



ESCUELA SUPERIOR POLITECNICA DEL LITORAL

Facultad de Ingeniería en Electricidad y Computación

“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 GRADO

Previo a la Obtención del Título de:

INGENIERO EN COMPUTACION ESPECIALIZACION EN SISTEMAS TECNOLOGICOS

Presentada por:

Miguel Stalin Aguirre Pazmiño
Paulo César Bedoya Piloza
Jorge Enrique Gallegos Grunauer

GUAYAQUIL – ECUADOR
AÑO - 2006

TRIBUNAL DE GRADUACIÓN

Ing. Holger Cevallos
SUBDECANO DE LA FIEC
PRESIDENTE

Ing. Fabricio Echeverría
DIRECTOR DE TÓPICO

Ing(a). Mónica Villavicencio
MIEMBRO DEL TRIBUNAL

Ing. Carlos Jordán
MIEMBRO DEL TRIBUNAL

DECLARACION EXPRESA

“La responsabilidad del contenido de esta Tesis de Grado, nos corresponden exclusivamente; y el patrimonio intelectual de la misma a la ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL”

(Reglamento de Graduación de la ESPOL)

Miguel Aguirre Pazmiño

Paulo Bedoya Pilozo

Jorge Gallegos Grunauer

TABLA DE CONTENIDO

TABLA DE CONTENIDO	I
ÍNDICE DE ILUSTRACIONES	V
INTRODUCCIÓN.....	I
1. ANÁLISIS DEL PROBLEMA.....	4
1.1. PLANTEAMIENTO DEL PROBLEMA.....	4
1.1.1. <i>Consideraciones del problema</i>	5
1.2. DEFINICIÓN DEL PROBLEMA Y PRESENTACIÓN DE LOS ASPECTOS DE LA SOLUCIÓN	6
1.3. SITUACIÓN ACTUAL DE LA PLANIFICACIÓN	9
1.4. RESULTADOS ESPERADOS.....	9
1.4.1. <i>Objetivos Generales</i>	14
1.4.2. <i>Objetivos Específicos</i>	15
1.5. DIFERENTES ASPECTOS DE LA SOLUCIÓN	15
2. DISEÑO DE LA SOLUCIÓN.....	20
2.1. ESPECIFICACIÓN UML DEL PROBLEMA.....	20
2.1.1. <i>DIAGRAMA DE CASOS DE USO</i>	22
2.1.2. <i>CASOS DE USO</i>	23
2.1.3. <i>ESCENARIOS</i>	25
2.1.4. <i>DIAGRAMA DE CLASES</i>	31
2.1.5. <i>DIAGRAMA DE SECUENCIAS</i>	32
2.1.6. <i>ESTIMACION DE PERSONAL NAVAL</i>	32
2.1.7. <i>PREDICCIÓN DE PERSONAL NAVAL</i>	34

2.2.	ALGORITMOS RELACIONADOS CON LA SOLUCIÓN DEL PROBLEMA.....	36
2.2.1.	<i>Modelo de Minería de Datos</i>	36
2.2.2.	<i>Tarea de Regresión</i>	36
2.2.3.	<i>Tarea de Clasificación y Estimación de Probabilidad</i>	37
2.3.	JUSTIFICACIÓN DE LOS ALGORITMOS UTILIZADOS	38
2.3.1.	<i>Algoritmo de Regresión Polinomial</i>	38
2.3.2.	<i>Algoritmo de Redes Bayesianas</i>	40
2.4.	DISEÑO MULTIDIMENSIONAL DE LA BASE DE DATOS.	41
2.5.	ANÁLISIS DISCRIMINANTE.	44
2.5.1.	<i>Objetivo</i>	44
2.5.2.	<i>Resultado</i>	45
2.5.3.	<i>Etapas del análisis discriminante.</i>	45
2.6.	FLUJO DE VENTANAS.....	46
2.6.1.	<i>INICIO DEL SISTEMA.</i>	46
2.6.2.	<i>MENU PRINCIPAL DEL SISTEMA.</i>	47
2.6.3.	<i>MENU PREDICCIÓN.</i>	48
2.6.4.	<i>MENU ESTIMACION.</i>	55
3.	CIRCULO VIRTUOSO DE LA MINERÍA DE DATOS	59
3.1.	ESTUDIO DEL COMPORTAMIENTO DE LA INFORMACIÓN EN BASE A MODELOS DE PREDICCIÓN.....	59
3.1.1.	<i>INTRODUCCIÓN.</i>	59
3.1.2.	<i>Modelos Predictivos: Regresión.</i>	63
3.1.3.	<i>Clasificación Bayesiana.</i>	66
3.2.	ANÁLISIS DE LOS DIFERENTES RESULTADOS Y SU IMPACTO	79
3.2.1.	<i>Ascensos Navales</i>	79

3.2.2. <i>Estimación del Personal Naval por años</i>	79
3.3. CALIDAD DE LOS RESULTADOS CON RELACIÓN A LOS DATOS	80
3.3.1. <i>Modelo Predictivo: Tarea de Regresión</i>	80
3.4. ESTUDIO BREVE DE LA VALIDEZ DE LOS MODELOS DE PREDICCIÓN A LARGO PLAZO	82
3.4.1. <i>Predicción: Modelo Bayesiano</i>	82
3.4.2. <i>Regresión</i>	85
4. ANÁLISIS ECONÓMICO	87
4.1. ANÁLISIS DE COSTO.....	87
4.1.1. <i>COSTO DEL PLANTEAMIENTO</i>	87
4.1.2. <i>COSTO DE LA SOLUCION</i>	94
4.1.3. <i>COSTO DE LAS HERRAMIENTAS INVOLUCRADAS</i>	98
4.1.4. <i>COSTO GLOBAL DE LA APLICACIÓN</i>	99
4.2. IMPACTO ORGANIZACIONAL POR AHORRO ECONOMICO DE PROCESOS	100
4.3. ANÁLISIS COMERCIAL.....	104
CONCLUSIONES Y RECOMENDACIONES.....	105
ANEXO A: LEYES DE ASCENSO (EXTRACTO DE ALGUNAS LEYES). 108	
ANEXO B: REGLAMENTOS (EXTRACTO DE ALGUNOS REGLAMENTOS)	110
ANEXO C: RANGOS	114

ÍNDICE DE TABLAS

<i>Tabla 2.3.1: Lista de valores que muestran la relación de los ascensos con los años para un grado en particular.....</i>	<i>39</i>
<i>Tabla #3.1.2 (b): Tabla que almacena por año y por rango (código) los números de elementos activos.</i>	<i>64</i>
<i>Tabla #3.1.3.1: Lista de calificaciones.</i>	<i>71</i>
<i>Tabla #3.1.3.2 (b): Lista de calificaciones discretizados para Rango: Capitanes de Fragata CPF.</i>	<i>75</i>
<i>Tabla #3.1.3.3 (a): Lista de calificaciones discretizados para Rango: Capitanes de Fragata CPF.</i>	<i>77</i>
<i>Tabla #3.1.3.3 (b): Registro de calificaciones.</i>	<i>77</i>
<i>Tabla #3.1.3.3 (b): Registro de calificaciones discretizado.....</i>	<i>77</i>
<i>Tabla #4.1.1: cuadro que representa los gastos básicos internos del grupo de desarrollo.</i>	<i>93</i>
<i>Tabla 1.1: Grados de Oficiales.....</i>	<i>114</i>
<i>Tabla 1.2: Grados de Tripulantes.....</i>	<i>114</i>

ÍNDICE DE ILUSTRACIONES

<i>Figura #2.1.2.1: Diagrama de casos de uso para estimación y ascenso de Personal.....</i>	<i>22</i>
<i>Figura #2.1.5.1: Diagrama de clases para la predicción de personal naval.</i>	<i>31</i>
<i>Figura #2.1.5.2: Diagrama de clases para la estimación de personal naval.</i>	<i>31</i>
<i>Figura #2.1.6.1: Diagrama de secuencia caso 1.....</i>	<i>32</i>
<i>Figura #2.1.6.1: Diagrama de secuencia caso 2 (a).</i>	<i>32</i>
<i>Figura #2.1.6.1: Diagrama de secuencia caso 2(b).</i>	<i>33</i>
<i>Figura #2.1.6.1: Diagrama de secuencia caso 2 (c).</i>	<i>33</i>
<i>Figura #2.1.7.3: Diagrama de secuencia caso 3.....</i>	<i>34</i>
<i>Figura #2.1.7.4: Diagrama de secuencia caso 4.....</i>	<i>34</i>
<i>Figura #2.1.7.5: Diagrama de secuencia caso 5.....</i>	<i>35</i>
<i>Figura #2.1.7.6: Diagrama de secuencia caso 6.....</i>	<i>35</i>
<i>Figura #2.1.7.7: Diagrama de secuencia caso 7.....</i>	<i>36</i>
<i>Figura #2.2.1(a): Fórmula para un polinomio de grado 4.....</i>	<i>37</i>
<i>Figura #2.2.1 (b): Representación gráfica de dispersiones aplicando regresión.</i>	<i>37</i>
<i>Figura #2.3.1: Representación gráfica de una regresión lineal.</i>	<i>40</i>
<i>Figura #2.3.2: Fórmula de la probabilidad de Bayes.</i>	<i>41</i>
<i>Figura #2.4 (a): Representación gráfica de la relación de la base de dato con el mundo exterior.....</i>	<i>42</i>
<i>Figura #2.4 (b): Modelo multidimensional (los atributos se mencionan en el capítulo 3).</i>	<i>43</i>

<i>Figura #2.6.1.1 Pantalla de ingreso al sistema.</i>	46
<i>Figura #2.6.1.2 Mal inicio de sesión.</i>	47
<i>Figura #2.6.2.1: Menú del sistema.</i>	47
<i>Figura #2.6.3.1: Menú predicción, configuración.</i>	48
<i>Figura #2.6.3.2: Menú predicción, configuración con valores erróneos.</i>	48
<i>Figura #2.6.3.3: Menú predicción, no hay información que procesar.</i>	49
<i>Figura #2.6.3.4: Menú predicción, si hay información a trabajar.</i>	49
<i>Figura #2.6.3.5: menú predicción, generación del árbol bayesiano.</i>	50
<i>Figura #2.6.3.6: Menú predicción, resultado de la generación del árbol bayesiano.</i>	50
<i>Figura #2.6.3.7: Menú predicción, intento de ejecutar el proceso de generación sin información.</i>	51
<i>Figura #2.6.3.8: Menú predicción, sección de evaluación.</i>	51
<i>Figura #2.6.3.9: Menú predicción, prueba individual.</i>	52
<i>Figura #2.6.3.10: Menú predicción, pruebas masivas.</i>	52
<i>Figura #2.6.3.11: Menú predicción, resultados pruebas colectivas por archivos.</i>	53
<i>Figura #2.6.3.12: Menú predicción, pruebas masivas por ingreso de tabla.</i>	53
<i>Figura #2.6.3.13: Menú predicción, pruebas masivas resultado del ingreso por tabla.</i>	54
<i>Figura #2.6.4.1: Menú estimación.</i>	55
<i>Figura #2.6.4.2: Menú estimación, configuración exitosa.</i>	55
<i>Figura #2.6.4.3: Menú predicción, configuración con valores erróneos.</i>	56
<i>Figura #2.6.4.4: Menú estimación, ventana de generación del modelo de regresión.</i>	56

<i>Figura #2.6.4.5: Menú predicción, resultados del modelo de regresión.</i>	<i>57</i>
<i>Figura #2.6.4.6: Menú predicción, intento de generación del modelo.</i>	<i>57</i>
<i>Figura #2.6.4.7: Menú predicción, resultado del cálculo del modelo de regresión.</i>	<i>58</i>
<i>Figura #2.6.4.8: Menú predicción, configuración con valores erróneos.</i>	<i>58</i>
<i>Figura #3.4.1.1 Poder predictivo de las redes bayesianas en función del número de casos de aprendizaje</i>	<i>85</i>

INTRODUCCIÓN

La presente tesis es un trabajo que se realizó para la Marina del Ecuador y que está enfocado a la estimación y predicción del personal 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.

Por lo tanto, la institución realiza de manera intuitiva ciertos procesos de planificación que son básicos y que están relacionados con los ascensos del personal militar los cuales pueden fundamentarse con “**Herramientas de Estimación y Predicción**” lo que permitiría generar información que ayude a tomar mejores decisiones.

Es de conocimiento que los datos con los cuales va a funcionar la aplicación son de tipo confidencial y por lo tanto reservado y exclusivo para la Marina del Ecuador.

Para el desarrollo del “*Predictor Básico Naval*” es necesario conocer los diferentes grados o rangos que maneja la Marina del Ecuador, los cuales se mencionan en el anexo del documento. Estos rangos se los tratará a lo largo de esta tesis y servirán para discutir situaciones particulares que están relacionadas a los mismos.

Los algoritmos usados en la aplicación se basan en la minería de datos los cuales se explican a continuación:

- Análisis por Discriminante: Herramientas basadas en la distinción de información y que ayuda a comprender las diferencias entre grupos; explica, en función de características observadas, por qué los objetos/sujetos se encuentran asociados a distintos niveles. En nuestro sistema usamos Redes Bayesianas como modelo clasificador discriminante de la información.
- Análisis por Estimación: Herramientas fundamentadas en la apreciación de la información y que mediante información de sucesos del pasado nos ayudan a conocer el comportamiento de los mismos en un futuro a corto plazo. En nuestro sistema usamos Regresión Lineal Multidimensional.

El “*Predictor Básico Naval*” se nutre de los datos procedentes de la base de datos con la que cuenta la institución para de esta manera ir actualizando periódicamente los valores previstos, resolviendo así el problema de la rapidez con la que se des-actualizan los resultados obtenidos, por medio de un análisis predictivo, cuya validez dependerá exclusivamente de la cantidad de datos con que se cuente.

Por medio de la herramienta que aplica métodos de predicción y de estimación, la Marina del Ecuador podrá tomar decisiones relacionadas a su personal por lo cual es necesario que se tomen en cuenta la mayor parte de atributos presentes en los reglamentos para que formen parte de esta tesis.

El presente trabajo realiza un análisis de la evidencia histórica recogida de la bases de datos de la Marina del Ecuador para explotar conocimiento que pueda establecer un conjunto de patrones y elementos que puedan ayudar a

solventar mejores planificaciones con bases tomadas en la información que la Institución posee.

Inicialmente el trabajo comienza describiendo como es la situación actual de la Institución enfocado a la planificación de los recursos militares y como éste se convierte en un problema que debe ser profundizado y estudiado para posteriormente establecer el mejor conjunto de datos que deban de alimentar a las herramientas que se han construido con el propósito de extraer conocimiento, este es el objetivo principal de este documento y tema de proyecto.

Como evidencia histórica para alimentar las herramientas usadas en este trabajo se hizo un seguimiento de las promociones y los años de ascenso, así como la búsqueda de atributos que determinen cualidades de un individuo como por ejemplo la composición de sus calificaciones anuales que permitan analizar comportamientos que indiquen situaciones que lo comprometan a futuro.

Por lo mencionado anteriormente esta tesis está basada en las leyes y reglamentos del personal interno y establece que si para cierto año se contará con cierto número de personas dependiendo del rango como primer punto y como segundo punto si una persona con rango actual puede ostentar un rango superior teniendo en cuenta sus calificaciones, las cuales serán analizadas con las tendencias que se marcan en el pasado.

CAPÍTULO 1

1. ANÁLISIS DEL PROBLEMA

1.1. Planteamiento del problema

La Marina del Ecuador basada 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 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. Este conocimiento será utilizado por las herramientas usadas en este trabajo.
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 de ser representativo del mismo. 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 tratada por medio de las herramientas de minería de datos.

3. - Analizar los diferentes casos o conductas que puedan 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.

1.1.1. Consideraciones del problema

Para realizar este estudio, se necesita determinar la cantidad de personal existente en cada año por cada división. Por ejemplo: el total de oficiales en 1950 y así con los siguientes años hasta el presente. Esta información histórica, obtenida de la base de datos existente, ayudará a determinar si existe una tendencia o un modelo al que se puedan ajustar los datos para realizar la estimación.

Es importante mencionar que la base de datos no muestra o no lleva un registro de los oficiales, tripulantes y empleados civiles de mediados de siglo pasado (se cuenta con evidencia que parte desde 1970); por lo que para desarrollar el presente proyecto es necesario

tener en cuenta que dichos registros afectarán la calidad de precisión que se desea porque el hecho de trabajar con mas campos o atributos no afectarán considerablemente lo que se desea en el caso de estimar personal naval para los siguientes años.

Necesitamos determinar el número de personal en los diferentes conjuntos formados por los rangos navales (véase Anexo C).

1.2. Definición del problema y presentación de los aspectos de la solución

Para lograr diferenciar entre los diversos conjuntos necesitamos agruparlos considerando características similares entre los diversos rangos. En cada uno de los grupos que se forme vamos a tener diversas variables que son las que determinan si un miembro sube o no de rango por lo que será necesario cambiar de dominio el tipo de dato usado por estas variables e ingresarlas a un análisis de correlación por medio de una regresión lineal que permita modelar las tendencias por medio de una representación lineal.

Ejemplos de las variables que influyen en sentido general en la subida o descenso de personal tenemos:

- Resultados de las evaluaciones que periódicamente realiza la Marina del Ecuador a sus miembros.
- Penalizaciones obtenidas por un determinado individuo.
- Antigüedad en el rango actual
- Condecoraciones obtenidas.
- Becas obtenidas, en el país o en el exterior

Es importante observar que el valor de estas variables es muy distinto para un nivel de rangos u otro, por lo que se puede presentar el fenómeno de que existan muchos oficiales que estén listos para subir de un rango a otro dejando vacante su rango anterior, que no podrán ser llenadas porque existen muy pocos oficiales que están listos para subir a ese rango. Este fenómeno puede causar que en un determinado momento exista una sobrecarga de elementos en un rango y muy pocos en otros, entonces, por medio de nuestro sistema lo que se pretende modelar es la situación a futuro de cómo estos casos puedan afectar la contingencia de personal.

Para conllevar este problema el sistema debe mostrar cuales serán las necesidades en cierto tiempo de tal manera que los directivos pertinentes puedan tomar medidas que eviten que se produzcan estos fenómenos.

La Marina del Ecuador utiliza sus reglamentos internos como fuente básica para decidir si un miembro asciende o no, los reglamentos que contienen artículos referentes son:

- Reglamentos de Carrera de Oficiales.
- Ley de Personal.
- Reglamento de Calificación de Requisitos de Ascenso y Establecimiento de Antigüedades del Personal Militar.

Los siguientes párrafos son ejemplos como la Marina del Ecuador regula los ascensos:

- La promoción de los oficiales está determinada por las fechas en que el oficial cumpla con los requisitos estipulados en la Ley, la

misma que establece el 20 de Diciembre y 20 de Junio como fechas únicas para los ascensos.

- La fecha de ascenso de los oficiales sancionados con suspensión de funciones o pena privativa de la libertad por menos de 90 días, es postergada al semestre posterior al que le correspondía normalmente
- El tiempo de permanencia en una unidad operativa es computado desde la fecha de presentación hasta la fecha de trasbordo.
- El cómputo del tiempo de embarque incluye el periodo en que el oficial sirve a bordo de un buque no incorporado a la Marina del Ecuador pero que está en fase de pruebas de mar o en fase de transferencia; se fija la fecha de inicio de la contabilización del tiempo mediante acta de autoridad competente.
- Así mismo, se contabiliza el tiempo de permanencia en crucero de instrucción o pasantías profesionales, en buque extranjero.

Las funcionalidades requeridas por el sistema serán las que:

- Producto de la aplicación de cada uno de los reglamentos para los diferentes rangos militares, obtenga valores que representen los mismos y sirvan de entrada para el sistema.
- Busque tendencias dentro de la gran base de datos actual para de esta manera estimar y predecir como se van a comportar las diversas variables que influyen al proceso.
- Reciba como información la población actual dominios y sus rangos de la Marina del Ecuador.

- Aplique los reglamentos considerando la estimación de las variables hechas.
- Genere conocimiento que permita encontrar soluciones a los problemas derivados del ascenso de personal.

1.3. Situación actual de la planificación

El proceso de ascenso de grado está establecido según el tiempo que el aspirante ha permanecido en su grado actual (5 años por lo general) y si no ha cometido faltas que impidan su ascenso.

Esta información se encuentra almacenada en la base de datos de La Marina del Ecuador y se actualiza continuamente. Un programa desarrollado en Visual Basic (usado en la actualidad por la institución) permite conocer el estado actual de los diferentes miembros de la Institución, pero no responde a las preguntas antes planteadas.

Para analizar las tendencias en cuanto a la cantidad de personal para cierto tiempo se necesitaría de un trabajo previo para establecer esta información, si se quiere analizar en base a las tendencias si un individuo es propenso a un ascenso observando patrones se tendría que analizar las hojas de individuos que hayan pasado por el rango de análisis a través del tiempo.

1.4. 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 si es que se analizará con datos históricos ricos en una tendencia oculta que no es posible saber a simple vista y por supuestos los resultados esperados.

Por lo tanto, a continuación, se presentan algunos análisis en base a los reglamentos [reglamento sección anexos literal A] que se han considerado como parte clave en nuestro proyecto.

A.-) Según el reglamento las personas con mayor antigüedad son las que tienen mayor opción a ascender, por lo que esto es una de las características a analizar en el modelo y adicionalmente se tiene que considerar su historia militar para conocer mediante ciertos parámetros su comportamiento a lo largo de su vida militar.

B.-) Siguiendo con el estudio de las normas más importantes para nuestro análisis encontramos [reglamento sección anexos Literal B] que el orden de ascenso está encaminado por un tiempo mínimo en la que un elemento activo pueda ascender y a su vez dispone que todo elemento becado en el extranjero pueda subir de rango en el doble de tiempo que le conllevaría a un oficial del mismo rango hacerlo. En nuestro análisis destaca la necesidad de mirar y analizar datos relevantes al caso como historia militar antes del tiempo estipulado en el caso que lo hubiere para que de esa forma se pueda tener un amplio margen de comparación no solamente de los que están dentro del periodo a considerar sino también a aquellos que no han subido de rango pero dentro de un rango de tiempo estipulado (es decir aquellos que según el reglamento tienen más tiempo como es el caso por citar de algún oficial que tiene 6 años de permanencia activa y según el reglamento los que tengan 5 años solo deben de ser considerado).

