos del algebra matricial para resolver el problema.

Para el caso de una sola variable independiente, el grado del polinomio de mejor ajuste se puede determinar al graficar un diagrama de dispersión de los datos que se obtienen de un experimento que da n pares de observaciones de la forma {(x<sub>i</sub>, y<sub>i</sub>); i = 1, 2, ..., n} en nuestro caso Año vs. Número de ascensos por año.

#### **5. Resultados y Conclusiones**

Los algoritmos usados en la aplicación muestran resultados que son dependientes de la cantidad de información con la que se cuente, de allí su exactitud o su aproximación.

Se necesita que todas las poblaciones de datos residentes en la base de datos de trabajo estén casi sin errores (sin valores nulos y sin valores extraños), por lo que en caso contrario se necesita aplicar técnicas de normalización y discretización, lo que demandaría tiempo y recursos.

Este proyecto es el resultado de un estudio de los diferentes procesos manuales para la ascensión de personal, los que demandan recursos y tiempo, por lo

que con nuestra aplicación se pretenden reducir dichos costos y optimizar los procesos llegando a obtener un práctico modelo predictivo basado en la minería de datos.

### **6. Agradecimientos**

Agradecemos a nuestros padres por su tarea infatigable de inyectarnos perseverancia y animarnos a alcanzar la meta que nos convierta en nuevos profesionales productivos para el mercado laboral cada vez más exigente.

También agradecemos a nuestros maestros que tanto se han esforzado por compartir su conocimiento y experiencias en cada una de las clases impartidas a lo largo de nuestra vida universitaria, cada una de sus clases será de gran utilidad en nuestra próxima vida laboral.

A la Escuela Superior Politécnica del Litoral en general por ser una institución de sólidos principios y capaz de formar profesionales de gran valía.

### **Referencias**

- Microsoft [www.microsoft.com](http://www.microsoft.com) con la búsqueda de licencias
- Monografías [www.monografias.com](http://www.monografias.com) con la búsqueda de modelos y herramientas de

desarrollo para proyectos de software (CASE).

### **Referencias bibliográficas a través de textos como:**

- “Introducción a la minería de Datos” de Orallo, Quintana y Ramírez.
- “Análisis de Datos Multivariantes” de Daniel Peña.
- “Probabilidad y estadística para ingenieros” de Millar y Freud.
- La documentación UML está basada en la obra “Diseño del Sistema de Tarjeta de Crédito Con UML” Javier Mendoza Navarro LIMA – PERÚ 2003 de la cual se extraen partes para denotar conceptos y por la cual constituyen las bases de dicha documentación.
- “Estimación de los rangos de los Oficiales Navales en Administración de Recursos Humanos para el Personal Naval basado en la implementación de modelos de Predicción y Clasificación” (Tesis, de la facultad de ingeniería en Electricidad y Computación. Escuela Superior Politécnica del Litoral, M. Aguirre, P. Bedoya, J. Gallegos, 2006)