C.-) En cuanto se refiere al requisito de haber sido calificado en la primera oportunidad [reglamento sección anexos literal C], para el ascenso a ciertos grados, y en el caso de no se pueda dar por falta de vacancia en el orgánico se le dará al elemento una segunda oportunidad, la cuál será con la siguiente promoción. En nuestro estudio se considera este caso como parte de los ya existente a seguir, solo que estos elementos se les debe de considerar un merecimiento para el puesto basado en calificación

(la cual supone que debe de estar entre las mejores por haber sido considerados como parte del proceso anterior en el cual no pudieron ascender) y la cual dictamina un valor por cada proceso en el cual haya estado involucrado, observándose que para cada caso habrá un valor que sea similar o en caso excepcionales marque la diferencia.

También podremos citar leyes mas profundas al reglamento de ascenso y hacer un análisis con estas leyes las cuales se mencionan a continuación:

D.-) En los literales donde se menciona requisitos [reglamento sección anexos literal D] para el ascenso de oficiales de arma y de servicios o técnicos a grado superior el proceso está enfocado en datos que representan cursos esenciales para el grado, asuntos disciplinarios, de diplomacia y finalmente académico en lo que corresponde a reprobaciones. Nuestro enfoque va hacia la misma perspectiva pero analizada en el pasado, es decir su comportamiento en los mismo parámetros pero abriendo los mismo para que sean tomados en cuenta más puntos de análisis como es el caso de los cursos de grado para subir de grado, si lo enfocamos en los diferentes cursos que ha tomado el elemento podemos analizar su desempeño intelectual.

En los literales d, e, f, g del mismo reglamento [reglamento sección anexos literal D] analizando el proceso de selección encontramos que el enfoque está dictaminado por vacantes para el ascenso, asuntos que se tratan en consejos de Oficiales para la ascensión y puestos que

obligatorios para hacer merecimiento a la ascensión, si analizamos uno de estos reglamentos nos damos cuenta que el problema más grave se da ante la falta de vacante para tal puesto lo que de por si es un problema que se pretende reducir también en nuestra solución. Si tomamos los mismo parámetros y hacemos un análisis en el pasado podemos ver que los resultados son más eficaces y de cierto punto más ideales de lo que se espera, aunque tal vez el conjunto de datos de salida sea mínimo debido al pobre conjunto de datos iniciales con los que se pretende alimentar al algoritmo.

Nuestro análisis para los literales (a, b, c) [reglamento sección anexos literal E] nos encamina hacia el pasado de los datos historiales en busca de información trascendental como en el caso del literal a con respecto al título obtenido, si añadimos las calificaciones obtenidas dentro del curso (siempre y cuando sea información que se pueda extraer de la base de datos) podremos obtener un amplio criterio de selección y así aplicar para los otros literales respetando los parámetros declarados en los mismos para haciendo énfasis en el pasado o historial.

El análisis con respecto a estos literales (a, b, c, d) [reglamento sección anexos literal F] nos da a entender que el proceso para estos casos son basados en su mayoría en cursos que son obligatorios y conforme a lo que los demás reglamentos dictaminan y a su vez también está el asunto disciplinario a excepción del literal d cuyo análisis solo es por parte del consejo establecido para dicho caso. Si buscamos los mencionados parámetros (entiéndase el contenido en los literales) en

el pasado y hacemos una inferencia a patrones de comportamiento y asuntos relacionados con las calificaciones, nos damos cuenta que estos procesos le dan más peso al proceso cotidiano proporcionando un esquema de tratamiento de datos más amplio evitando la generalización.

Para respaldar nuestra solución se presenta a continuación los objetivos en los que se basó para crear la aplicación Predictor Básico Naval:

1.4.1. Objetivos Generales

- Proveer a la Marina del Ecuador una aplicación que contenga herramientas con la capacidad de estimar comportamientos y predecir posibles eventos que puedan acontecer en el presente, a corto o a largo plazo.
- Dotar de Conocimiento como producto de los resultados de la aplicación a las planificaciones regulares que mantiene la Marina del Ecuador relacionados con la preparación de individuos.
- Establecer a través de los resultados obtenidos mediante el uso de la aplicación métricas que ayuden a fomentar correctivos en diversas áreas para minimizar eventos a futuro que tengan relación con sucesos del pasado.
- Mejorar el panorama actual de las posibles decisiones a tomar mediante el conocimiento obtenido como producto del análisis tomando en consideración la información histórica que se tiene.

1.4.2. Objetivos Específicos

- Predecir la cantidad de personal con que va a contar la Marina del Ecuador discernidos por rangos y por tiempo a corto plazo, basándose en el comportamiento histórico que ha tenido la Marina del Ecuador a lo largo de un periodo de tiempo establecido.
- Estimar si los individuos que tienen un conjunto de atributos y un rango actual pueden ser capaces de ostentar un rango superior tomando en consideración el conjunto de tendencias evidenciado en un análisis de ascensos por rango observados en un intervalo de tiempo.

Cabe destacar que la aplicación “*Predictor Básico Naval*” solo toma en cuenta las calificaciones que representen cada una de estas políticas y que puedan tomarse del repositorio de datos que tiene la institución, para luego proceder mediante técnicas de adaptación a alimentar los algoritmos de minería de datos.

1.5. Diferentes aspectos de la solución

Los parámetros que acarrea tomar dichas decisiones contemplan diferentes puntos de vistas, que en algunos casos es común, pero en casos especiales sólo se enfocan en una parte que generalmente es la más relativa al caso, pero no la más relevante.

Las soluciones anteriores hacen referencia a un análisis en merecimientos otorgados y en calificaciones obtenidas a lo largo del tiempo en el servicio, entonces se comienza a estudiarlas y con ayuda de la información histórica almacenada se procede a obtener un

panorama más amplio que contiene tendencias que no son evidenciables inmediatamente.

En el literal A, la solución que se infiere asociado al proceso de ascensión se enfoca en el historial que pudieren tener todos los candidatos fuera de rango estipulado, en un enfoque más práctico esto sirve para avalar que el elemento a considerar no solo sea tomado en cuenta por su antigüedad, también está el hecho de su comportamiento a lo largo de su tiempo de vida activa en la institución. Nos damos cuenta que el hecho de analizar un historial le da mas peso y hace que el proceso de selección sea mas detallado y más exacto que el proceso cotidiano ya que le está añadiendo mas dependencias y carga al proceso lo que lo hace mas robusto.

Por otro lado en un proceso manual considerar esto conllevaría más tiempo del ordinario debido a que si se quiere tener un amplio margen de consideración entre elementos se tendría que analizar uno a uno su comportamiento lo que por supuesto acarrea errores e insatisfacción a la hora de analizar los resultados.

El caso que citamos es el que se puede dar cuando para un conjunto de 20 elementos se desea saber si van a ascender; para el efecto se ha predispuesto la vida activa de los participantes documentado en por lo menos 10 partes de las diferentes disciplinas encaminadas por cada uno de los candidatos; en números se ve algo tedioso ya que 200 partes a ser analizadas conllevaría un tiempo dedicado al ciento por ciento para determinar cuales de estos 20 elementos son los candidatos idóneos son para un proceso de selección final.

Con nuestro modelo automatizado esto solo consistirá en hacer una referencia a los datos para hacer su posterior análisis enfocados en los parámetros mencionados evitando la redundancia en la clasificación de patrones a analizar, simplemente la tendencia de evidencia en el pasado nos muestra si los individuos tendrán la oportunidad de ostentar un rango superior.

En el literal B, lo que se quiere hacer es contemplar a las personas que están fuera del periodo a considerar pero para aquellas que estuvieran dentro de la prórroga que dictaminase el consejo superior como es el caso de los que tuvieron 6 años de vida activa dentro del rango actual y en el cual pueden ser considerados para el rango superior que se ubica para los elementos de 5 años de vida activa. El hecho está en que tener una mayor serie de elementos a considerar le da al proceso de selección un amplio margen de análisis dejando a un lado los vacíos que podrían tenerse al solo contar con los que cumplen solamente la disposición inicial de los 5 años. Manualmente analizando el comportamiento de los que pudieran ingresar dentro del análisis se tendría que estudiar caso por caso del total de la promoción inmediata al merecimiento agregando los individuos que no entraron en el proceso anterior por falta de vacante.

Con nuestro modelo se pretende es sencillamente analizar si estos individuos que entran junto con los otros elementos se adaptan a las tendencias evidenciadas en las promociones anteriores y justificar mayormente su ascenso con los de la promoción en la que se encuentra (de la promoción que no es original).

En el literal C, el proceso de análisis trata de hacer referencia a los problemas que pudiere tener cada candidato obteniendo así un amplio estudio no solo por calificaciones obtenidas, haciéndose un proceso más riguroso. Esto permite que el proceso de selección no solo sea enfocado desde un punto de vista que es el de la calificación por elemento dando más peso cuando se agrega otro factor ponderante que es el del historial de los problemas o antecedentes, también esta el hecho de añadir otros elementos de consideración siempre y cuando un consejo superior dictamine que deban de tomarse en cuenta, haciendo que el modelo tenga mayores atributos por lo cual deba de discriminar.

En lo que corresponde a la solución propuesta en el literal D primera parte (literales a, b, c) lo que se propone es hacer énfasis al pasado que pudiera estar registrado en las bases de datos y el cual pueda ser estudiado minuciosamente para así dar un mejor criterio de selección, claro está el hecho de que la fidelidad de los datos pueda en los mejores de los casos dar un mejor resultado.

En estos literales (d, e, f, g) nuestra solución ya debe de acarrear el problema de las vacantes debido a que es un problema que se va corrigiendo de acorde al uso que se le de a nuestra aplicación y su interpretación a la hora de planificar posibles sucesos en el presente.

Para el análisis del literal E nuestro enfoque tomando en cuenta el historial así como los demás estudios mostrados anteriormente nos da una mejor perspectiva de lo que se quiere en otras palabras reforzar un proceso que ha sido siempre enfocado con asuntos netamente

académicos, en lo que corresponde a títulos, darle más notoriedad analizando las calificaciones y asuntos relacionados con este último parámetro como el rendimiento, pero claro está de que si hay datos que avalen esto último para extraer posteriormente y convertirlo e información trascendental al proceso.

Nuestro enfoque con respecto al literal F nos permite obtener una mejor comprensión para este proceso debido a que evitamos una generalización a este nivel, es decir, al hecho de que el número de cursos obtenidos conlleva la mayor parte de méritos dentro del proceso, por lo que el análisis inferido como parte de la solución hace rescatar a los mejor elementos para el proceso que al final mostrará un reporte con el mayor grado de aceptación, claro que siempre y cuando se pueda contar con datos que respalden dicho proceso.

CAPÍTULO 2

2. DISEÑO DE LA SOLUCIÓN

2.1. ESPECIFICACIÓN UML DEL PROBLEMA

En esta parte del capítulo mediante el uso de técnicas de modelamiento se definirá el concepto de solución aplicado a una especificación formal y práctica.

Con el uso de las técnicas del desarrollo del software se pretende crear un concepto más formal de cómo las herramientas (algoritmos de minería de datos) usadas en este trabajo se comportan ante diferentes patrones de datos y como deben de poseer atributos que puedan ser medibles, para que dichas herramientas arrojen resultados concretos.

El Lenguaje de Modelaje Unificado UML es el lenguaje de modelaje orientado a objetos estándar de la industria para: especificar, visualizar, construir y documentar los elementos de los sistemas de software, así como para modelaje del negocio y de otros sistemas que no son de software. Simplifica el proceso complejo de análisis y diseño de software, facilitando un plano para la construcción. En este trabajo se plasma la utilización del UML y su aplicación en el análisis y diseño de sistemas.

El UML representa una colección de las mejores prácticas de ingeniería que han probado ser exitosas en el modelaje de sistemas grandes y complejos. En la actualidad existen una serie de empresas que se

dedican al desarrollo de herramientas que utilizan como base la notación y sintaxis integrada en el UML.

A lo largo de los siguientes puntos se especificará casos de uso en el que se definirá todos los casos contemplados por los algoritmos de minería de datos en los que se especifica como deben de tomarse los datos para alimentar mencionados algoritmos, luego se hará la recreación de los escenarios que puedan contemplarse en los casos de uso y en el que se detallará como un conjunto de acciones pueden definir un proceso que forma parte de una tarea como cuando se ingresan los datos, que tipo de información se espera analizar. Posteriormente se definirá el diagrama de casos de uso que contemplará una visualización de todos los componentes actuando entre sí para formar una interacción que representa a los algoritmos funcionando con las especificaciones de usuario.

2.1.1. DIAGRAMA DE CASOS DE USO.

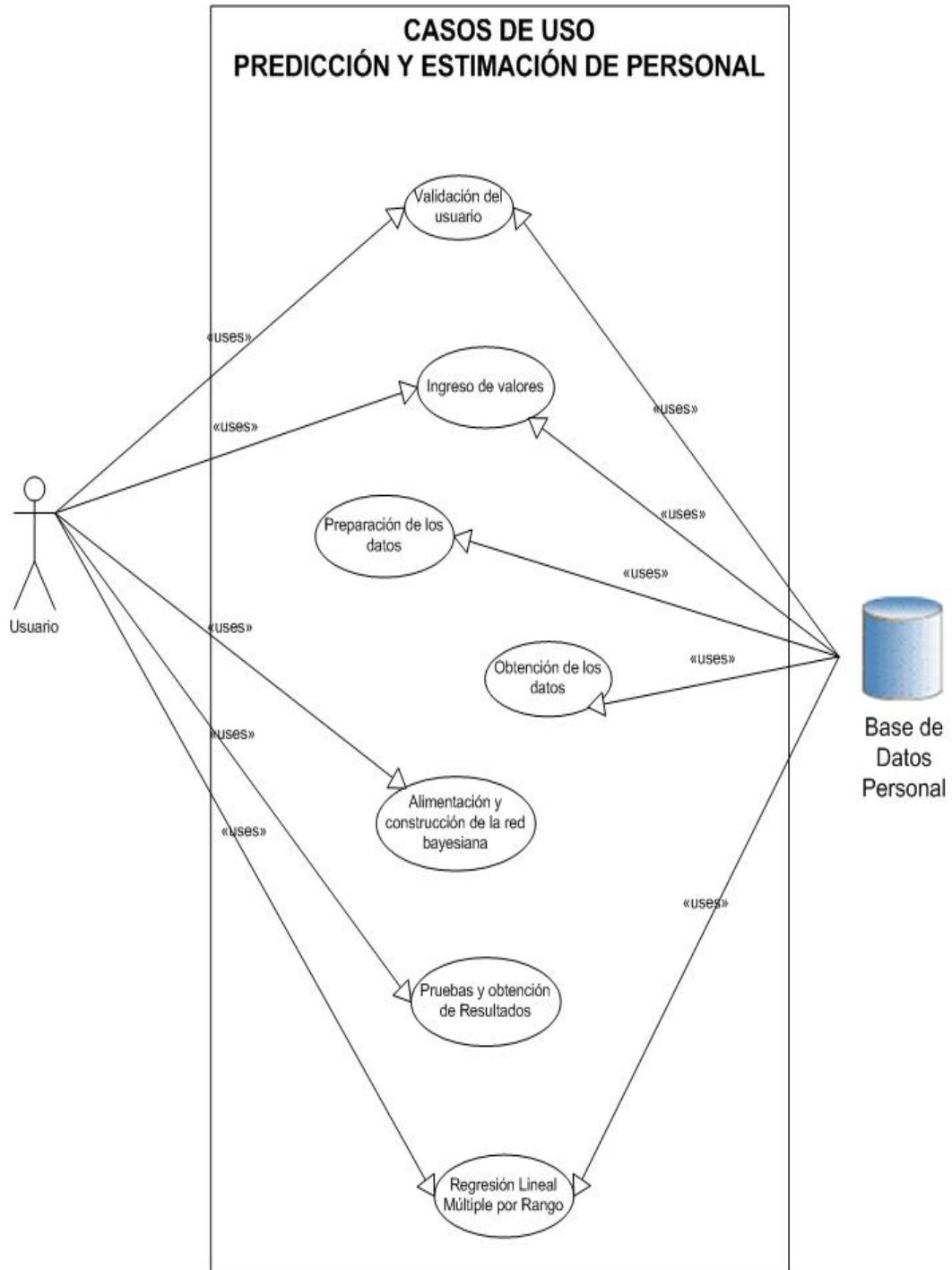


Figura #2.1.2.1: Diagrama de casos de uso para estimación y ascenso de Personal.

2.1.2. CASOS DE USO.

2.1.2.1. ACTORES.

Base de datos: es el ente que interviene a lo largo de los procesos contemplados en nuestro sistema Predictor Básico Naval y es el que lleva a cabo la operación de recuperación de datos iniciado por el usuario.

Usuario: es la persona que ingresa al sistema con un usuario y contraseña y que administra las herramientas según sea el uso.

2.1.2.2. CASOS DE USO.

El objetivo de los casos de uso en esta sección es la de dar a conocer cuales son las situaciones que se obtendrán con el uso de nuestro sistema y en particular con la preparación, configuración y obtención de resultados dependiendo de la herramienta a estudiar.

Caso de uso 1: Validación del usuario.

Especificación: Mediante este caso de uso se valida si el usuario que quiere acceder al sistema puede ingresar y hacer uso de las operaciones descritas posteriormente.

2.1.2.3. ESTIMACION DE PERSONAL POR RANGO.

Caso de uso 2: Regresión Lineal Múltiple por Rango.

Especificación: Permitir a un usuario realizar una regresión lineal y obtener una aproximación a fecha futura.

2.1.2.4. PREDICCIÓN DE ASCENSO DE PERSONAL.

Caso de uso 3: Ingreso de los valores en los parámetros iniciales.

Especificación: Se ingresan los valores que servirán de entrada al algoritmo de predicción.

Caso de uso 4: Preparación de los datos.

Especificación: Mediante este caso se prepara los datos que se encontraron como resultado del ingreso de los valores en el caso anterior.

Caso de uso 5: Obtención de los datos.

Especificación: En este caso se genera un conjunto de pasos que servirán para alimentar el algoritmo de predicción y está más enfocado a alimentar las estructuras que usa el algoritmo.

Caso de uso 6: Alimentación y construcción de la red bayesiana.

Especificación: Resultado de los casos anteriores se usan los datos cargados en las estructuras para generar los procesos que usa la RED BAYESIANA para comenzar a formar las probabilidades condicionales y establecer así un conjunto de criterios que servirán para la sección de pruebas.

Caso de uso 7: Pruebas y obtención de Resultados.

Especificación: El caso de uso muestra que mediante el conjunto de datos que se quieran ingresar ya sea individual o colectivo se generará como resultado para el caso

individual si tiene posibilidades de ostentar un rango superior o si el conjunto colectivo de individuos tendrán opciones de ascender, para esto último se pronostica la salida de dos grupos en el que se muestra los que tienen mas porcentajes de subir y los que no.

2.1.3. ESCENARIOS.

Con las especificaciones de casos de uso realizadas en el punto anterior lo que se pretende en esta parte es dar a conocer todos los posibles eventos que pueda ocurrir dada las situaciones especificados por los casos de uso.

Caso de uso 1: Validación del usuario.

Escenario 1.1: El usuario introduce clave incorrecta. Este paso consiste en la autenticación de todos los usuarios que quieran acceder al sistema mediante una clave válida dentro del dominio del sistema y por lo tanto se necesitará que un administrador de la aplicación brinde dicha clave para los propósitos deseados.

Escenario 1.2: El usuario introduce un usuario incorrecto. Al igual que el escenario anterior no se puede ingresar si no se ha autenticado correctamente un usuario cuyos datos a validar se los hace internamente en la base por lo que es deber del administrador proveer de dicha información dependiendo de los propósitos anhelados.

Escenario 1.3: El usuario intenta ingresar sin ingresar usuario y/o contraseña. El sistema como tal obliga a que se ingrese los dos

datos y por lo tanto no se justifica el ingreso sin uno de los dos ya que la autenticación es hecha en la base de datos de manejo.

2.1.3.1. ESTIMACION DE PERSONAL POR RANGO.

Caso de uso 2: Regresión Lineal Múltiple por Rango.

Escenario 2.1: Selección del rango e intervalo de análisis. Por medio del ingreso del rango y los años inicial y final se debe crear un límite para la extracción de información. Luego de presionar “Preparar” esta información quedará almacenada para generar el modelo.

Escenario 2.2: Resultado del Modelo. Aquí se tomará la información previamente preparada para realizar la generación del modelo de regresión. Aquí se presenta tanto la ecuación resultante como las gráficas que ayudan a entender el resultado.

Escenario 2.3: Resultados. Una vez generado el modelo, se puede seleccionar una fecha (a futuro) para obtener una predicción del personal en dicha fecha.

2.1.3.2. PREDICCIÓN DE ASCENSO DE PERSONAL

Caso de uso 3: Ingreso de los valores en los parámetros iniciales:

Escenario 3.1: Selección del rango o grado. Se selecciona el rango con el que se va a trabajar para lo cual se muestra un conjunto de grados posibles a escoger.

Escenario 3.2: Ingreso de tiempo a analizar. Se ingresa un periodo de tiempo que servirá para buscar y extraer información en las tablas de sucesos.

Escenario 3.3: Ingreso datos que no generan información. Es casi probable que se encuentre con esta situación debido a que se trabaja con una tabla que reúne evidencia histórica por lo que tal vez lo que si se pueda encontrar es poca o casi nada de información relevante al caso.

Caso de uso 4: Preparación de los datos.

Escenario 4.1: Extracción de los datos. Una vez que se han ingresado correctamente los valores en los parámetros de entrada se procede a relacionar toda la información que se tiene en las tablas de evidencia histórica para así trabajar con un conjunto de datos predefinidos.

Escenario 4.2: Formato e ingreso en la tabla de trabajo. Obtenida la información a trabajar se procede a dar el formato con el que va a trabajar el algoritmo dando un nivel de discretización a todos los posibles valores encontrados que corresponderán a las calificaciones con las que se van a evaluar y obtener patrones para estudiar.

Escenario 4.3: Problemas encontrados en la extracción de información. No hay suficiente evidencia histórica como para comenzar a analizar y esto puede estar relacionado al conjunto de años ingresados en los pasos anteriores.

Escenario 4.4: Problemas encontrados en el formateo e ingreso en la tabla de trabajo. Se encontraron demasiados valores nulos (sin valor) o valores aberrantes (valores extraños) lo que hace que en este caso se ingresen valores extraños a la tabla de trabajo.

Caso de uso 5: Obtención de los datos.

Escenario 5.1: Obtención segura de la información a trabajar. Se procede a obtener toda la información relacionada en la tabla de trabajo con una consulta hecha internamente.

Escenario 5.2: Llenado de las estructuras. Se comienzan a llenar las estructuras que se usarán a lo largo del proceso del algoritmo de predicción y las cuales tendrán como información lo que repose en la tabla de trabajo.

Escenario 5.3: Problemas con el llenado de las estructuras. Esto solo ocurrirá cuando se ha hecho un mal formateo de la información por problemas acarreados anteriormente lo que desencadenará en una serie de errores que puedan ser admitidos por el algoritmo pero dejando como resultado valores lejanos de lo esperados.

Caso de uso 6: Alimentación y construcción de la red bayesiana.

Escenario 6.1: Uso de las estructuras. Una vez que las estructuras están llenas correctamente se procede a repartir información que está contenida en estas y las cuales son usadas en todos los procesos del algoritmo.

Escenario 6.2: Proceso de análisis y estudio del algoritmo de predicción. Esta parte nos muestra que lo hecho anteriormente no ha dado problemas y que por lo tanto el algoritmo se encuentra trabajando correctamente con los datos almacenados en las estructuras, por lo que internamente se hace un conjunto de procesos que obedecen a la minería de datos.

Escenario 6.3: Ejecución de salidas. Como producto de los pasos anteriores y finalización de todos los procesos el algoritmo de predicción nos dará información que servirá para los pasos posteriores. Se espera la salida de un modelo de predicción para que sea visto por el usuario.

Escenario 6.4: Problemas con el algoritmo de predicción. Esto solo se dará debido a que se encontró valores aberrantes que no fueron filtrados o a valores que no concuerdan con la información que se extrajo anteriormente, lo que desencadena en el entorpecimiento del algoritmo que pueda provocar que se detenga.

Caso de uso 7: Pruebas y obtención de Resultados.

Escenario 7.1: Ingreso de parámetros a probar. Esto solo se habilitará cuando el usuario ya ejecutó los pasos anteriores y se visualizó el resultado producto del uso del algoritmo de predicción. Se deben de ingresar las calificaciones que se deseen evaluar correspondientes al individuo o al conjunto de individuos a analizar.

Escenario 7.2: Resultados de las pruebas. Como resultado del ingreso anterior el algoritmo de predicción mostrará si el individuo tiene probabilidades de una aspiración a un rango superior. Igual sucede con el conjunto de individuos ingresados solo que se visualizarán en esta situación dos grupos: los propenso a subir y los que no.

2.1.4. DIAGRAMA DE CLASES

2.1.4.1. PREDICCIÓN DE PERSONAL NAVAL

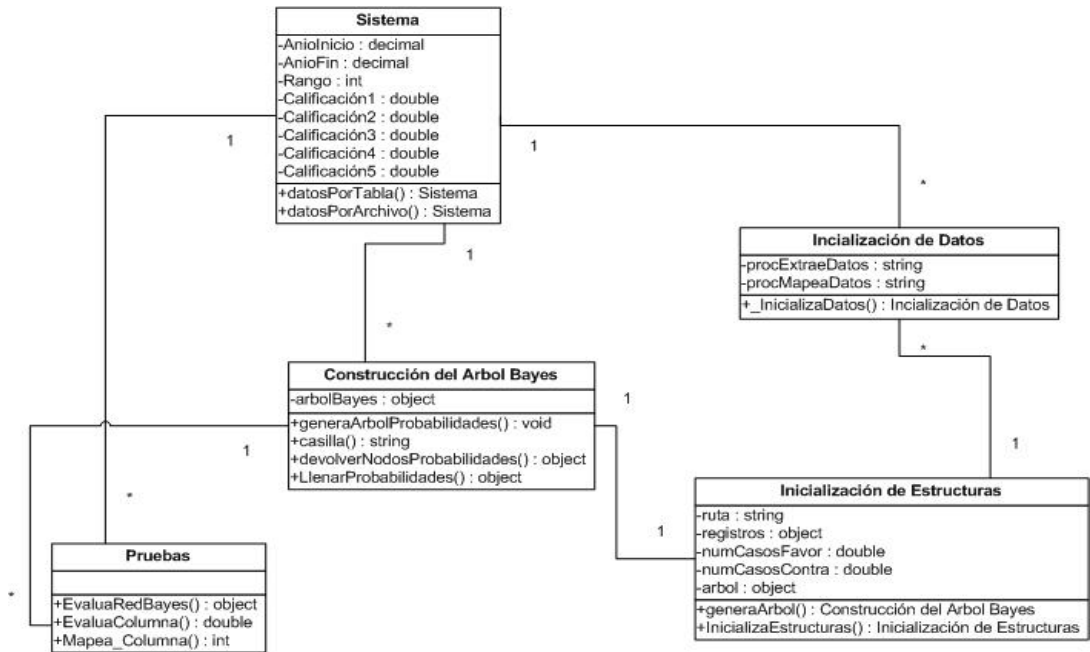


Figura #2.1.5.1: Diagrama de clases para la predicción de personal naval.

2.1.4.2. ESTIMACION DE PERSONAL NAVAL.

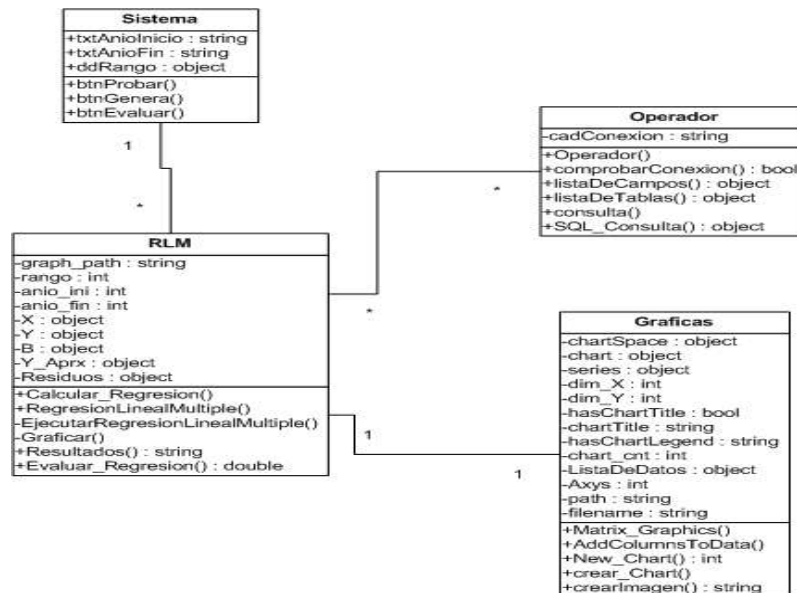


Figura #2.1.5.2: Diagrama de clases para la estimación de personal naval.

2.1.5. DIAGRAMA DE SECUENCIAS.

2.1.5.1. CASO 1: VALIDACIÓN DEL USUARIO.

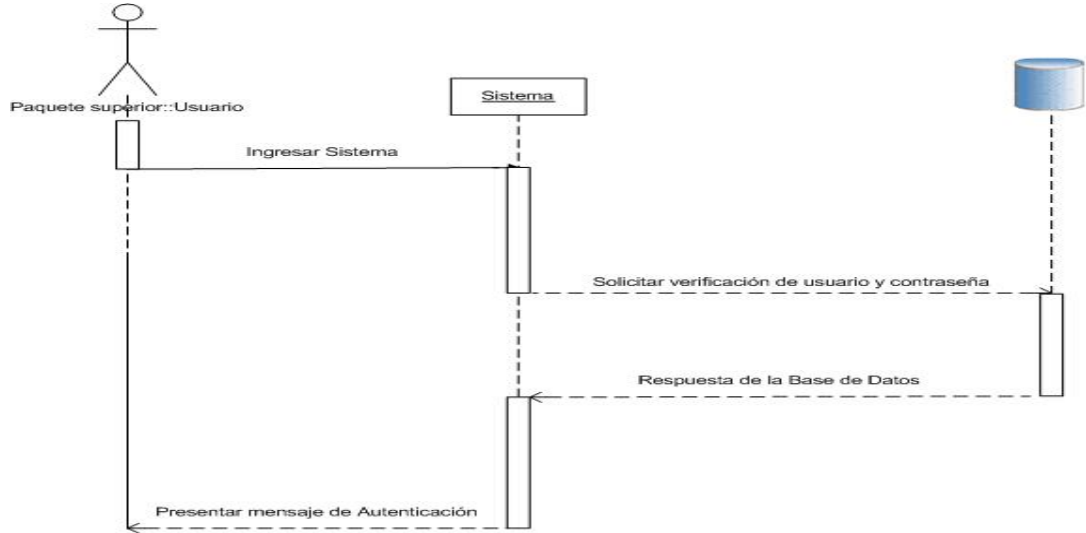


Figura #2.1.6.1: Diagrama de secuencia caso 1.

2.1.6. ESTIMACION DE PERSONAL NAVAL.

2.1.6.1. CASO 2 (a): INICIAR REGRESIÓN.

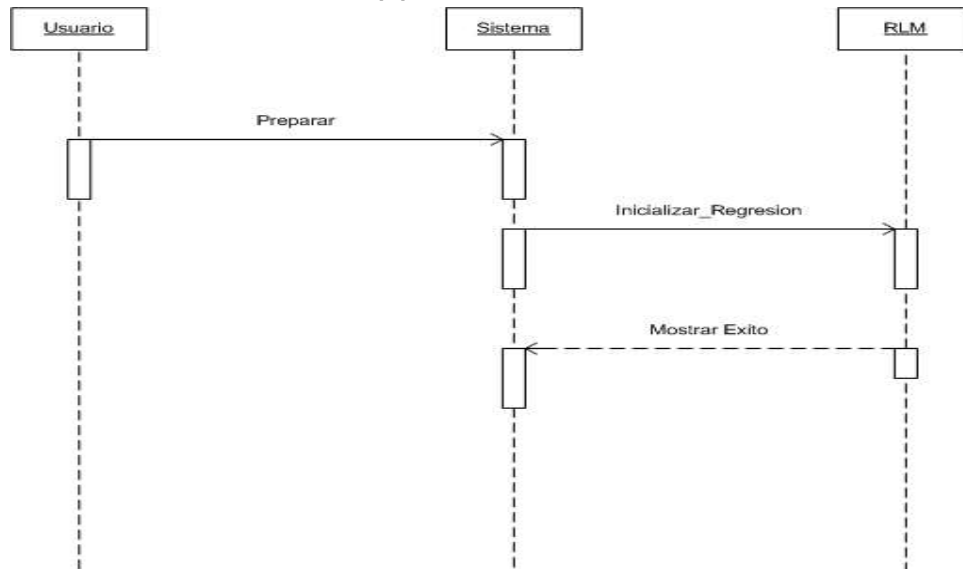


Figura #2.1.6.1: Diagrama de secuencia caso 2 (a).

2.1.6.2. CASO 2 (b): MOSTRAR RESULTADOS Y GRÁFICOS.

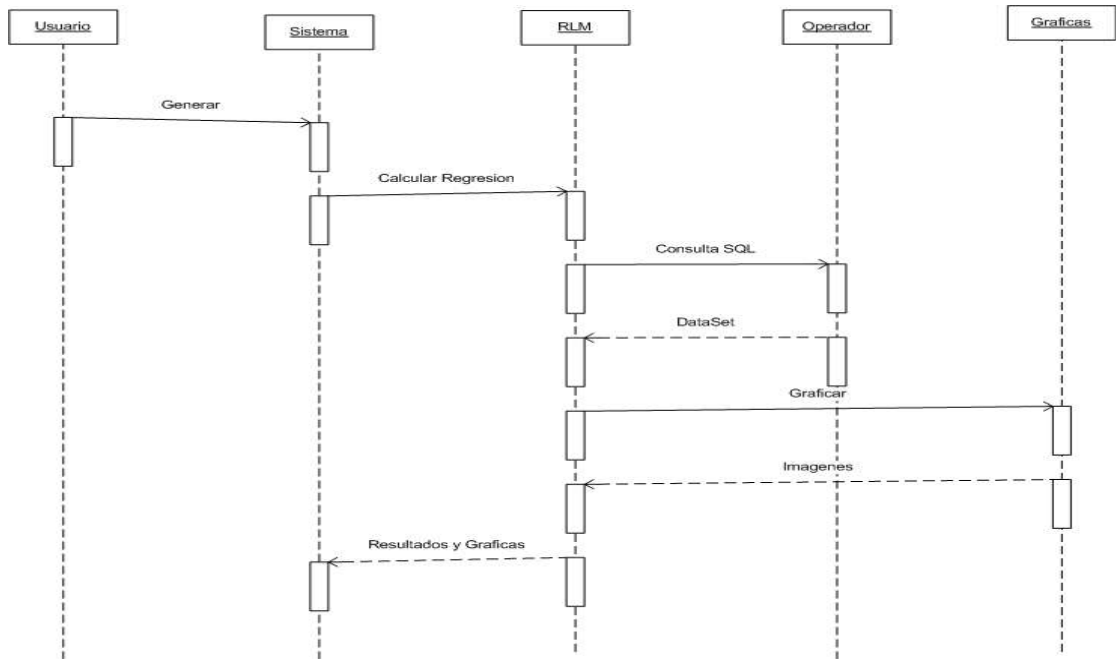


Figura #2.1.6.1: Diagrama de secuencia caso 2(b).

2.1.6.3. CASO 2 (c): EVALUAR REGRESIÓN.

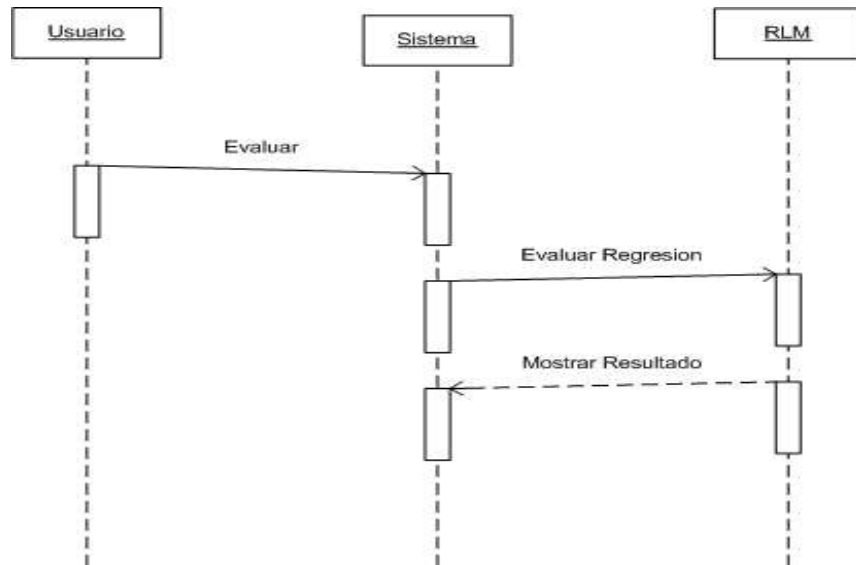


Figura #2.1.6.1: Diagrama de secuencia caso 2 (c).

2.1.7. PREDICCIÓN DE PERSONAL NAVAL.

2.1.7.1. CASO 3: Ingreso de los valores en los parámetros iniciales.

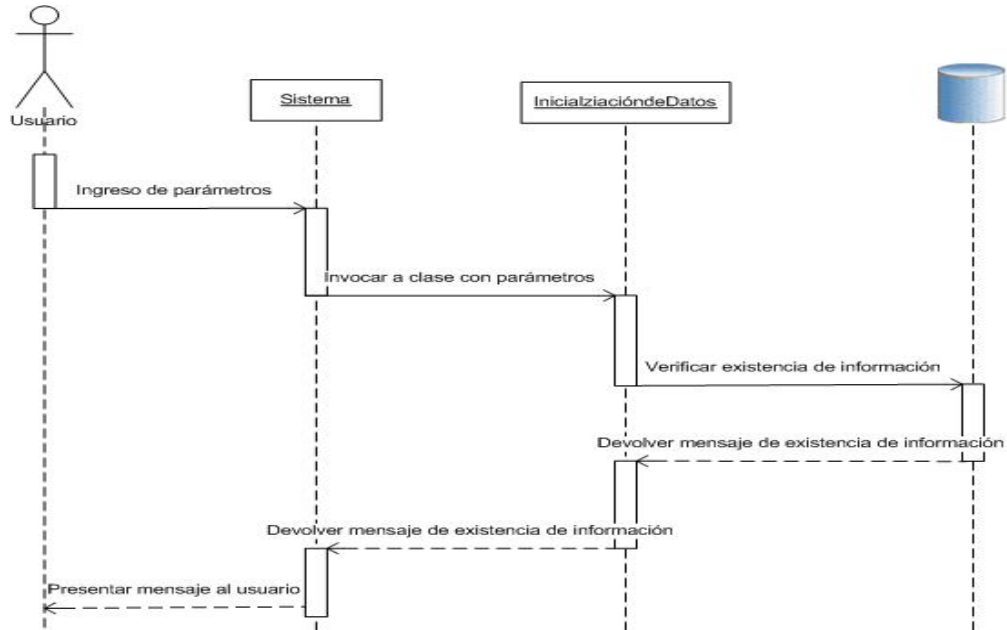


Figura #2.1.7.3: Diagrama de secuencia caso 3.

2.1.7.2. CASO 4: Preparación de los datos

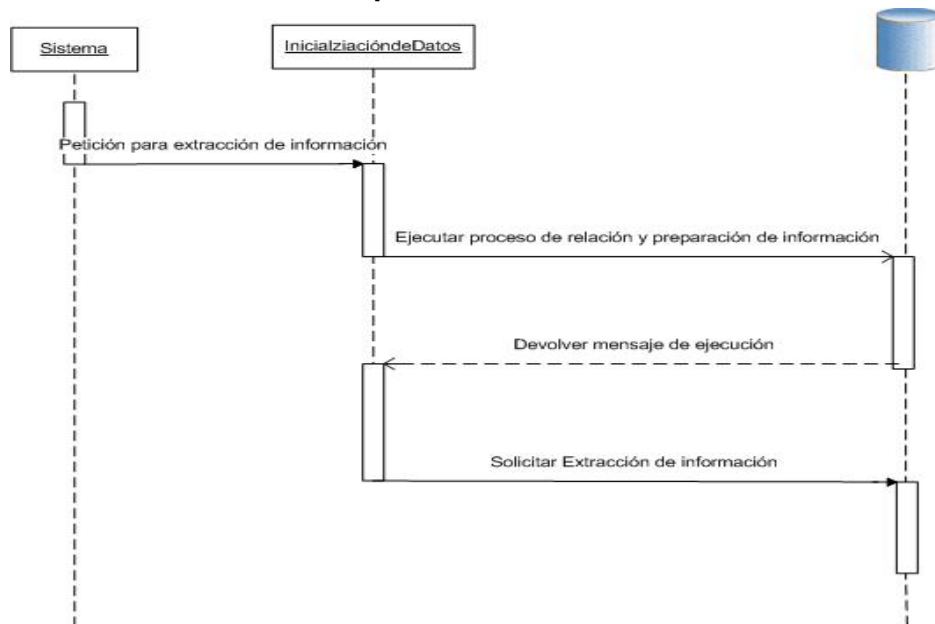


Figura #2.1.7.4: Diagrama de secuencia caso 4.

2.1.7.3. CASO 5: Obtención de los datos

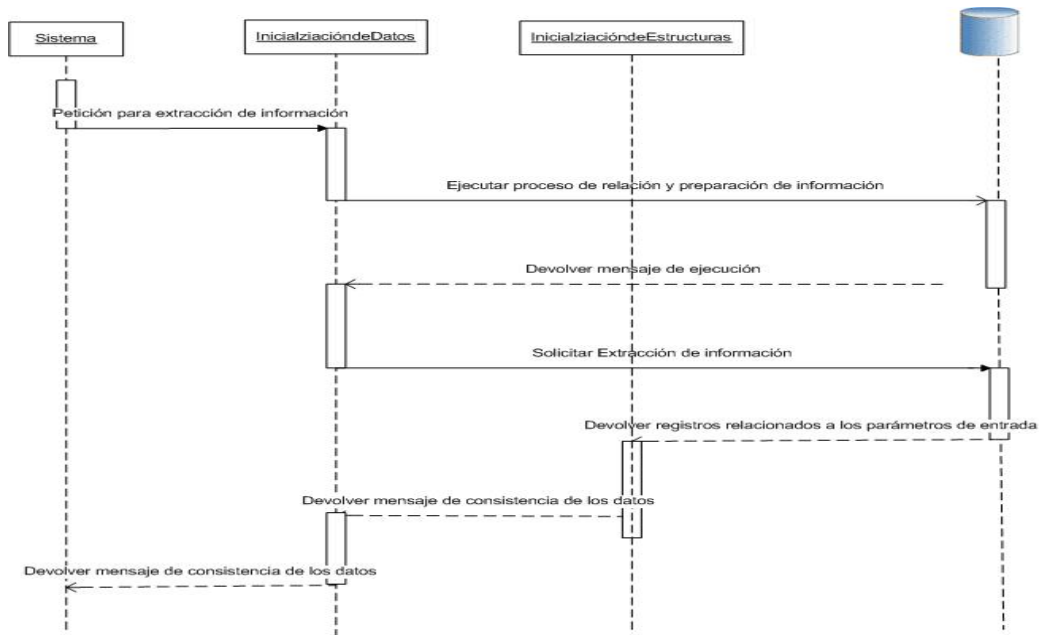


Figura #2.1.7.5: Diagrama de secuencia caso 5

2.1.7.4. CASO 6: Alimentación y construcción de la red bayesiana.

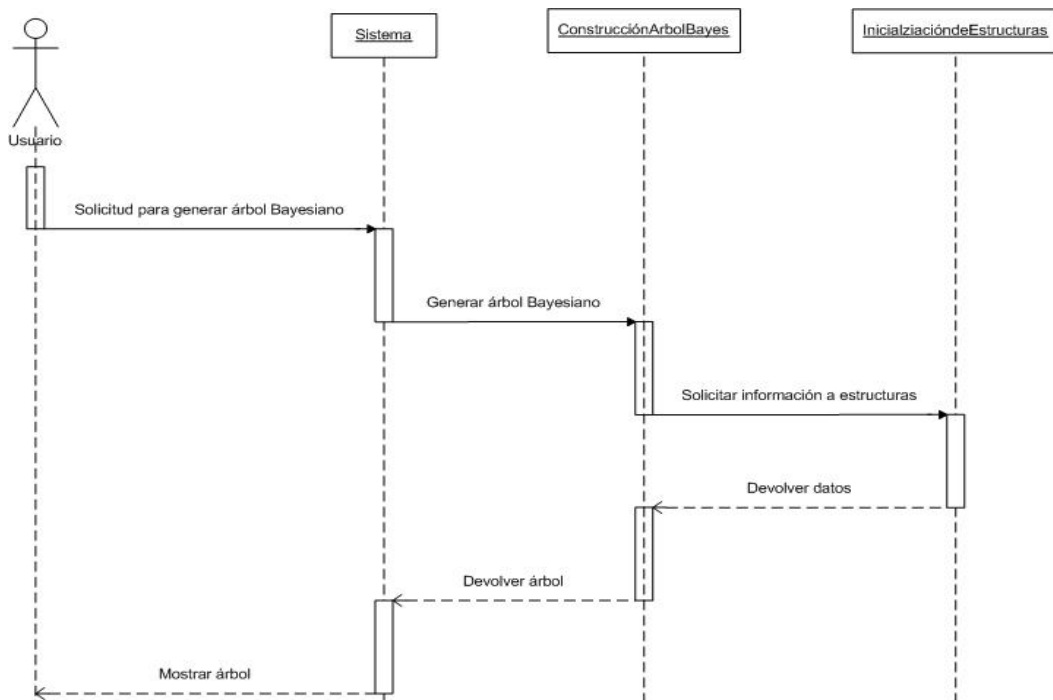


Figura #2.1.7.6: Diagrama de secuencia caso 6.

2.1.7.5. CASO 7: Pruebas y obtención de Resultados.

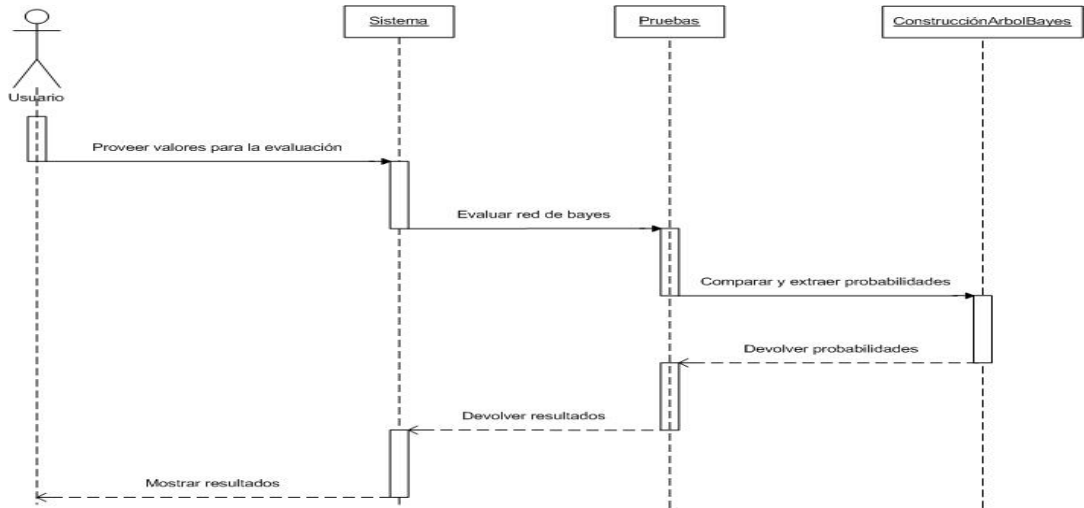


Figura #2.1.7.7: Diagrama de secuencia caso 7.

2.2. Algoritmos relacionados con la Solución del Problema

2.2.1. Modelo de Minería de Datos

La estimación de personas que van a estar dentro de un rango específico dentro de la Marina del Ecuador es posible predecirla por medio de Algoritmos de Minería de Datos enfocados dentro de la línea de modelos predictivos, las tareas que se usan en el proyecto son la de Regresión y la de Clasificación.

2.2.2. Tarea de Regresión

Asociada a la tarea de Regresión usamos el método de “Regresión Polinomial” que se encuentra catalogado como una técnica algebraica y estadística ya 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 cumplen el siguiente esquema matemático.

$$y = a_0 + a_1x^{(1)} + a_2x^{(2)} + a_3(x^{(1)})^2 + a_4(x^{(2)})^2$$

Figura #2.2.1(a): Fórmula para un polinomio de grado 4.

Para luego llegar a estimar una función matemática que será capaz de predecir valores a futuro de la siguiente manera.

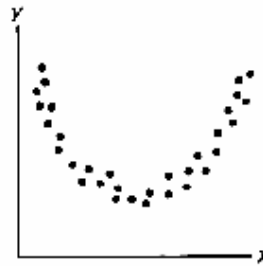


Figura #2.2.1 (b): Representación gráfica de dispersiones aplicando regresión.

2.2.3. Tarea de Clasificación y Estimación de Probabilidad

Las tareas de clasificación se encuentran asociadas 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 un conjunto de la población y el conjunto de dichas características nos dan una respuesta 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. El funcionamiento de dicha red es similar al siguiente esquema.

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 están reflejados en una red bayesiana.

2.3. Justificación de los Algoritmos Utilizados

2.3.1. Algoritmo de Regresión Polinomial

Dentro de la base de datos que utiliza la Marina del Ecuador se encuentran los datos históricos que muestran la cantidad de personal que ha ascendido de un rango a otro desde 1980 hasta el año 2002, a partir de estos datos se construye un modelo predictivo

en base a una Regresión polinomial que estima una función algebraica que modela con la mayor precisión posible el comportamiento de los datos a 5 años, si se realizara una estimación a un mayor número de años volvemos a la predicción en poco confiable puesto que el modelo usado no nos permite poseer datos fidedignos a largos plazos. Como la Marina del Ecuador necesita conocer cuáles serán las proyecciones de ascenso máximo a 5 años entonces puede fácilmente utilizar nuestro modelo para conocer dicha cifra y realizar su planificación.

También es conveniente indicar que la regresión introduce errores en la estimación, estos errores son cuantificados dentro del proyecto con la evaluación de los residuos los que son presentados a manera de gráficas que muestran que una nube de puntos dispersos y por medio de las áreas de mayor concentración es posible identificar las zonas de problemas en la predicción.

Por ejemplo, el número de ascensos dentro de un rango dado en la Marina del Ecuador para diferentes años se los puede apreciar en la siguiente tabla:

Año	Ascensos	Año	Ascensos
1960	1	1969	16
1962	10	1971	5
1972	10	1978	67
1979	59	1986	63
1980	45	1994	27
1989	32	1995	14
1993	17	1998	69
1997	66	2001	27
1999	42		

Tabla 2.3.1: Lista de valores que muestran la relación de los ascensos con los años para un grado en particular.

Al ser aplicado el algoritmo de regresión polinomial a estos datos vamos a obtener una gráfica similar a la siguiente:

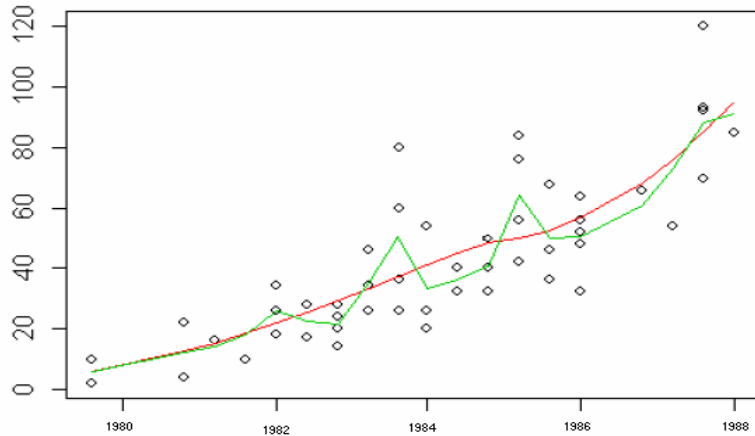


Figura #2.3.1: Representación gráfica de una regresión lineal.

Donde cada una de las observaciones corresponde a los ascensos de personal naval en los diferentes años de su promoción por cada rango, por medio de los algoritmos implementados en el proyecto generamos la función polinomial que cambia en el tiempo para cada rango tomando en cuenta los datos particulares de cada uno de ellos.

Del mismo modo se obtiene un gráfico de los residuos causados por los errores propiciados por el uso del método de regresión.

2.3.2. Algoritmo de Redes Bayesianas

Las redes bayesianas son un método práctico para realizar inferencias a partir de los datos, induciendo modelos probabilísticos que después serán usados para razonar o formular hipótesis sobre nuevos valores observados. Además, permiten calcular de forma explícita la probabilidad asociada a cada una de las hipótesis posibles, esto constituye una gran ventaja sobre otras técnicas.

El teorema de Bayes expresa lo siguiente: “Dadas dos Variables X e Y, tales que $P(x) > 0$ para todo x y $P(y) > 0$ para todo y, se cumple:

$$P(x | y) = \frac{P(x) \cdot P(y | x)}{\sum_{x'} P(x') \cdot P(y | x')}$$

Figura #2.3.2: Fórmula de la probabilidad de Bayes.

En nuestro proyecto tenemos un objetivo bien definido al cual le aplicamos el teorema de Bayes por medio de la construcción de la red Bayesiana, dicho objetivo es determinar si un miembro de la Marina del Ecuador ascenderá o no, lo que al final obtenemos es una red Bayesiana que se ve entrenada con datos históricos provenientes de las variables puntuales a en las que se fundamenta la institución para decidir si un elemento asciende o no.

Una vez entrenada la red Bayesiana se encuentra lista para predecir en base al cálculo de probabilidades.

Al final la red se ve alimentada con los nuevos datos que pasan a formar parte de su estructura y son tomados en cuenta para la próxima predicción.

2.4. Diseño Multidimensional de la Base de Datos.

Para realizar una tarea de Minería de Datos realmente efectiva y válida es necesario llegar a crear un almacén de datos o Data Warehouse sobre el cual es posible realizar transacciones OLAP y generar conocimiento.

El almacén de datos se forma a partir de bases de datos relacionales y de fuentes de datos externas que incluyen información del ambiente que es ajena a las bases transaccionales.

La fuente de datos del proyecto de análisis de ascensos de la Marina del Ecuador se nutre para realizar su almacén de datos de la base transaccional principal de manejo de personal, esta base presenta gran cantidad de información no muy organizada haciendo muy difícil practicar técnicas de minería directamente sobre ella, por lo cual optamos por crear un almacén de datos sobre ella.

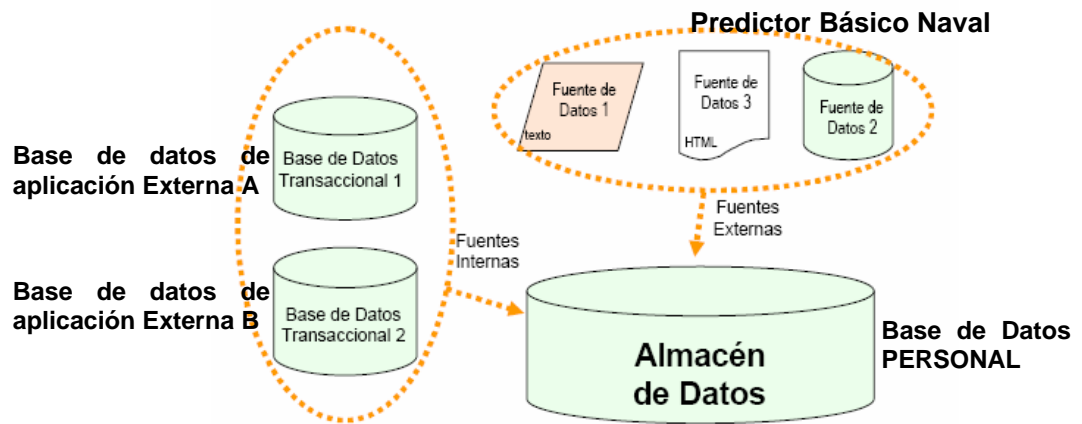


Figura #2.4 (a): Representación gráfica de la relación de la base de dato con el mundo exterior.

Dentro de la organización de los datos, se ha organizado el almacén de datos en torno a los hechos que a su vez tienen atributos o medidas que se ven en mayor o menor detalle según ciertas dimensiones. Nuestro objetivo de estimación son los ascensos y sobre ellos existen varias medidas importantes como son: cantidad de ascendidos por año, rangos con mayor población, etc. estas medidas se pueden detallar o agregar en varias dimensiones como cursos y calificaciones, disciplina, antigüedad en el rango, unidades asignadas, condecoraciones ganadas, etc.

De esta manera llegamos a estructurar un modelo que nos permite de una manera sencilla obtener información sobre hechos a diferentes

niveles de agregación y podemos aplicar métodos de minería de datos a partir de los datos recuperados del almacén que posee datos estructurados y limpios que son fáciles de modelar y utilizar.

El modelo multidimensional del almacén de datos construido para el proyecto Ascensos Navales es el siguiente:

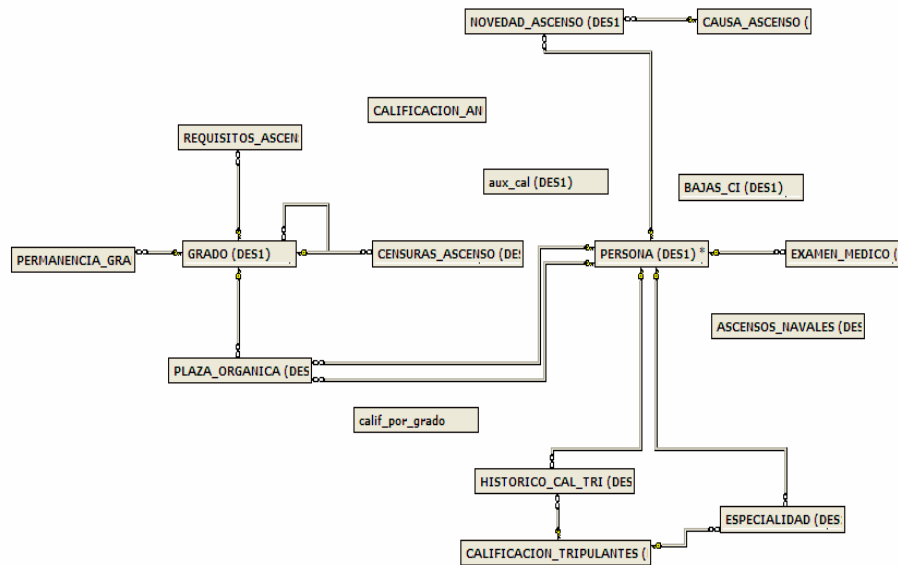


Figura #2.4 (b): Modelo multidimensional (los atributos se mencionan en el capítulo 3).

En una base de datos multidimensional la información se almacena de forma dimensional y no relacional. Las dimensiones determinan la estructura de la información almacenada y definen adicionalmente caminos de consolidación. La información almacenada se presenta como variables que, a su vez están caracterizadas por una o mas dimensiones. De esta manera la información puede analizarse dentro de un “cubo” formado por la intercepción de cada una de las dimensiones de la variable particular. Es decir podemos analizar la variable “Puntaje de ascenso” y podemos determinar patrones al revisar los valores históricos que contiene dicha variable. A partir de la información de tendencias que obtenemos al analizar el

comportamiento de la variable la ingresamos en una red Bayesiana y estimamos la probabilidad de que el resultado para un individuo en particular sea éxito o fracaso, para la dimensión de esa variable.

También las dimensiones pueden estructurarse jerárquicamente, de modo que se pueden construir caminos que permitan analizar la información de lo más general hacia lo específico, es decir que a partir de las calificaciones de ascensos con la que cuenta cada miembro de la Marina del Ecuador podemos inferir conocimiento de el comportamiento en otras, como pueden ser el aprovechamiento académico del militar, los resultados de los análisis físicos, las sanciones disciplinarias, los méritos o condecoraciones, etc. Todo este conjunto de variables se complementan para llegar a predecir si un determinado miembro asciende o no.

2.5. Análisis Discriminante.

El análisis discriminante ayuda a comprender las diferencias entre grupos. Explica en función de características métricas observadas, porque los objetos/sujetos se encuentran asociados a distintos niveles de un factor. En nuestro proyecto cada uno de las etapas del análisis discriminante se las puede describir de la siguiente manera.

2.5.1. Objetivo.

Deseamos predecir si un individuo ascenderá o no de un rango inferior a uno superior teniendo como base las variables que se especifican en los reglamentos de la Marina del Ecuador como requisitos fundamentales.

Se explota el fichero histórico de calificaciones y evaluaciones de los marinos, se analizan los reglamentos asociados a los procesos de ascensos, se observan las variables cuantitativas potencialmente explicativas:

2.5.2.Resultado.

Al aplicar el modelo discriminante a los datos históricos podremos ser capaces de anticipar la probabilidad de que un individuo ascienda o no a un rango superior.

2.5.3.Etapas del análisis discriminante.

2.5.3.1. Selección de Variables Dependientes e Independientes

Para realizar el análisis discriminante a los datos procedentes de la base de datos histórica de la Marina del Ecuador debemos primero identificar la variable dependiente que en nuestro caso es de tipo binario, el marino asciende o no asciende. Las otras variables tomadas en cuenta por medio del análisis de las leyes y reglamentos de la Marina del Ecuador son tomadas de manera que sean mutuamente excluyentes y además cumplen las siguientes características:

- No son excesivas, son las necesarias para realizar el procedimiento de minería de datos.
- Deben atender siempre al objetivo conceptual.
- Pueden someterse a un test univariante de diferencia de medias o un test ANOVA.

2.5.3.2. Selección del tamaño muestral.

El tamaño muestral se encuentra dentro de los parámetros recomendados para realizar una eficiente tarea de minería de datos, esto aproximadamente 15 a 20 observaciones por variable.

2.6. FLUJO DE VENTANAS.

En esta sección se explica el manejo del sistema con un conjunto de gráficos que representan los caminos a seguir para el uso de las herramientas desarrolladas.

2.6.1. INICIO DEL SISTEMA.

2.6.1.1. INICIO



Figura #2.6.1.1 Pantalla de ingreso al sistema.

Como se puede observar en el gráfico para acceder al sistema se requiere de un usuario y una contraseña que será autenticada en la base de datos.

2.6.1.2. INICIO (USUARIO O CONTRASEÑA INCORRECTOS).



Figura #2.6.1.2 Mal inicio de sesión.

En esta gráfica un usuario ha querido acceder al sistema, como ha proporcionado datos que no existen en la base de datos, se lo rechaza, como se muestra en el mensaje.

2.6.2. MENU PRINCIPAL DEL SISTEMA.

2.6.2.1. MENU PRINCIPAL

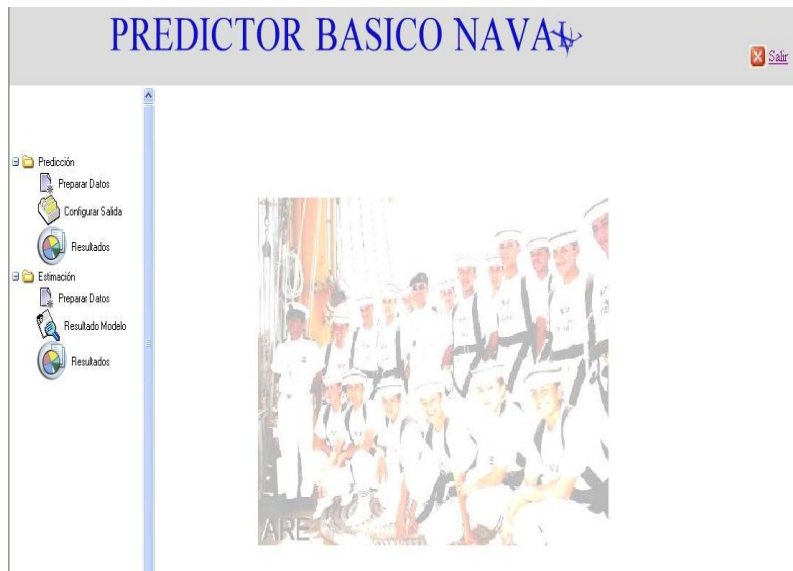


Figura #2.6.2.1: Menú del sistema.

En esta parte se observa el menú de navegación que contiene los tres grandes grupos de interacción: Predicción y Estimación.

2.6.3. MENU PREDICION.

2.6.3.1. PREPARACION DE LOS DATOS.

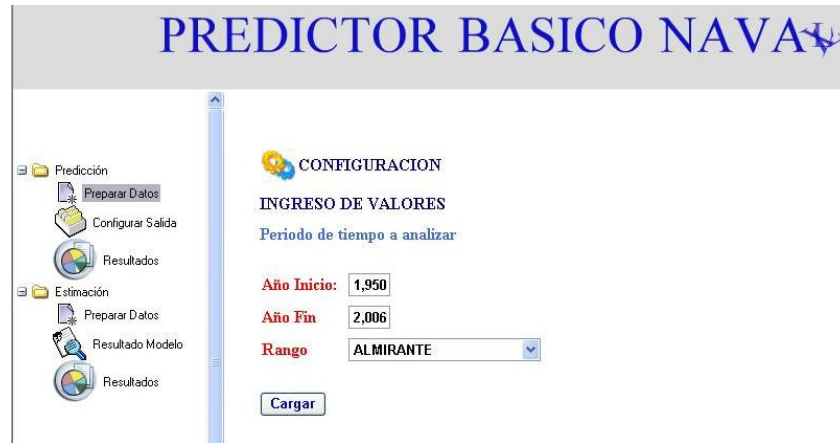


Figura #2.6.3.1: Menú predicción, configuración.

A través de este menú se procede a configurar los valores con los cuales se obtendrá la población de datos a estudiar.

2.6.3.2. PREPARACION DE LOS DATOS CON VALORES INCORRECTOS.



Figura #2.6.3.2: Menú predicción, configuración con valores erróneos.

En esta parte se ha intentado iniciar el proceso con rango de años inválidos por lo que inmediatamente se muestra un mensaje de error.

2.6.3.3. SIN INFORMACION QUE PROCESAR.



Figura #2.6.3.3: Menú predicción, no hay información que procesar.

Como podemos apreciar, este error nos indica que los datos ingresados como parámetros no son los correctos ya que no generan información para continuar con el proceso normal.

2.6.3.4. PROCESO EXITOSO.



Figura #2.6.3.4: Menú predicción, si hay información a trabajar.

Este ingreso como vemos ha generado información que sirve para continuar con el proceso.

2.6.3.5. MENU SALIDA (ARBOL INICIAL).



Figura #2.6.3.5: menú predicción, generación del árbol bayesiano.

En esta parte se mostrará el árbol de Bayes como producto de la población de datos obtenida a través de los valores de entrada. Inicialmente se muestra un árbol sin información debido a que no se ha presionado el botón generar.

2.6.3.6. MENU SALIDA (GENERACION DEL ARBOL).



Figura #2.6.3.6: Menú predicción, resultado de la generación del árbol bayesiano.

Como vemos este es el resultado que ha provocado la información obtenida como producto de los parámetros de ingreso. Cada Nodo verde indica el atributo o campo de análisis,

mientras que las hojas indican probabilidades de ocurrencia según sea el caso.

2.6.3.7. MENU SALIDA (GENERACION DEL ARBOL).



Figura #2.6.3.7: Menú predicción, intento de ejecutar el proceso de generación sin información.

Como se aprecia se ha querido intentar generar el árbol sin haber ingresado los parámetros y haber generado la información con la que se debe de construir el árbol.

2.6.3.8. MENU RESULTADO (PRUEBAS).

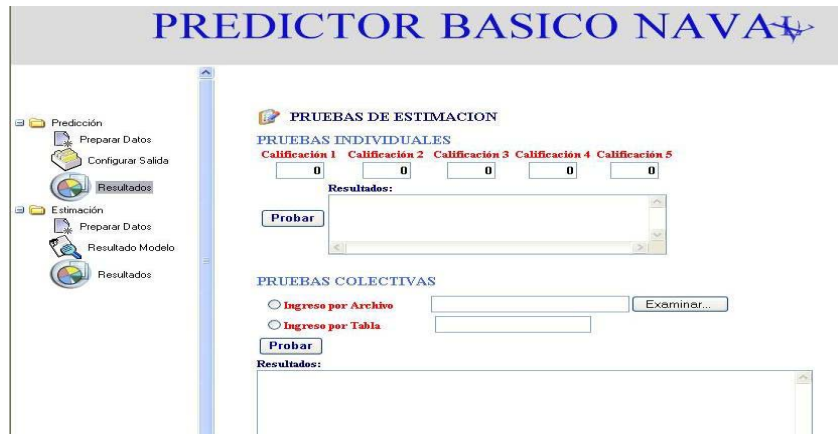


Figura #2.6.3.8: Menú predicción, sección de evaluación.

Como se aprecia en el gráfico se procederá a hacer pruebas individuales o colectivas según sea el caso, reportando resultados

de valores que contemplarán posibilidades de ascenso en cada caso.

2.6.3.9. MENU RESULTADO (PRUEBAS INDIVIDUALES).

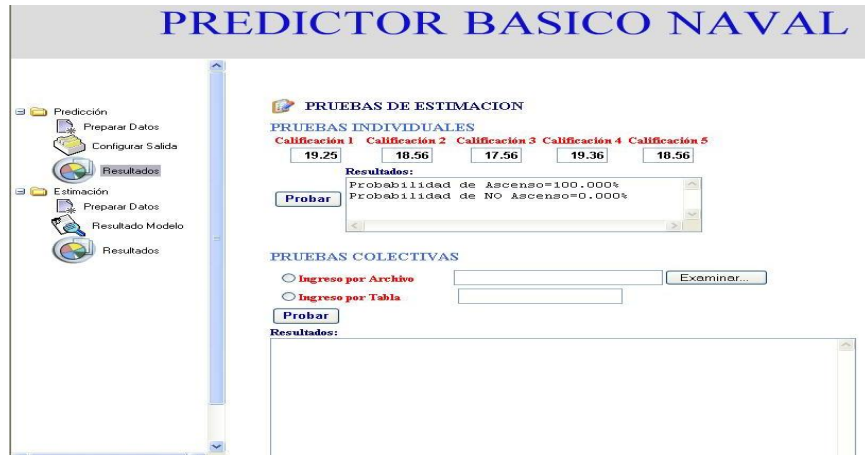


Figura #2.6.3.9: Menú predicción, prueba individual.

En esta parte se ingresaron valores en el área de pruebas individuales arrojando resultados que son producto de la evaluación del árbol bayesiano.

2.6.3.10. MENU RESULTADO (PRUEBAS COLECTIVAS POR ARCHIVO).

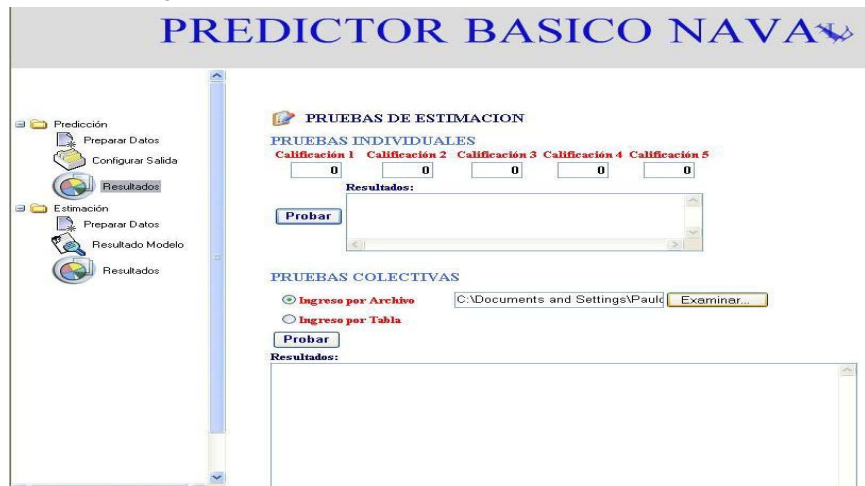


Figura #2.6.3.10: Menú predicción, pruebas masivas.

Como vemos en la sección de pruebas colectivas se pueden hacer pruebas de dos formas por archivo o por tablas (bases de

datos). En esta situación se ha escogido por archivo y se oculta el cajón de texto que admite el nombre de la tabla.

2.6.3.11. MENU RESULTADO (PRUEBAS COLECTIVAS POR ARCHIVO, RESULTADOS).

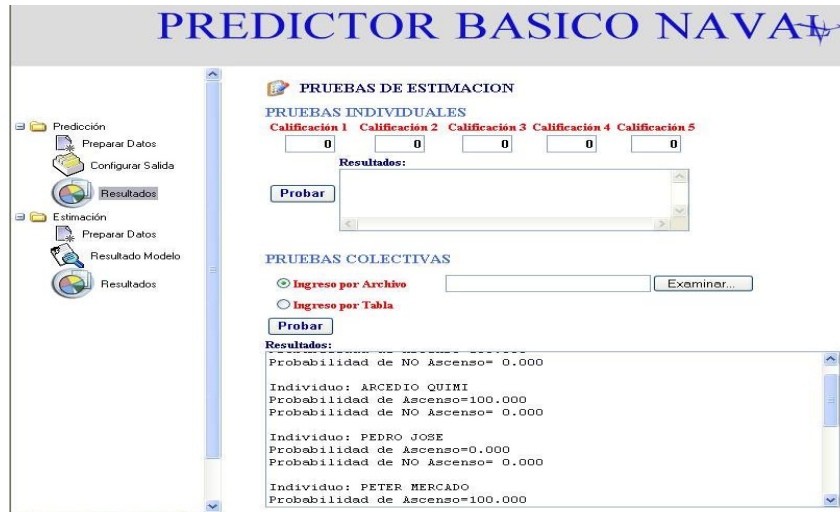


Figura #2.6.3.11: Menú predicción, resultados pruebas colectivas por archivos.

Como se aprecia se puede ver los resultados de haber ingresado un conjunto de registros que han sido extraídos de un archivo proporcionado por el usuario.

2.6.3.12. MENU RESULTADO (PRUEBAS COLECTIVAS POR TABLA).



Figura #2.6.3.12: Menú predicción, pruebas masivas por ingreso de tabla.

Podemos ver que en esta parte se ha seleccionado el ingreso de pruebas masivas por tabla dejando oculto el ingreso por archivo. En esta parte se debe de proveer un nombre de tabla que exista en la base de datos con información, caso contrario se mostrará un mensaje de error.

2.6.3.13. MENU RESULTADO (PRUEBAS COLECTIVAS POR TABLA, RESULTADOS).



Figura #2.6.3.13: Menú predicción, pruebas masivas resultado del ingreso por tabla.

Como se podemos ver, éste ha sido el resultado de haber ingresado la tabla donde se encuentran registros con calificaciones que deben de ser analizadas por la red bayesiana.

2.6.4. MENU ESTIMACION.

2.6.4.1. MENU PREPARAR DATOS.



Figura #2.6.4.1: Menú estimación.

Como se ve en esta parte lo que se espera es un rango de tiempo y un rango para hacer el análisis, para así poder seleccionar el conjunto de registros con el cual trabajará la regresión.

2.6.4.2. MENU PREPARAR DATOS (PREPARACIÓN DE DATOS EXITOSO).



Figura #2.6.4.2: Menú estimación, configuración exitosa.

Se aprecia en la gráfica que el ingreso del rango de años más el rango ha permitido inicializar los valores de trabajo correctamente.

2.6.4.3. MENU PREPARAR DATOS (PARAMETROS INCORRECTOS).

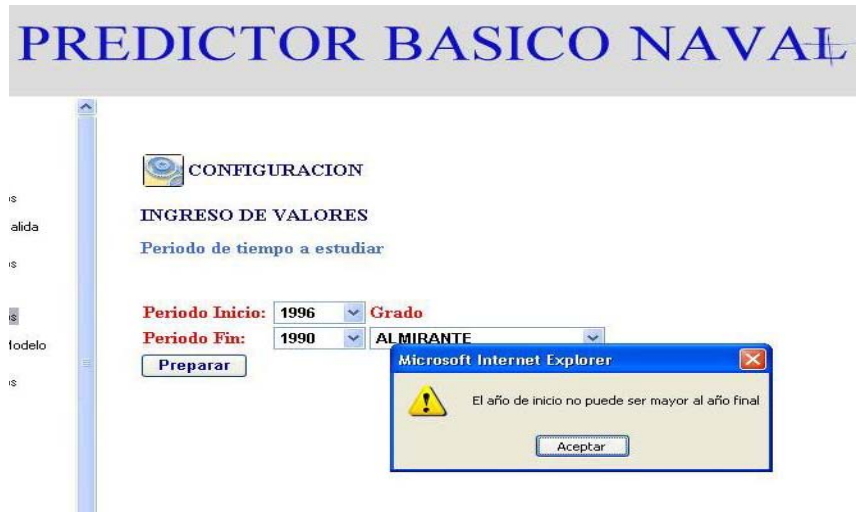


Figura #2.6.4.3: Menú predicción, configuración con valores erróneos.

Como vemos se ha intentado preparar la regresión con un rango de años erróneo, por lo que se muestra el respectivo mensaje de error.

2.6.4.4. MENU RESULTADO MODELO.

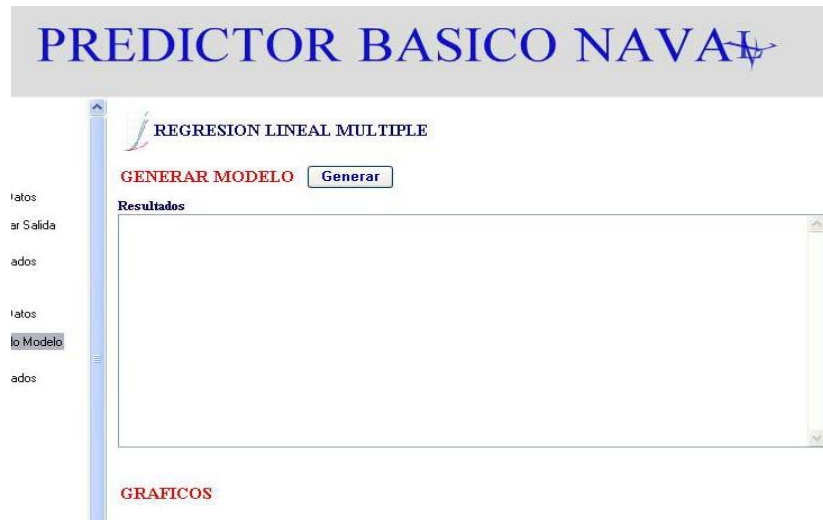


Figura #2.6.4.4: Menú estimación, ventana de generación del modelo de regresión.

En esta pantalla se mostrará el resultado de proceso de regresión, cuyos valores servirán a un especialista en tratamiento de información estadística corroborar los resultados.

2.6.4.5. MENU RESULTADO MODELO (RESULTADOS Y GRAFICOS).

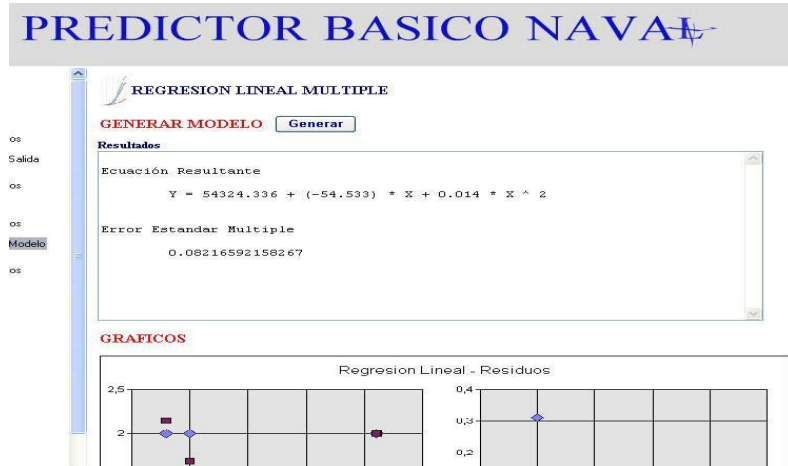


Figura #2.6.4.5: Menú predicción, resultados del modelo de regresión.

Como vemos una vez que se han inicializado los valores de los parámetros se procede a ejecutar el cálculo que genera el modelo mostrando en esta parte la fórmula de regresión y los gráficos asociados a las mismas.

2.6.4.6. MENU RESULTADOS (GENERACIÓN NO VÁLIDA).

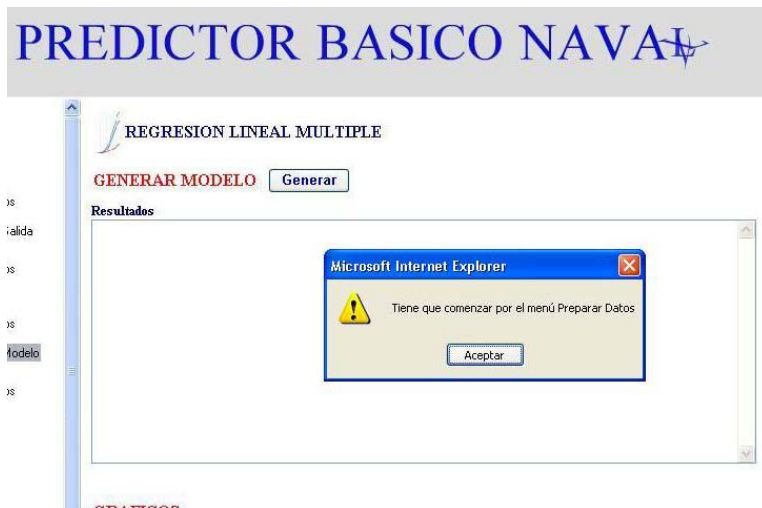


Figura #2.6.4.6: Menú predicción, intento de generación del modelo.

No se puede iniciar la regresión si no se ha inicializado los datos en el menú anterior. El mensaje de error es producto de este intento.

2.6.4.7. MENU RESULTADOS (PRUEBAS DEL MODELO DE REGRESIÓN).



Figura #2.6.4.7: Menú predicción, resultado del cálculo del modelo de regresión.

Como se aprecia en esta ventana, se ha iniciado la evaluación con un año ingresado, mostrando el resultado respectivo a la evaluación.

2.6.4.8. MENU RESULTADOS (PRUEBA NO VÁLIDA).



Figura #2.6.4.8: Menú predicción, configuración con valores erróneos.

Como se ve se ha intentado hacer una evaluación sin iniciar el proceso por el menú de datos, el mensaje de error muestra esta acción fallida. Igualmente ocurrirá si no se genera el modelo aunque si se hayan inicializado los datos.

3. CIRCULO VIRTUOSO DE LA MINERÍA DE DATOS

3.1. ESTUDIO DEL COMPORTAMIENTO DE LA INFORMACIÓN EN BASE A MODELOS DE PREDICCIÓN.

3.1.1. INTRODUCCIÓN.

La minería de datos tiene como objetivo analizar los datos para extraer conocimiento. Este conocimiento puede ser en forma de relaciones, patrones o reglas inferidos de los datos y previamente desconocidos, o bien en forma de una descripción más concisa. Son estas relaciones las que constituyen el modelo de los datos analizados y de los modelos depende el tipo de técnica que puede usarse para inferirlos.

Los modelos dentro de la minería de datos pueden ser predictivos o descriptivos. Los modelos predictivos pretenden estimar valores futuros o desconocidos de variables de interés o variables objetivo o dependientes, a partir de otras variables o campos de la base de datos denominadas variables independientes o predictivas. En cambio los modelos descriptivos identifican patrones que explican o resumen los datos.

El presente trabajo se basa en modelos predictivos cuyos métodos más representativos son la clasificación y la regresión.

Los procesos de minería de datos conducen a la obtención de patrones y modelos que luego son evaluados e interpretados con el fin de inferir conocimiento. Para realizar este procedimiento es necesario realizar sobre los datos un proceso de selección, limpieza,

transformación y proyección de los datos. Básicamente se pasa la información por cinco fases como son:

- Fase de integración y recopilación de datos.
- Selección, limpieza y transformación
- Minería de datos
- Evaluación e interpretación
- Difusión y uso

A final lo que se obtiene es conocimiento útil que sirve para inferir conclusiones que conducen a tomar las mejores decisiones dentro de un modelo de negocios.

Durante la fase de integración se usa la tecnología “data warehousing” por medio de la cuál se usa el análisis de diversas fuentes de datos llegando a integrarlas y haciéndolas accesibles para aplicar minería de datos y luego tomar decisiones.

Esencialmente, los almacenes de datos se utilizan para poder agregar y cruzar eficientemente la información de manera sofisticada. Por ello, los datos se modelan con una estructura de base de datos multidimensional, donde cada dimensión corresponde a un atributo o conjunto de atributos en torno a hechos que almacenan el valor de alguna medida agregada.

Durante la fase de selección y preparación el subconjunto de datos que se va a minar, se convierten en lo que se conoce como vista minable. A partir de la base de datos proporcionada por la Armada del Ecuador se realizan procesos de Integración y de selección usando las herramientas proporcionadas por el motor de bases de

datos SQL Server™ para el proyecto y se llega a obtener la vista minable sobre la que se aplican métodos de minería de datos. En ciertos casos se presentan problemas por datos anómalos que no se ajustan al comportamiento general de la información, dichos datos se los puede considerar errores, otros pueden ser descartados al ser considerados ruido. Para manejar datos erróneos se distingue entre la detección y el tratamiento de los mismos. La detección depende de conocer el formato o los posibles valores de los campos. Los datos erróneos, más difíciles de detectar, son aquellos que sí se ajustan al formato.

La alta dimensionalidad se nos presenta como un gran problema para manejar el conocimiento que se infiere de los datos. Al tener muchas dimensiones (atributos) respecto a la cantidad de instancias o ejemplos, nos encontraremos con una situación poco deseable al existir tantos grados de libertad, por esto puede ocurrir que los patrones extraídos puedan ser caprichosos y poco robustos. Tendríamos entonces un espacio casi vacío y, por tanto, los patrones no tienen datos donde apoyarse a la hora de tomar una u otra forma.

Otro aspecto importante es la discretización o numerización de los datos puesto que ciertos algoritmos dentro de la minería de datos requieren de datos tratados y transformados a los requerimientos particulares de cada uno de ellos.

El Proyecto trabaja con una base de datos histórica en la que se encuentran registrados los datos de 5 años del personal de la Marina

del Ecuador, se aplican procesos de selección y limpieza como los anotados en párrafos anteriores y se procede a construir un modelo multidimensional que pasa a formar un “Almacén de Datos” sobre el cual se aplican los métodos de minería.

Básicamente se utilizan algoritmos correspondientes al modelo predictivo para dar solución a el problema planteado en el proyecto.

Los modelos predictivos tratan problemas y tareas en los que hay que determinar uno o más valores que luego será necesario presentar por medio de una salida que ayudará a inferir el conocimiento. En este modelo el conjunto de evidencias son correspondencias entre dos conjuntos de la siguiente manera:

$\delta: E \rightarrow S$, donde S es el conjunto de valores de salida. Los ejemplos, al ir acompañados de un valor de S, se denominan comúnmente ejemplos etiquetados y δ es un conjunto de datos etiquetado. El objetivo es aprender una función $\lambda: E \rightarrow S$ que represente la correspondencia existente en los ejemplos, es decir, para cada valor de E tenemos un único valor para S.

Para el caso de los problemas que resolvemos con nuestro sistema usamos:

- Estimación de personal naval para los próximos 5 años:
Regresión polinomial.
- Predicción por medio de clasificación probabilística: Redes Bayesianas.

3.1.2. Modelos Predictivos: Regresión.

En la primera parte de este proyecto usamos un modelo de regresión, es posible aplicarlo porque tanto la variable de respuesta como las variables explicativas son cuantitativas. Si solo se dispone de una variable explicativa estamos hablando de regresión simple, mientras que cuando disponemos de varias variables explicativas se trata de una regresión múltiple. Así mismo la regresión puede ser lineal o polinomial, la regresión lineal es demasiado imprecisa y está sujeta a errores grandes por esto optamos por utilizar una Regresión Polinomial. El proyecto se vale de las herramientas proporcionadas por el motor de base de datos *Microsoft Sql Server* para generar una tabla (vista minable) donde constan los valores de los ascensos de personal por rango diferenciadas por año.

Se construyó un procedimiento almacenado que recopila las cantidades de personal naval que han ascendido a partir de las tablas históricas que almacenan la información de ascensos navales en los diferentes años.

El proceso realizado para hacer la recopilación de datos se la puede resumir con el siguiente esquema:

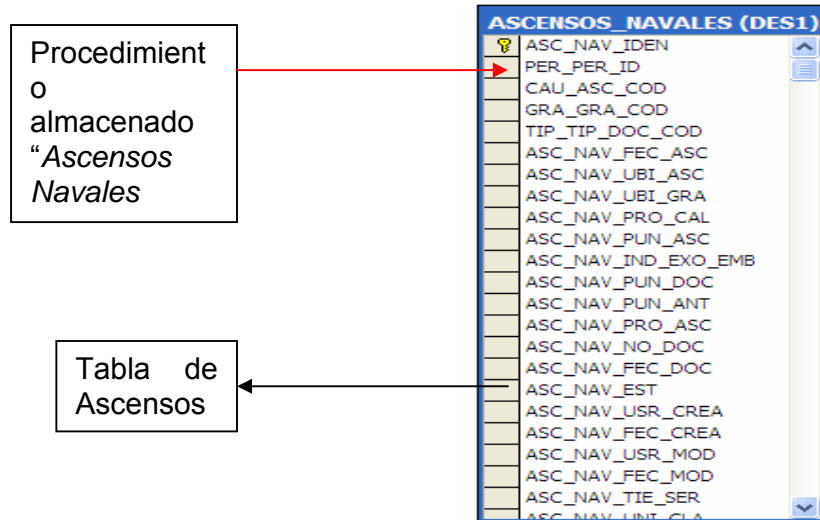


Figura #3.1.2 (a): Una vista de los procedimientos almacenados en la base de datos Personal.

La tabla generada muestra la cantidad de ascensos de personal para cada uno de los grados de la institución militar como se muestra en la siguiente tabla.

Grado/Año	002	004	006	008	010
1980	4	12	23	32	19
1981	6	13	24	23	3
1982	4	12	22	24	4
1983	7	14	23	2	31
1984	4	15	25	3	12
1985	2	23	22	34	8
1986	4	23	26	5	23
1987	6	14	23	4	37
1988	5	23	25	34	3
1989	4	12	23	23	7
1990	3	16	19	5	34
1991	5	18	23	5	39
1992	4	14	25	32	1
1993	5	12	27	23	3
1994	4	15	29	4	12
1995	3	17	30	6	56
1996	4	12	18	34	2
1997	5	18	23	32	3
1998	5	19	24	3	4
1999	6	17	27	3	12
2000	4	15	26	31	12

Tabla #3.1.2 (b): Tabla que almacena por año y por rango (código) los números de elementos activos.

A partir de esta tabla se aplican procedimientos de regresión polinomial consistente en establecer una función polinómica que

permita predecir la cantidad de personal en cierto rango para los siguientes 5 años.

El sistema construye un modelo matemático para estimar los parámetros “bi” desconocidos y establecer la curva representativa del rango de acuerdo al siguiente modelo matemático:

$$\hat{y} = b_0 + b_1x + b_2x^2 + \dots + b_r x^r .$$

Figura #3.1.2 (b): Fórmula polinómica de grado r.

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 siguiendo el siguiente esquema:

Para el caso de una sola variable independiente, el grado del polinomio de mejor ajuste se puede determina al graficar un diagrama de dispersión de los datos que se obtienen de un experimento que da n pares de observaciones de la forma $\{(x_i, y_i); i = 1, 2, \dots, n\}$ en nuestro caso Año vs. Número de ascensos por año. Matricialmente tenemos lo siguiente:

$$\begin{bmatrix} n & \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i^r \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i^3 & \sum_{i=1}^n x_i^{r+1} \\ \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i^3 & \sum_{i=1}^n x_i^4 & \sum_{i=1}^n x_i^{r+2} \\ \vdots & \vdots & \vdots & \vdots \\ \sum_{i=1}^n x_i^r & \sum_{i=1}^n x_i^{r+1} & \sum_{i=1}^n x_i^{r+2} & \sum_{i=1}^n x_i^{2r} \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ \vdots \\ b_r \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_i y_i \\ \vdots \\ \sum_{i=1}^n x_i^r y_i \end{bmatrix}$$

Figura #3.1.2 (c): Matriz de regresión polinomial.

Al resolver estas $r + 1$ ecuaciones, obtenemos las estimaciones b_0, b_1, \dots, b_r y por ello generamos la ecuación de predicción de regresión polinomial

$$\hat{y} = b_0 + b_1x + b_2x^2 + \dots + b_r x^r .$$

Figura #3.1.2 (c): fórmula resultante de una regresión polinomial.

Que representa un polinomio el cual al ser graficado nos permite estimar en un lapso máximo de 5 años (limite tolerante de error). Esta estimación la podemos apreciar por medio de las gráficas que realiza el sistema donde se puede apreciar las funciones polinómicas para cada rango.

3.1.3. Clasificación Bayesiana.

3.1.3.1. Fases de la minería de datos aplicadas al proyecto.

El proceso de extracción de conocimiento es iterativo y por lo tanto la salida de alguna de las fases puede hacer volver a pasos anteriores y a menudo son necesarias varias iteraciones para extraer conocimiento de alta calidad.

La extracción de conocimiento en nuestro proyecto se realiza de acuerdo a cinco fases representadas en la siguiente figura.

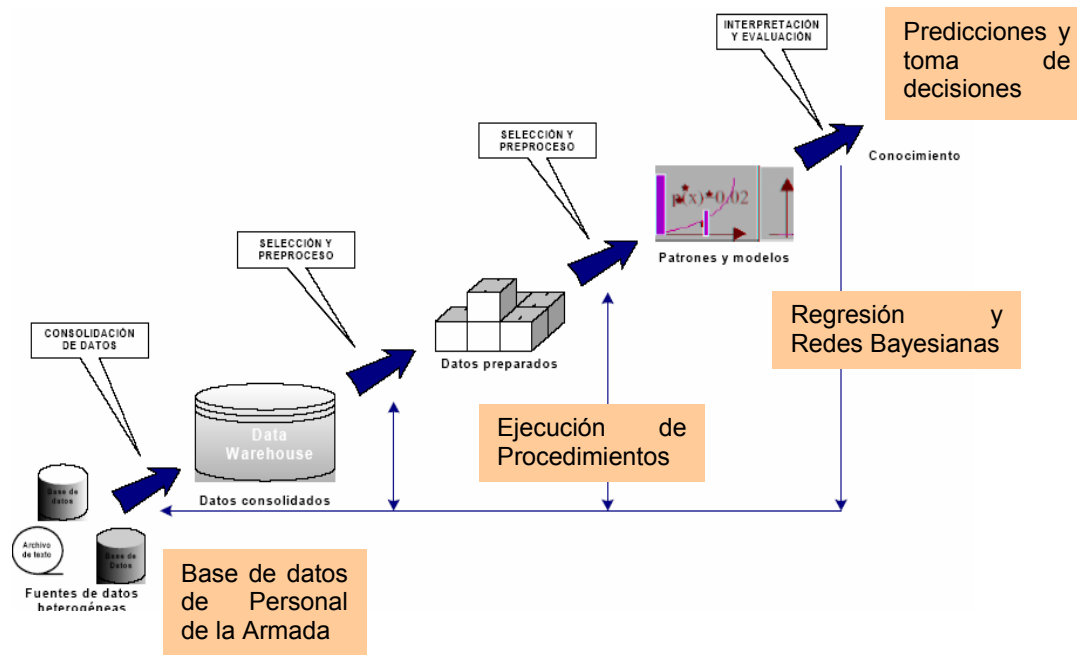


Figura #3.1.3 (c): Procesos a seguir dentro de la minería de datos.

3.1.3.2. Datos Iniciales.

Contamos para la elaboración del proyecto de la base de datos provista por la dirección de personal de la Marina del Ecuador (DIGPER) que cuenta con aproximadamente de 380 tablas, 1057 vistas y 215 procedimientos almacenados. Dado que los datos son información almacenada años anteriores en los que se han podido cometer errores que provoquen anomalías es necesario realizar una fase de selección, limpieza y transformación en la que se eliminan o corrigen los datos incorrectos decidiendo la estrategia a seguir para un mejor manejo de los datos.

Primeramente, con la base de datos proporcionada realizamos el proceso de integración y recopilación, para lo cual seguimos los siguientes pasos:

- Ubicación de las tablas que contienen los datos necesarios para poder discernir el personal militar que ascenderá encontrando las tablas que almacenan la información histórica de calificaciones, estos datos son básicos para lograr ascender en la Marina del Ecuador.
- Verificación de enlaces de datos válidos que conduzcan a mantener la información debidamente relacionada.

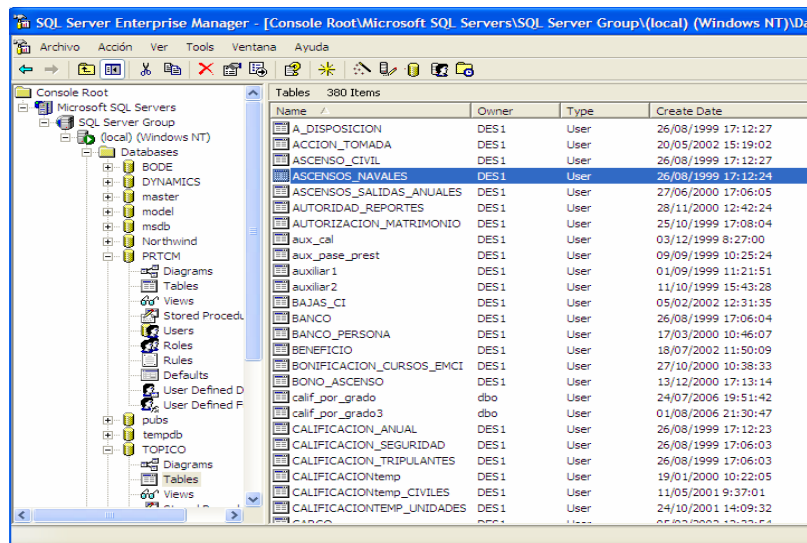


Figura #3.1.3.1 (a): Vista del explorador de SQL mostrando la tabla de evidencia histórica.

- Exploración de las vistas y procedimientos creados ya en la base de datos de la Marina del Ecuador para lograr obtener información recopilada a partir de la cual elaborar nuevas tablas con estructura idónea para la minería de datos.

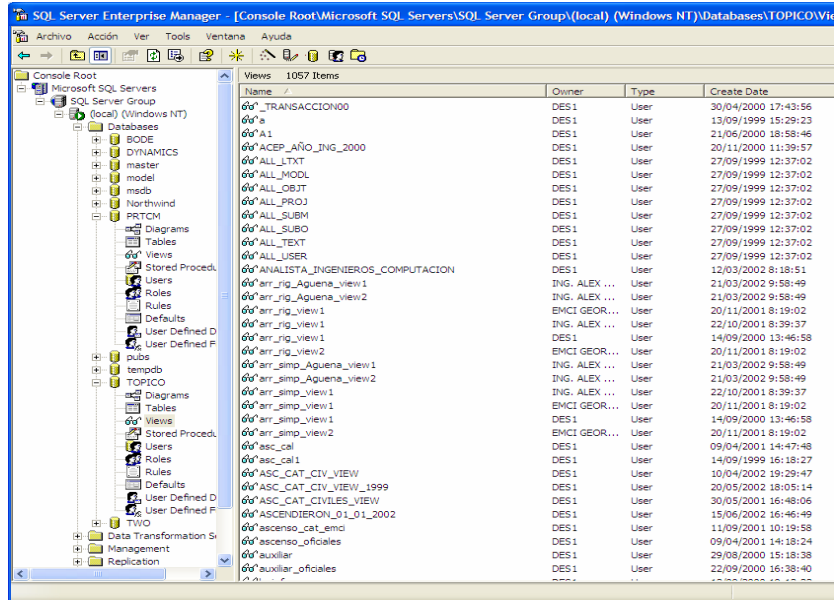


Figura #3.1.3.1 (b): Vista del explorador de SQL mostrando la vistas de la base de datos Personal.

- Uso del Lenguaje de consultas estructuradas (SQL) para la creación de procedimientos almacenados que construyen nuevas tablas donde se recopila la información detallada en los reglamentos y que aglutinan y describen los requisitos de ascenso de los militares.

3.1.3.3. Tablas Seleccionadas:

En esta primera etapa seleccionamos las siguientes tablas:

- **Tabla Persona:**

Almacena información referente al militar, esta tabla contiene gran cantidad de relaciones con otras tablas de la base.

Campos útiles recopilados de esta tabla son:

- El Grado actual del militar.
- Fecha de ingreso a la institución.
- Ubicación en el grado dentro de la promoción de cada persona.

- Indicador de baja de la institución.
- Promoción a la que pertenece la persona*.
- Especialidad del militar.
- Ubicación para el ascenso.

- **Tabla Calificación Anual:**

Almacena información referente a las calificaciones anuales de los miembros de la Marina del Ecuador, estas calificaciones son primordiales para discriminar entre el individuo que tiene posibilidades de ascender y cuales no.

Campos útiles recopilados en esta tabla son:

- Identificador de la unidad administrativa a la que estuvo asignado el militar al momento de obtener la calificación.
- Fecha de la calificación.
- Las calificaciones anuales en si, obtenidas por el individuo desde que ingresó a la institución hasta el presente en que se encuentra postulando por ascender a un rango superior.

- **Tabla Lista Calificación:**

Almacena el significado de cada una de las nomenclaturas de calificaciones que usa la Armada para manejar los resultados de las evaluaciones. Catalogar una calificación (“Exonerado” o “Lista 1”), es un ejemplo de los datos guardados en esta tabla:

* Cada miembro de la marina posee una promoción que es similar al año de ingreso de cada uno

De:	Hasta:	Equivalencia
19.000	20.000	LISTA 1
17.000	18.999	LISTA 2
14.000	16.999	LISTA 3
12.000	13.999	LISTA 4
00.000	11.999	LISTA 5

Tabla #3.1.3.1: Lista de calificaciones.

- **Tabla unidad Administrativa:**

Almacena información relativa a las unidades administrativas dentro de la Armada, esta información es vital al momento de decidir un ascenso ya que uno de los requisitos para el ascenso es prestar servicio por ciertos lapsos de tiempo en las unidades administrativas correspondientes de acuerdo al rango. Campos útiles recopilados en esta tabla son:

- Identificador de la unidad administrativa.
- Tipo de la unidad.
- Unidad administrativa siguiente.
- Descripción de la unidad.
- Rangos asociados a las unidades administrativas.

- **Tabla Ascensos Navales:**

Almacena información histórica correspondiente a todos los ascensos que han ocurrido en la institución para todos los rangos, sean de personal de tropa como de oficiales, esta tabla es vital y dentro de esta primera etapa de Integración y recopilación es la tabla que mayor información nos aporta al

almacén de datos que se construye. Campos útiles recopilados en esta tabla son:

- Grado desde el cual se realizó el ascenso.
- Fecha en que el militar ascendió, es decir, recopilamos todas las fechas en que ascendió un individuo según que tan alto ha llegado en el escalafón naval.
- La ubicación en el ascenso, que nos indica en que posiciones ha quedado un cierto individuo cada vez que ascendió.
- Las ubicaciones en el grado que ha tenido un cierto individuo.
- El puntaje de calificaciones anuales del individuo.
- El puntaje total de ascenso obtenido por el individuo en cada uno de sus ascensos.
- La calificación de la documentación e historial del marino en cada uno de sus ascensos.
- El puntaje de antigüedad obtenido en sus diversos ascensos.

- **Tabla Grado:**

Almacena los detalles particulares correspondientes a cada grado. Campos útiles recopilados en esta tabla son:

- Código del grado.
- Descripción del grado
- Grado siguiente.

- Promedio pedido por la institución para cada uno de los grados de acuerdo a los reglamentos internos.

3.1.3.4. Selección, Limpieza y Transformación.

Siguiendo el esquema de minería de datos, el siguiente paso es la selección, limpieza y transformación de datos. Esta fase ayuda a depurar la información recopilada en la fase anterior siendo más selectivo al momento de recuperar los parámetros que definen y deciden el ascenso, eliminando ruido, discretizando datos, estableciendo un estándar para manejar la información.

Para realizar este procedimiento realizamos varios Procedimientos almacenados compilados en la base de datos que nos recuperan la información de las distintas calificaciones por año, agrupadas por rango, esto nos permite hacer un análisis de patrones al poder observar como se han ido definiendo las puntuaciones y cuales son las variables más preponderantes.

Para terminar el proyecto definimos básicamente 5 variables fundamentales las cuales se usan para definir si una persona asciende o no. Estas son:

- Promedio de la calificación anual
- Nota de grado
- Promedio de méritos
- Promedio de deméritos

- Promedio de antigüedad. *

El procedimiento almacenado “*calif_personal*” realiza una consulta a las tablas nombradas anteriormente buscando los diferentes parámetros enunciados en los reglamentos de ascensos y crea una nueva tabla denominada “*calif_por_grado*” donde recopilamos todas las calificaciones que les sirvieron a los marinos para ascender a lo largo del tiempo.

Por ejemplo, para el individuo “A” que desde su ingreso ha tenido ya tres ascensos almacenaríamos en la tabla (de acuerdo a una consulta a las tablas que aglutinan los requisitos) las calificaciones obtenidas para sus tres ascensos cada una con una fecha diferente.

Por medio del procedimiento almacenado “*calif_personal*” determinamos las calificaciones obtenidas por los marinos que no alcanzaron el ascenso, en este caso se tienen sus calificaciones hasta el momento en que ascendieron y la última calificación corresponde a la ocasión en que no pudo ascender.

Así llegamos a confeccionar un gran tabla histórica que almacena las calificaciones del personal militar por rango que contiene datos exitosos (el individuo ascendió) y datos fallidos (el individuo no ascendió).

Calif_por_grado realiza una depuración de los datos que va recopilando desechando valores nulos o fuera de rango que pueden provocar resultados no válidos. De modo que todo el ruido

* La antigüedad es una calificación que varía de acuerdo a puntos que va ganando el militar y al final los ascensos se realizan comenzando desde el marino que posee el mayor valor de antigüedad en la promoción

presente en la base de datos es limpiado por medio del procedimiento llegando a tener un conjunto de datos conciso, similar a la siguiente tabla:

Persona	Calif1	Calif2	Calif3	Calif4	Calif5	Asciende
A	19.45	19.01	19.45	19.56	18.75	Si
B	18.45	17.04	16.90	17.40	18.34	No
C	17.65	16.09	17.90	18.90	16.87	No
D	17.87	18.90	18.99	19.23	19.34	Si
E	19.80	19.75	17.90	17.89	16.75	Si
F	20.00	19.89	17.86	18.90	19.87	Si
G	16.10	16.89	19.76	17.89	16.23	No

Tabla #3.1.3.2 (a): Lista de calificaciones para Rango: Capitanes de Fragata CPFG.

El siguiente paso es realizar una transformación (discretización) de los datos, este proceso transforma a las calificaciones (datos continuos) en valores discretos cada uno con un valor de resultado objetivo, Asciende o no Asciende. La discretización se lleva a cabo usando procedimiento "Sp_Discretiza" que se vale de verificación de rangos numéricos para acreditar un valor discreto a cada calificación, al final nos queda similar a la siguiente tabla:

Persona	Calif1	Calif2	Calif3	Calif4	Calif5	Asciende
A	2	2	2	1	3	1
B	4	5	6	5	4	0
C	4	7	4	3	6	0
D	4	3	3	2	2	1
E	1	1	4	4	6	1
F	1	1	4	3	1	1
G	7	6	1	4	7	0

Tabla #3.1.3.2 (b): Lista de calificaciones discretizados para Rango: Capitanes de Fragata CPFG.

De esta manera contamos con tablas creadas por rango con los históricos de calificaciones exitosas y fallidas a partir de las cuales podemos inferir conocimiento y establecer patrones que nos

indiquen si un individuo ascenderá o no según el análisis de la proyección de sus calificaciones en los próximos años.

3.1.3.5. Minería de datos: Métodos Bayesianos.

A partir de las tablas obtenidas en la fase anterior aplicamos el modelo predictivo por medio de los 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 de las tablas obtenidas en el punto anterior, en esta fase de entrenamiento se calculan las probabilidades de éxito o fracaso para las variables usando el teorema de Naive Bayes.

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

Figura #3.1.3.3: Fórmula de la probabilidad de Bayes

Aplicando la fórmula creamos un Árbol (red bayesiana) que posee en sus hojas las probabilidades de 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.

El cálculo de probabilidades se realiza, con los datos de la tabla, de la siguiente manera:

Persona	Calif1	Calif2	Calif3	Calif4	Calif5	Asciende
A	2	2	2	1	3	1
B	4	5	6	5	4	0
C	4	7	4	3	6	0
D	4	3	3	2	2	1
E	1	1	4	4	6	1
F	1	1	4	3	1	1
G	7	6	1	4	7	0

Tabla #3.1.3.3 (a): Lista de calificaciones discretizados para Rango: Capitanes de Fragata CPGF.

Entonces, ¿Cuál es la probabilidad de que un individuo al que se le ha hecho una recopilación de sus calificaciones históricas, se las ha filtrado, y se les ha extraído patrones pueda ascender? Si sus calificaciones están representadas por el siguiente registro:

H	19.23	18.53	19.26	19.58	17.23	?
---	-------	-------	-------	-------	-------	---

Tabla #3.1.3.3 (b): Registro de calificaciones.

Que luego de ser discretizado pasa a ser:

H	2	3	2	1	6	?
---	---	---	---	---	---	---

Tabla #3.1.3.3 (b): Registro de calificaciones discretizado.

Bien, a partir de la estructura de nuestra red Bayesiana presentada en el gráfico #10, podemos tabular la probabilidad de

que el individuo H ascienda o no, de manera que podemos predecir si esta persona ascenderá al siguiente rango con cierta precisión según el valor de la probabilidad resultante.

A continuación se detalla el procedimiento matemático del cálculo de la probabilidad de que este individuo ascienda:

$$p(1/H) = [p(\text{calif}1/1)p(\text{calif}2/1)p(\text{calif}3/1)p(\text{calif}4/1)p(\text{calif}5/1)]p(1)/p(H)$$

Entonces:

$$P(1/H) = 2/4 * 1/4 * 1/4 * 1/4 * 1/4 * 4/7 = 0.0011$$

Que representa la probabilidad de que este individuo ascienda.

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 pasarán 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 con casos en los que se encuentran 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.

3.2. Análisis de los diferentes Resultados y su Impacto

3.2.1. Ascensos Navales

Una vez obtenidas las funciones discriminantes, nuestro objetivo es establecer la contribución relativa de las distintas variables a la discriminación, o lo que es lo mismo, determinar cuáles son las variables que más contribuyen a discriminar entre un grupo y otro. En nuestro caso, tratamos de determinar cual es la variable que mas prevalece al momento de determinar si una persona asciende o no, se trata entonces de averiguar cuáles son las variables que mejor representan las diferencias existentes entre los marinos. De esta forma, podremos identificar, por ejemplo, los puntos “fuertes” que caracterizan a quienes han conseguido llegar hasta altos sitios dentro de la institución y los puntos débiles de aquellos que no han logrado ascender y que optan por retirarse a la vida civil.

3.2.2. Estimación del Personal Naval por años

El hecho de tener una herramienta que realice automáticamente una regresión y calcule con un margen de error muy pequeño cuantas personas estarán en un determinado rango en los próximos 5 años provee a la Marina del Ecuador un mecanismo regulador que ayudará a planificar el manejo de las **vacantes orgánicas** que se deben crear a futuro para poder asignar plazas de trabajo a aquellas personas que en un futuro cercano ascenderán.

De esta manera la institución está un paso adelante a los posibles problemas de falta de vacantes o ingresos desmedidos desde la

Escuela Naval y sus directivos simplemente examinando las gráficas de dispersión y de residuos proporcionadas por el sistema pueden darse cuenta de las áreas donde podrían estar fallando.

3.3. Calidad de los resultados con relación a los datos

3.3.1. Modelo Predictivo: Tarea de Regresión.

La calidad de los resultados depende de la calidad de los datos con los que se realice la tarea de minería de datos, y del modelo escogido. Si el modelo es correcto, los residuos deben ser ruido blanco, esto es, deben ser una simple fluctuación aleatoria y no deben de estar relacionados o contener información. Para manejar este comportamiento residual que interfiere con la calidad de los resultados obtenidos realizamos los siguientes aspectos:

- Normalidad de los residuos
- Observaciones mal ajustadas, alejadas e influyentes

3.3.1.1. Normalidad de los Residuos.

La forma en que verificamos si los residuos se comportan de una forma normal es graficándolos en un plano estándar y observando el grado de dispersión que poseen con respecto a un marco de referencia, de esta manera podemos detectar la existencia de residuos altos que pueden ser provocados por la necesidad de efectuar sobre la variable de respuesta una transformación o bien es posible que no se hayan introducido todas las variables relevantes en el modelo. Un residuo alto indica una observación a

la que el modelo no se ajusta bien. Estas observaciones no son desechadas, sino que son localizadas en la base de datos y se procura entender el porque de su mal ajuste al modelo ya que es posible que contengan información valiosa que contribuya a estimar con mayor precisión el número de persona que ascenderán.

3.3.1.2. Observaciones mal ajustadas.

Otra fuente de residuos altos pueden ser las observaciones atípicas, ya que estas invalidan la matriz de correlaciones entre las variables explicativas y por tanto todas las estimaciones del modelo. Las observaciones alejadas del resto, como en el caso de notas de puntuación realizadas con otra escala o errores en los operadores al momento de registrar una nota de ascenso o calificación de aptitud física o disciplinaria se denominan en minería de datos observaciones influyentes. Por consiguiente, previamente al modelo de regresión hemos tenido cuidado de realizar una buena exploración de los datos recogidos, de cara a detectar observaciones atípicas y subsanar errores que, por el carácter de los datos y la naturaleza de la base de datos, que no contiene datos de muchos años atrás, generen residuos que se encuentren dentro del margen de tolerancia aceptado.

3.4. ESTUDIO BREVE DE LA VALIDEZ DE LOS MODELOS DE PREDICCIÓN A LARGO PLAZO

3.4.1. Predicción: Modelo Bayesiano

Un aspecto muy relevante en la extracción de conocimiento es que los modelos extraídos sean comprensibles. Para ayudar a resolver este aspecto es muy importante construir un modelo (en nuestro caso un modelo predictivo). El hecho de realizar las etapas de la minería de datos nos permite generar un modelo predictivo bayesiano que después del aprendizaje genera un red que queda entrenada y constituye una función, un modelo, que puede aplicarse a nuevos ejemplos sin necesidad de mirar los anteriores.

Un aspecto también fundamental que se ha considerado en el proyecto es el esfuerzo computacional que se requiere y lógicamente hemos preferido el modelo que obtiene patrones de una manera más rápida, lo que depende mucho de la eficiencia del aprendizaje que a su vez se encuentra en función del número de ejemplos utilizados para inicializar la red, el número de atributos o variables que se tomen, del espacio de hipótesis que se está considerando, del conocimiento previo existente, del nivel de error permitido, de las expectativas que se tengan para los patrones generados.

En nuestro caso al trabajar sobre ascensos navales en La Marina del Ecuador uno de los factores que más influyen en la eficiencia del aprendizaje es el tamaño de los datos (históricos dentro de la base de datos de personal) de esto va a depender el número de ejemplos o casos de aprendizaje con los que se alimentará la red bayesiana,

también influyen la cantidad de variables o atributos que se tomen en cuenta para los datos puesto que no es posible realizar una reducción significativa de variables ya que debemos regirnos estrictamente a las leyes y reglamentos de La Marina del Ecuador es en este punto donde hemos tenido mucho cuidado a la hora de seleccionar las variables para no caer en redundancias tomando en cuenta variables no linealmente independientes o no significativos que compliquen el modelo y puedan causar disminución en la calidad de los resultados al trabajarse con un espacio de alta dimensionalidad.

Sea por el número de ejemplos o el de atributos, en muchos casos hemos notado que el tiempo de aprendizaje depende de una manera no lineal respecto al tamaño de los datos.

Otro aspecto que influye en la validez y eficiencia del modelo que hemos elegido para el proyecto es la calidad de los datos y la existencia de conocimiento previo (representado dentro de la base de datos por las tablas historias que guardan información de ascensos por años)

Poder Predictivo

La metodología utilizada para llevar a cabo pruebas del poder predictivo de las redes bayesianas con la base de datos del personal naval se detalla a continuación.

- Dividimos la base de datos en dos. Una de control o entrenamiento (aproximadamente $2/3$ de la base total) y otra de contraste o validación (con los datos restantes).

- Procesamos la base de datos de entrenamiento mediante el algoritmo de aprendizaje de redes bayesianas para obtener el subconjunto de variables que conformarán la red.
- Repetimos para el 10%, 20%, ..., 100% de los datos de la base o Repetimos 30 veces.
- Tomamos al azar el X% de la base de datos de control según el porcentaje que corresponda a la iteración.
- Mediante ese subconjunto de casos de la base de control, realizamos el aprendizaje estructural y paramétrico de las redes bayesianas.
- Evaluamos el poder predictivo de la red utilizando la base de datos (lo medimos como el porcentaje de casos clasificados correctamente sobre el total de casos evaluados). Cuando el porcentaje de certeza sobre la clasificación de un caso no sea del 100%, se considera la clase con mayor probabilidad.

Si graficamos las curvas que representan el poder predictivo en función de la cantidad de casos de entrenamiento observamos que obtenemos una función creciente. Este fenómeno se da independientemente del dominio de datos utilizado y del método evaluado. Esto se traduce en la conclusión de que cuanto mayor sea la base de datos que se disponga para la tarea de entrenamiento, mejor será el poder predictivo de la red bayesiana obtenida ya que la misma, será una representación más fidedigna del dominio real que intenta representar el modelo.

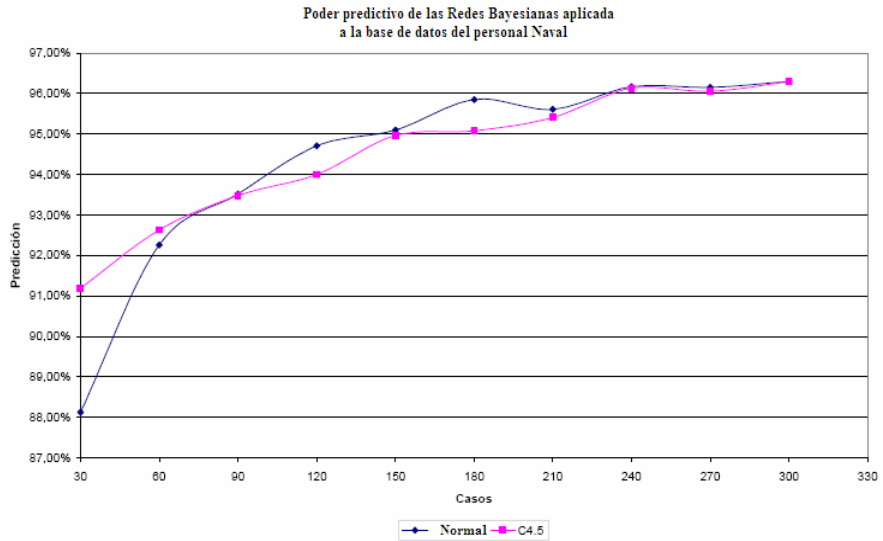


Figura #3.4.1.1 Poder predictivo de las redes bayesianas en función del número de casos de aprendizaje

3.4.2. Regresión

En minería de datos el objetivo final de los modelos de regresión como el aplicado en nuestro proyecto es realizar buenas predicciones de individuos anónimos, es decir con valor de la variable de respuesta desconocido. Por lo tanto, usamos un modelo con errores de predicción lo más pequeños posible.

Como una métrica que nos da una idea de la eficiencia y la validez de la regresión polinomial podemos calcular el error cuadrático medio que se calcula de la siguiente manera:

$$y^i = r(x) + \varepsilon$$

Si $y^i = x^i \cdot b$ es el valor predicho por el modelo lineal, entonces el error cuadrático medio de predicción puede descomponerse en el error en la predicción de la función de regresión más el error debido a la predicción individual y^i .

Sin embargo podemos fácilmente tener estimadores sesgados: basta que no se cumpla alguna de las hipótesis del modelo o simplemente no haber incluido todas las variables relevantes. Por ello es prudente incluir el mayor número posible de variables explicativas en el modelo a fin de minimizar el riesgo de sesgo. Sin embargo añadir variables no relevantes significa añadir ruido al modelo y aumentaríamos la varianza de las predicciones. Tratamos entonces que el manejo de los datos del personal naval confluya en un equilibrio entre sesgo y varianza. Esto se consigue disponiendo de suficientes datos y dividiéndolos en dos grupos, uno conforma la muestra de aprendizaje ósea el conjunto de entrenamiento y otra muestra, independiente de la anterior, que formará la muestra de prueba que a su vez sirve para estimar el error de predicción del modelo.

4. ANÁLISIS ECONÓMICO



Figura #4.0 Procesos

En este capítulo se discutirá sobre como una solución planteada en base a un problema puede afectar económicamente a corto plazo a una institución. El análisis y planteamiento de la solución no busca necesariamente obtener la mejor respuesta a un problema planteado al mejor costo

sin estudiar a fondo lo que este demandará en el desarrollo del mismo.

4.1. ANÁLISIS DE COSTO

Desde el momento mismo que se concreta el desarrollo de la solución existen una serie de pasos que conllevan a un levantamiento de información eficaz, para luego hacer un estudio exhaustivo de los mismos y determinar cuales son las necesidades reales, y así poder establecer un cronograma de trabajo para resolver el mencionado problema. Es entonces que desde el primer paso el sistema empieza a tener un “COSTE” significativo y es por el cual en los siguientes puntos se hará una referencia hacia los mismos los cuales en conjunto con los demás gastos asociados a la solución determinarán el costo total de los mismos.

4.1.1. COSTO DEL PLANTEAMIENTO

Como se explicó anteriormente el inicio de una nueva aplicación comienza a contraer costos desde que se inicia el primer paso que es el de la definición formal del problema, para luego incurrir en un

levantamiento de la información la cual generan “COSTES” que se van sumando a los gastos a incurrir en los siguientes pasos, claro está que mencionados pasos contienen tareas las cuales conllevan a procesos que generan una alta demanda de recursos tanto económicos, organizacionales y sobre todo de tiempo.

En el primer momento que se reúnen el grupo de expertos a plantear el problema en conjunto con los potenciales usuarios finales los denominados costes están influenciados en como:

1.- Van a lucir los escenarios actuales usando herramientas que



Figura # 4.1.1 (a):
paquete instalador
VISIO.

permitan modelar el planteamiento inicial.

Generalmente hablamos de herramientas “moldeadoras”, las cuales para efectos de visualización pueden recomendarse el uso de:

POWERPOINT, PAINT como herramientas

básicas con costes significativamente bajo con respecto a las herramientas industriales a usar, debido a que son herramientas de uso diario y las cuales se cuenta con las licencias del mismo.

También están herramientas que permiten una mejor interpretación de los problemas actuales y es el caso de DREAMWEAVER que permite una mejor apreciación y visualización de los escenarios actuales (son de consideración inicial para poder entender el problema) para así poder presentar a los usuarios si es efectivamente lo que está ocurriendo, claro que en este caso esta herramienta comienza a generar un coste proporcionalmente más alto con respecto a las

anteriores mencionadas debido al costo que bordea los \$300 aproximadamente y se debe a que esta es una herramienta de diseño gráfico y por lo tal su uso comercial genera un coste relativamente mayor a las herramientas de oficinas. En fin podemos citar un conjunto de herramientas y justificar su uso pero ese no es el objetivo de este punto, más bien dar a entender de cómo una interpretación de un escenario que es en realidad parte del “gran problema” debe de ser moldeado técnicamente y mencionar los gastos por uso del mismo.

2.- Se va incurrir en herramientas que permitan obtener una mejor representación visual de los problemas incurridos como consecuencia de los escenarios anteriormente mencionados. Específicamente estamos comentando el hecho de cómo relucir los problemas “ocultos” que inicialmente no son visibles con los escenarios iniciales y es el caso por mencionar algunos:

a.- Quienes van acceder a la aplicación (usuarios finales, creación de perfiles, permisos).

b.- Como debería de actuar la aplicación frente al usuario (lo que el usuario espera).

c.- El tiempo de respuesta a una acción ejecutada por el usuario

(cuanto es lo permisivo en tiempo de ejecución-respuesta).

En nuestro caso la mejor representación fue hecha en un esquema realizado en VISIO claro



Figura # 4.1.1 (b): paquete instalador Visual Studio 2003.

que su precio que está por los \$250 si justifica su uso, que es una aplicación que tiene varios niveles de visualización que van desde: el esquema de un edificio, hasta el diagrama de procesos de un conjunto de tareas. Esta herramienta nos permitió establecer algunos de los problemas que si se veían y que algunos eran dependientes de otros y claro los que se mencionaron anteriormente y se catalogaron como “ocultos”. Como tal esta aplicación genera un gasto ya que como se describió la herramienta de diseño gráfico su uso está ligado a licencias para trabajos profesionales y no de oficina.

3.- Concebir el modelo final, el que mejor represente los problemas anteriormente mencionados. Una vez que están claros y sin problemas de entendimiento se debe de definir el conjunto de escenarios finales los cuales deben de recibir la aprobación de los usuarios y es en este paso en el que se cierra el planteamiento del problema y se recibe la aprobación para el levantamiento de la información.

El uso de herramientas como DREAMWEAVER, VISIO, y hasta herramientas de desarrollo como Visual Basic 6.0, C#, Visual Basic.Net, no estarían demás a la hora de establecer tan importante logro, siempre y cuando cumpla con los gastos estipulados en una partida presupuestaria destinada a incurrir en dichas herramientas. Dichas herramientas podrían ser usadas en conjunto con aplicaciones tipo CASE para moldear el modelo final en conjunto con los requisitos básicos para

establecer métricas de desarrollo. Se puede definir a las Herramientas CASE como un conjunto de programas y ayudas que dan asistencia a los analistas, ingenieros de software y desarrolladores, durante todos los pasos del Ciclo de Vida de desarrollo de un Software (Investigación Preliminar, Análisis, Diseño, Implementación e Instalación).

En el mercado hay herramientas como ERwin (diagramación de Bases de Datos), EasyCASE (esquemas de base de datos e ingeniería reversa), SystemArchitect (integrador de herramientas y metodologías), todas estas herramientas están entre \$700 y \$2000 aproximadamente, por lo que depende de la calidad del trabajo a realizar se justificará su precio.

Así como mencionamos el cómo se va a visualizar los problemas que acarrea el planteamiento, es necesario especificar el porque:

1.- Interpretar el desarrollo del problema enfocado a una solución la cual determine si su coste es viable para el desarrollo del mismo. Un proyecto que desde el inicio fue contemplado como ideal y exitoso sin medir el costo real puede llegar a ser un fracaso a corto plazo debido a que el camino del desarrollo se encontrará con “dificultades” que se debieron de presupuestar.

Por lo mencionado anteriormente es esencial definir el conjunto de elementos que deberán de participar en la concepción del desarrollo de la solución midiendo con métricas los costes que se incurrirán como:

- a.- Recursos humanos que levanten la información. Por cada uno un gasto de \$300 mensuales.
- b.- Recursos humanos que depuren y establezcan las verdaderas necesidades. Por cada uno un gasto de \$250 mensuales
- c.- Recursos técnicos para la elaboración, descripción y especificación de la información a usar. Por cada uno un gasto de \$250 mensuales.
- d.- Recursos técnicos para el desarrollo, pruebas, depuración de la aplicación. El grupo anterior puede ayudar en esta parte con una ligera bonificación.
- e.- Espacios físicos de trabajos los cuales deben de facilitar el acceso a la información que se requiera. Esto está sujeto a que si lo hace en un lugar que no incurra gastos por espacios físicos de alquiler, que puede ser la casa de uno de los desarrolladores

Por lo que se menciona anteriormente se incurren en gastos reales que se deben de presupuestar ya que como se mencionó al principio el éxito de un proyecto dependerá enteramente del control de los gastos que se mencionaron anteriormente y los cuales se consideran dentro de los básicos.

En nuestro caso los puntos anteriormente mencionados solo se resumen en:

Gastos en Transporte.- Para dirigirse al lugar donde se procedió al levantamiento de la información; 2 veces por semana desde el inicio del proyecto.



Gastos en equipo.- Para el desarrollo del proyecto; desde el inicio del proyecto.



Consumos por alimentación.- Alimentación fuera del hogar, y sesiones prolongadas de trabajo; mensualmente.



Gastos en útiles de oficina.- Para uso del proyecto en particular para elaboración de borradores, informes, de uso personal para escribir ideas, analizar tareas, etc; desde el inicio del proyecto.



RUBRO	TOTAL
Gastos en transportación.	\$6.00 (semanales).
Gastos en equipo.	\$800 (un solo pago).
Consumos por alimentación.	\$20 (mensuales).
Gastos en útiles de oficina.	\$200 (un solo pago).

Tabla #4.1.1: cuadro que representa los gastos básicos internos del grupo de desarrollo.

2.- Seleccionar las mejores herramientas de desarrollo para la elaboración de la aplicación. Un proyecto encaminado a ser exitoso debe de optimizar recursos técnicos específicamente de desarrollo y es que escoger una herramienta costosa que no justifique su impacto podría situar al proyecto en un ambiente incontrolable con relación al gasto ya que podría desviarlo de la intención inicial que es el desarrollar bajo un conjunto de

esquemas “predefinidos”. En nuestro proyecto usamos la herramienta Visual Studio 2003 con SQL2000 Server la cual nos permitió elaborar los ambientes planificados en los escenarios especificados anteriormente. El coste por el uso de esta herramienta está en los \$300 aproximadamente por el paquete completo.

4.1.2.COSTO DE LA SOLUCION

Una vez definido los escenarios de desarrollo y elaborado las pantallas que fueron aprobados por los usuarios finales se debe de incurrir en los gastos operativos que fueron mencionados anteriormente pero no descritos con mayor detalle que es lo que se va a hacer en esta parte.

El hecho de que este planteado los problemas y el esquema a adoptar para el desarrollo de la solución no implica que se quiera seguir o incurrir en gastos temporales. La planificación de los gastos que se van a seguir debe de estar contemplado como parte del problema ya que como tal no es ajeno a la solución.

Analizar como deben de incurrirse en mencionados gastos dependerá de que tan extenso o que tan grande o que tanto abarcará el desarrollo de la solución, y es que pesará más el saber que no se contó con todos los detalles que deben de abarcar mencionados gastos.

Desde que se escoge y se aprueba la metodología de desarrollo de la solución se deben de comenzar a controlar los gastos que se

presupuestaron al principio como parte del problema. Tenemos costos básicos como:

1.- Gastos en transportación: son los gastos entendidos entre el lugar de trabajo y el lugar donde está la información que es de importancia para el desarrollo. Es esencial definir que el gasto sube cuando no hay una atención adecuada desde el levantamiento de la información hasta la terminación del mismo. Como se explica en la gráfica este tiende desde un principio a mantenerse en un nivel alto y tiene a bajar de acorde al interés del usuario conforme se avanza en la elaboración del mismo.

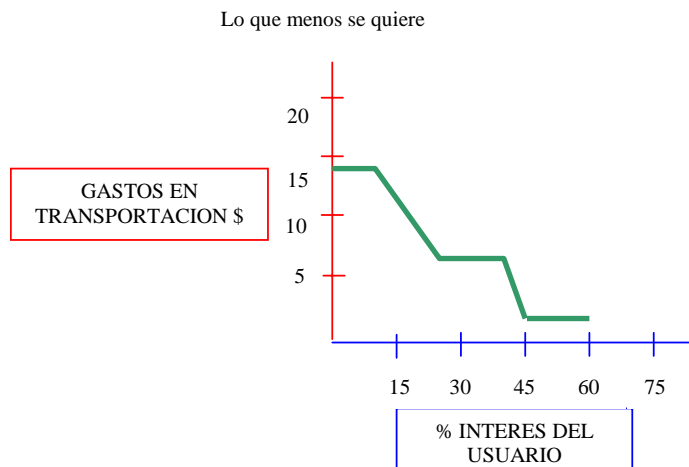


Figura # 4.1.1 (g): Representación de los gastos en transportación vs. Porcentaje de interés del usuario.

En el gráfico anterior se visualizan los niveles de gastos obtenidos en el proyecto. Cuando el proyecto comienza los gastos se dispararon hasta un nivel razonable y es por el levantamiento de información y procesos relacionados al mismo, interpretándose en visitas continuas que representan gastos en transportación.

Cuando el usuario comienza a recibir información acerca de cómo se está llevando los procesos relacionados al desarrollo del sistema las visitas comienzan a disminuir debido a que el usuario pone un interés y entrega la información pertinente del caso que sirven como parte vital en la elaboración de la aplicación. Poniendo un caso está que desde el principio las visitas fueron continuas hasta establecer claramente cual era la intención de elaborar una aplicación que tuviera un impacto dentro los “procesos manuales”.

Conforme se fue elaborando informes que describían cual debería de ser el compromiso a entregar las visitas fueron disminuyendo ya que se proporcionaron datos importantes y relevantes al caso. Esto hizo que se comenzaran a bajar discretamente los gastos de transporte hasta llegar a un nivel estable como de dos visitas cada 3 semanas.

2.- Gastos en útiles de Oficina: estos costes corresponden a todo el material como: resmas para impresión de información crucial para el desarrollo, cartuchos de impresión, cuadernos así como plumas. Estos elementos básicos corresponden al uso de recursos comprendidos desde que el tema fue propuesto hasta la impresión de los documentos de especificación en uso de la aplicación.

Así como en el caso anterior conforme se tiene un claro entendimiento de lo que se va a hacer este gasto considerablemente baja debido a que se usa menos material

para por ejemplo aclarar una idea o la diagramación de una tarea en especial cuando no se tiene entendido por donde comenzar.

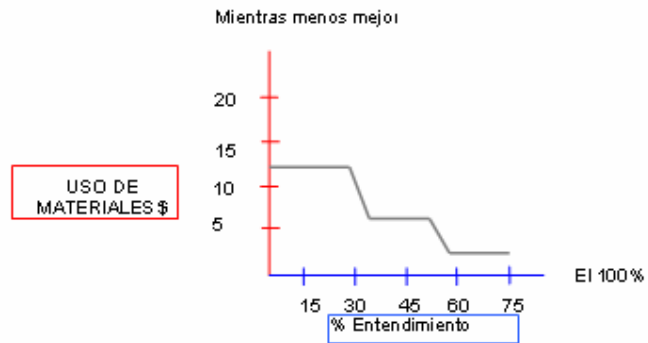


Figura # 4.1.1 (h): Representación de los gastos en materiales vs. Porcentaje de entendimiento.

Como apreciamos en esta gráfica se ve claramente que los gastos incurridos en materiales de oficina disminuyen circunstancialmente conforme se tenga claro como lo que se quiera hacer. Esta gráfica no refleja el gasto del uso en impresiones cuando es por reportes o documentación de usuario y/o especificaciones técnicas a revisar, ya que considerablemente es alto. Esto se debe a que conforme se vaya avanzando y se vaya entendiendo que es lo que hay que hacer se va usando menos papeles para hacer diagramaciones, pruebas, planificaciones y otro tipo de acciones ya que cuando se ponen en claro todas las situaciones y eventos es cuestión de armar bien el escenario, visualizarlo y desarrollarlo.

3.- Gastos en licencias para programas: esto debe de ser un gasto que dependerá exclusivamente del usuario como cliente ya que no se negocia la aplicación y la herramienta en la que se

construyó. Como tal el grupo de desarrollo no debe infringir en gastos que no sean parte de la solución y es el caso de las licencias para uso de herramientas de desarrollo.

Las obligaciones actuales estipulan que el permiso para el uso de una herramienta de desarrollo que tiene licencia normalmente está regulado a que sea cancelado por el cliente y no por el grupo de personas que se comprometen a entregar una aplicación siempre y cuando se convenga dicho acuerdo.

4.1.3. COSTO DE LAS HERRAMIENTAS INVOLUCRADAS

Como mencionamos anteriormente los gastos incurridos tanto por la parte inicial en la elaboración de un proyecto no está ligado a gastos simples y poco prácticos y es en esta parte que se trata de aclarar el asunto del alto costo que puede representar adquirir licencias por parte del grupo que va a desarrollar la aplicación.

Una licencia de software otorga al usuario derecho legal a utilizar un software. Por cada programa de software de Microsoft que se utiliza, se otorga una licencia al usuario y ésta se documenta en el Contrato de Licencia de Usuario. El software está protegido por la ley de

derechos de autor, que establece que el producto no puede copiar sin autorización del dueño de derechos de autor.

En el caso de que el grupo de desarrollo de la solución acarrearía el gasto este aumentaría significativamente debido a



Figura # 4.1.3: Paquete de instalación de Visual Studio 2003

que la licencia no saldría por volumen que es aplicado a empresas y

por el cual hay un descuento en comparación a que si se adquiere para usuario final, en el caso de empresas este costo es de \$300 por computadora claro por el uso de Visual Studio 2003 (Este precio fue consultado en el laboratorio de computación de la FIEC-ESPOL).

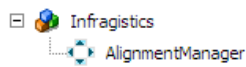


Figura # 4.1.3:

También está el uso de herramientas de control web como Infragistic 2005 la cual tiene un precio estimado en el mercado de \$500.

Entonces se ve que hay una relación de dependencia de costos si es que se adquiere licencias ya sea por la empresa del cliente o por usuario final. Definido este punto en el proceso de definición de problemas se define cuanto sería el costo agregado si se quiere conocer ambos lados.

4.1.4. COSTO GLOBAL DE LA APLICACIÓN

Definidos todos los posibles gastos que se pueda incurrir, hay que definir cuales de estos se pueden cargar en el proyecto y así obtener un gasto real y global por el cual se deba de cobrar.

Acordémonos que no todos los gastos que salen inmediatamente sin haberlo definido se consideran como un gasto práctico, específicamente aquellos que no tienen su razón de ser como es el caso de llamadas telefónicas las cuales se debe de solventar en casos regulares.

Definidos los gastos que se van a incurrir en el proceso de desarrollo se procede a sacar un gasto global que debe de ir acompañado de los costes por licencias y uso de programas para desarrollo lo cual

dependerá si amerita invertir en una solución que genera costes desde que se hace la planificación.

Los gastos que hay que tomar en cuenta van desde el planteamiento del problema como tal, pasando por la especificación de escenarios, modelación de los requisitos, elaboración y diseño de ventanas, adquisición de las licencias, desarrollo de la solución, transportación, gastos alimenticios, útiles de oficina. En este documento no se puede dar un gasto debido a que fue un proyecto de tesis de grado como parte del tópico “Minería de Datos”, pero si se puede especificar los gastos incurridos como parte del proceso de desarrollo los cuales se mencionaron anteriormente.

4.2. IMPACTO ORGANIZACIONAL POR AHORRO ECONOMICO DE PROCESOS

Esta parte trata de analizar como procesos manuales y a veces complicados hacían incurrir en procesos complicados que generaban gastos económicos.

Con la aplicación se pretendió eliminar el conjunto de procesos los cuales hacían una búsqueda exhaustiva de todos los requisitos para que un individuo pueda cumplir con las leyes para que pueda ascender.

El solo hecho de pensar que se quiera analizar un conjunto aproximado de 100 individuos que soliciten el ascenso conlleva a que se tenga que analizar las hojas de vidas de los mismos, seleccionar los que cumplen, después seleccionar los mejores para posteriormente elegir los que ocuparán las plazas orgánicas.

Como vemos esto resulta en una práctica muy poco optimizada ya que se tiene un conjunto de personas que están tratando de obtener los mejores resultados aplicando las leyes y a su vez seleccionando los mejores individuos para luego decidir quienes ocuparán dichos puestos.

Con la aplicación que se desarrolló se ahorra todo ese conjunto de procesos. Basta con que se seleccione el conjunto de individuos especificados por el rango y la promoción se hará el mismo proceso anterior pero sin la intervención humana, lo que dará como resultado una lista de individuos que estará organizada por una jerarquía que establece el orden de importancia y significancia en el que se van a escoger para ocupar las plazas orgánicas.



Figura # 4.2: Procesos.

Todo este proceso es el resultado del análisis del modelo de negocio que lleva la institución para el ascenso y el cual está reflejado en las leyes que contempla el estado y es quien regula como se debe de hacer.

Por lo que hemos descrito el proceso que antes duraba un conjunto de días estipulados ahora solo va a tardar un par de minutos, ahorrando tiempo (días que normalmente toma el proceso cotidiano), dinero (costos en: recursos humanos, equipos, etc.) y recursos (lugar donde se hará el proceso, útiles de oficina, etc.), los cuales pueden ser destinados al siguiente proceso que es el de escoger a los que van a ocupar la vacante orgánica. Se espera que con el uso de esta aplicación que tiene elementos de predicción y estimación se tienda a

disminuir gastos y por consiguiente el ahorro económico dentro de la institución.

En el mercado también existen herramientas más poderosas que la que presentamos en este trabajo y que están enfocadas en un análisis minucioso cuyos objetivos son las de estimar y predecir información mediante una serie de indagaciones proveniente de un repositorio.

Hay muchos aspectos a mencionar y los cuales pueden ser propósito de discusión pero para efectos prácticos podemos basarnos en dos puntos de vistas que disciernen el trabajo realizado por estas “herramientas empresariales”.

Por un lado si hablamos de aspectos positivos nos enfocamos en la inmensa utilidad que puede demandar en la institución, mediante la cual no solo se lo puede emplear para fines de estimación y/o predicción ya que su uso va mucho más allá y es el de extraer conocimiento numérico acerca de patrones o comportamientos ocultos como por ejemplo conocer información de la población de personal naval que se le pueda asociar un perfil antes de ser ascendido lo que daría como resultado un conjunto de individuos que tienen como elemento adjunto sus conocimientos académicos para que de esta forma sean aprovechados sus niveles intelectuales y académicos ganados en el transcurso de su vida militar.

Por otro lado también está el asunto de la personalización de la herramienta que se quiera introducir en la institución para fines específicos. El hecho está en que se deben de aplicar un conjunto de “adecuaciones” para que la herramienta sea 100% apegada a los datos

de la institución de allí que nace el paradigma de que “los algoritmos se adaptan a los datos o los datos a los algoritmos”. Esto representaría un gasto adicional ya que se compra la herramienta como tal pero se le tiene que agregar un precio que es el de la adaptación al medio para que funcione con las características deseadas. Esto comúnmente se encuentra en el mercado ya que por su costo una herramienta de uso general y considerada como uso empresarial no son ni siquiera 60% adaptables al negocio inicialmente, de allí que hay que seguir un conjunto de pasos para adaptar a las características del mismo, lo cual demanda un gasto adicional pero los beneficios son establecidos inicialmente y los cuales son el motivo de la compra del producto.

Haciendo una reseña en el mercado encontramos herramientas básicas como lo son: newHammerTap™ con un precio de \$179.00, SPSS con un precio promedio de \$800 (dependiendo de la característica que se desea), [Synapse](#) de Peltarion con un precio aproximado de \$1200 (todos los precios citados en este párrafo corresponden a las páginas de los productos; costo sin licencias).

Claro si consideramos las librerías de tratamiento de información que circulan en el INTERNET encontraremos algunas que son aceptadas por sus beneficios pero también demandan un gasto.

Para concluir este capítulo de análisis económico nuestra herramienta como tal reporta solo los gastos mencionados anteriormente pero sin asumir costo de profesionalidad en la elaboración del mismo (como se mencionó anteriormente este rubro es significativo si se lo quiere considerar) pero por ser un trabajo de tesis para una institución del

ESTADO aquello no se lo puede considerar en práctica por lo que no es ético ni profesional de nuestra parte.

4.3. ANÁLISIS COMERCIAL.

Esta sección debe detallar cómo la aplicación representa un valor económico para la empresa que quiera usarlo y cómo estaría constituido, para de esta forma obtener un monto el cual determina su valor comercial.

Desde que se comienza a mencionar quien es la figura empresarial que manejará dicho producto en este caso una institución del gobierno se deja de hablar de un valor comercial porque este estudio como tal partió de una necesidad de la Marina del Ecuador la cual salió como propuesta de tesis.

Adicionalmente a lo mencionado no se puede negociar ni vender los derechos de este trabajo porque es un estudio particularizado en los datos que la Marina del Ecuador ofreció para la elaboración del mismo, y por lo tanto se prohíbe su difusión y uso sin consentimiento de la institución.

CONCLUSIONES Y RECOMENDACIONES.

Conclusiones:

- Esta tesis es el resultado del estudio de los diferentes procesos necesarios para el ascenso de personal, estos procesos demandan recursos y mucho tiempo, por lo que podemos concluir que, al usar nuestra aplicación se ahorran dichos costos.
- Para poder realizar nuestra aplicación es necesario conocer el funcionamiento interno de la Marina del Ecuador por lo que se hizo un estudio de las leyes y reglamentos que maneja.
- La base de datos que nos proporcionaron tenía gran cantidad de información y algunas de sus tablas tenían muchos campos. Fue necesario utilizar mucho tiempo de análisis, para conocer como estaba diseñada la base de datos, como se relaciona cada tabla y cada campo con las leyes y reglamentos antes estudiados.

Recomendaciones:

- Las aplicaciones externas que alimentan el repositorio de datos, deben de impedir que se almacenen datos nulos (sin valor).
- 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.
- Todos los usuarios de la aplicación están siendo autenticados en la base de datos PERSONAL, por lo que se recomienda que dichos usuarios sean creados en la base de datos.

BIBLIOGRAFÍA

- Las imágenes fueron proporcionadas a través de INTERNET con el repositorio de imágenes de google de www.google.com.

Referencias bibliográficas en INTERNET a través de:

- Microsoft www.microsoft.com con la búsqueda de licencias.
- Monografías 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.
 - Documentos acerca de “Aprendizaje Bayesiano” de Oscar Javier Prieto Izquierdo y Raúl Casillas Díaz.
 - Documento de análisis de “Redes Bayesianas” de Eduardo Morales.
- 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.

ANEXO A: LEYES DE ASCENSO (EXTRACTO DE ALGUNAS LEYES)

A continuación se presenta un extracto de algunas leyes que usaron para la elaboración del sistema, las cuales por razones de seguridad de estado solo mostramos las que sirvieron como complemento.

REQUISITOS COMUNES PARA EL ASCENSO

Los requisitos comunes que debe reunir el militar para el ascenso en todos los grados son los siguientes:

- a) Acreditar el puntaje mínimo que para cada grado se determine en la presente Ley;
- b) Aprobar el correspondiente curso;
- c) Haber cumplido funciones en unidades correspondientes a su clasificación, por lo menos durante un año en el grado, para oficiales superiores, suboficiales y sargentos primeros, y dos años para el resto de jerarquías;
- d) Haber sido declarado apto para el servicio, de acuerdo a la ficha médica; y,
- e) Haber cumplido con el tiempo de permanencia en el grado.

De las calificaciones para el ascenso

La calificación para el ascenso es el resultado de la evaluación del desempeño del militar en el grado, en base al análisis de las calificaciones anuales, cursos O exámenes de promoción y de los méritos y deméritos.

Para el caso de los oficiales en el grado de Coronel o Capitán de Navío, y de tropa en el grado de Suboficial Segundo, la evaluación indicada en el inciso precedente se realizará en todos los grados anteriores, de acuerdo con el Reglamento.

El puntaje mínimo requerido para el ascenso al grado inmediato superior será el siguiente:

a) General de División o sus equivalentes	19,00
b) General de Brigada o sus equivalentes	18,50
c) Coronel o su equivalente	18,00
d) Teniente Coronel o su equivalente	17,50
e) Mayor o su equivalente	17,00
f) Oficiales subalternos	16,50
g) Suboficiales 1ros.	18,00
h) Suboficiales 2dos.	17,50
i) Sargentos 1ros.	17,00
j) Sargentos 2dos.	16,50
k) Cabos	16,00

ANEXO B: REGLAMENTOS (EXTRACTO DE ALGUNOS REGLAMENTOS)

A) Conforme a lo que señala el reglamento militar aprobado por el gobierno militar y el poder ejecutivo establece que la sucesión en el Comando de una Fuerza o Unidad de Combate cuya reglamentación se menciona en el anexo, recaerá sobre el militar de arma de mayor grado o antigüedad; a falta de oficiales de arma, la sucesión seguirá el orden establecido en este artículo respetando la antigüedad.

B) El tiempo de permanencia en el grado para el personal de Oficiales de Arma y de Servicios o Técnicos, es el siguiente:

- Subteniente o Alférez de Fragata 4 años
- Teniente o Teniente de Fragata 5 años
- Capitán o Teniente de Navío 6 años
- Mayor o Capitán de Corbeta 6 años

C) Los coroneles y capitanes de navío de arma declarados idóneos que no fueren considerados para su ascenso en la primera oportunidad, podrán esperar a una segunda oportunidad con la siguiente promoción, situación en la que permanecerán hasta un máximo de un año. En caso de no producirse dicha calificación en el lapso anteriormente señalado, deberán solicitar su disponibilidad o serán colocados en tal situación.

D) Requisitos para el ascenso de oficiales de arma y de servicios o técnicos

Los oficiales de arma y de servicios o técnicos a más de los requisitos comunes para su ascenso, cumplirán los siguientes, según su grado:

- a) Para el ascenso a Teniente, Capitán y Mayor o sus equivalentes en la Fuerza Naval, aprobar el respectivo curso de estudios militares, establecidos en los reglamentos pertinentes de cada Fuerza;
- b) Para ascender a Teniente Coronel o Capitán de Fragata, haber aprobado el curso de Estado Mayor en las respectivas academias de Guerra nacionales;
- c) Para ascender a Coronel o Capitán de Navío se requiere:
 - 1) No haber sido sancionado con suspensión de funciones durante su carrera militar;
 - 2) Ser diplomado de Estado Mayor en las academias de guerra nacionales y no haber repetido dicho curso por ninguna circunstancia; y,
 - 3) No haber sido reprobado en un Curso Militar o Técnico en el país o en el exterior, de acuerdo al Reglamento de cada Fuerza.

E) Requisitos para el ascenso de los oficiales especialistas

Los oficiales especialistas, a más de los requisitos comunes para su ascenso, cumplirán con los siguientes:

- a) Para el ascenso a Teniente o Teniente de Fragata, acreditar el título profesional de su especialidad para aquellos oficiales que se reclutaron como egresados de los Institutos de Educación Superior;
- b) Para el ascenso a Capitán o su equivalente en la Fuerza Naval, aprobar el respectivo curso de estudios en las áreas de su especialidad, de conformidad a lo establecido en el Reglamento pertinente de cada Fuerza;
- c) Para el ascenso a Mayor, o su equivalente en la Fuerza Naval, aprobar el respectivo curso de estudios en las áreas de su especialidad, y acreditar el título de Post-Grado, conforme a lo establecido en el Reglamento respectivo de cada Fuerza.

F) Requisitos para el ascenso del personal de tropa de arma y de servicio o técnicos

El personal de tropa de arma y de servicios o técnicos, a más de los requisitos comunes para su ascenso, según el grado, cumplirá con los siguientes:

- a) Para el ascenso hasta el grado de Sargento Primero, inclusive, haber aprobado los cursos establecidos en los pertinentes reglamentos de cada Fuerza;
- b) Para ascender a Suboficial Segundo, haber realizado un curso de administración militar y, no haber sido sancionado con suspensión de funciones durante su carrera militar;
- c) Para ascender a Suboficial Primero, haber presentado un trabajo de investigación o práctico en su especialidad, de interés para su Fuerza, de acuerdo al respectivo Reglamento.

ANEXO C: RANGOS

ACEP	Alumnos de Cursos Especiales
GAMA	Guardiamarina
ALFG	Alférez de Fragata
TNFG	Teniente de Fragata
TNNV	Teniente de Navío
CPCB	Capitán de Corbeta
CPFG	Capitán de Fragata
CPNV	Capitán de Navío
CALM	Contraalmirante
VALM	Vicealmirante
ALMI	Almirante

Tabla 1.1: Grados de Oficiales

GRTE	Grumete
MARO	Marinero
CBOS	Cabo Segundo
CBOP	Cabo Primero
SGOS	Sargento Segundo
SGOP	Sargento Primero
SUBS	Sub-Oficial Segundo
SUBP	Sub-Oficial Primero
SUBM	Sub-Oficial Mayor

Tabla 1.2: Grados de Tripulantes