

## Modelo de eficiencia de las terminales de contenedores del sistema portuario español

**GONZÁLEZ-CANCELAS, NICOLETTA**

*nicoletta.gcancelas@upm.es*  
*Universidad Politécnica de Madrid, Ingeniería Civil. Transportes*

**SOLER-FLORES, FRANCISCO**

*f.soler@upm.es*  
*Universidad Politécnica de Madrid, Ingeniería Civil. Transportes*

**CAMARERO-ORIVE, ALBERTO**

*alberto.camarero@upm.es*  
*Universidad Politécnica de Madrid, Ingeniería Civil. Transportes*

Recibido (27/02/2013)

Revisado (19/10/2013)

Aceptado (21/12/2013)

RESUMEN: Los puertos cumplen un rol para el desarrollo del comercio y en general para la competitividad del país, especialmente en el contexto de globalización de las últimas décadas. En el sector portuario, el avance tecnológico internacional ha determinado una reducción en los costos del intercambio comercial entre grandes distancias. El resultado de estos cambios ha sido que el sector portuario ha iniciado un proceso de rápida renovación de sus instalaciones para que éstas respondan adecuadamente a la demanda por sus servicios. Mientras que en el pasado se identificaba la existencia de cierto poder monopólico de los puertos en un territorio determinado debido a los altos costo que implicaba movilizar carga por un puerto alternativo, en el marco de estas transformaciones, han aumentado las posibilidades de optar por puertos diferentes, generándose así un entorno cada vez competitivo. En este contexto, la eficiencia portuaria ha adquirido mayor importancia como determinante de los costos y la calidad del transporte, elementos que afectan la dinámica del comercio exterior y el grado de competitividad que alcanza un país.

El objeto del presente artículo es determinar los escenarios de eficiencia de las terminales de contenedores de los puertos que integran el actual Sistema Portuario Español. Estos escenarios permitirán a los gestores portuarios tomar las medidas oportunas para mejorar la gestión de los puertos. Para el análisis de los escenarios de eficiencia de las terminales de contenedores y mediante el empleo de modelos gráficos probabilísticos, Redes Bayesianas concretamente, se han definido las principales variables y se ha realizado inferencia en escenarios virtuales.

*Palabras Clave:* eficiencia, puertos, terminales de contenedores, redes bayesianas.

ABSTRACT: Ports play a role in the development of trade and in general for the country's competitiveness, especially in the context of globalization of the last decades. Specifically in the port sector, international technological advancement has resulted in a reduction in the costs of trade between large distances. The result of these changes has been that the port sector has begun a process of rapid renewal of its facilities so that they are responsive to the demand for their services. While in the past the existence of some monopoly in a given area is identified due to the high cost involved mobilizing load an alternate port, as part of these changes have increased the possibilities of choosing different ports, thus creating an increasingly competitive environment. Port efficiency has become more important as a determinant of the cost and quality of transport, factors affecting the dynamics of foreign trade and the degree of competition that reaches a country.

The purpose of this paper is to determine the stage efficiency of container terminals in the ports that make up the Current Spanish Port System. These scenarios enable port managers take appropriate measures to improve port management. For the analysis of scenarios efficiency of container terminals and by the use of probabilistic graphical models, specifically Bayesian networks, we have defined the main variables and inference is performed in virtual environments.

*Keywords:* efficiency, ports, container terminals, Bayesian networks.

## 1. Introducción

Los puertos son nodos de intercambio modal que han evolucionado atendiendo a las necesidades del mercado. Se encuentran en el sistema de transporte de manera que actúan como un elemento más de la cadena de transporte por lo que deben ser eficientes para no constituir el cuello de botella de dicha cadena. Los puertos están inmersos, a la vez que son impulsores, en los procesos de globalización de las sociedades, y es por ello que no son ajenos a la potencialidad de nuevas metodologías como pueden ser los sistemas de inteligencia artificial, este proceso de apertura constante impulsa el crecimiento económico (Rodríguez-Dapena [56]). En particular, se deben resaltar dos sucesos: la mayor utilización de contenedores para la carga y el desarrollo de naves más grandes y más especializadas. Respecto al transporte marítimo mundial, más de 8.000 millones de toneladas se transportan por esta vía (año 2010) para abastecer a los 6.700 millones de personas que habitan el planeta, de los cuales los contenedores superan los 500 millones de TEUs transportados. El sistema portuario español mueve más de 400 millones de toneladas (año 2011), lo que representa aproximadamente un 3% del tráfico portuario total del mundo y más de un 10% del tráfico portuario de la Unión Europea, por lo que es uno de los puntales de la economía (Lehavy et al. [41], Song et al. [61]).

La literatura sobre eficiencia en la industria portuaria es relativamente nueva ya que las primeras investigaciones aparecen a mediados de los años 90 y pobre, sobre todo si se compara con los estudios realizados en otros servicios públicos, incluido el sector del transporte, donde la mayoría de las publicaciones corresponden a publicaciones relativas al sector ferroviario y al aéreo (Serrano y Castellano [60]).

En los últimos años se ha asistido a un avance importante de los trabajos que analizan la eficiencia y la productividad del sector portuario. Los procesos de innovación tecnológica acaecidos en las industrias marítima y portuaria, y los cambios en la organización y gestión de los puertos, han condicionado una modificación en la naturaleza de las operaciones, propiciando una mayor especialización de los factores. Estos hechos han tenido un gran impacto en la productividad y en la eficiencia de las operaciones portuarias. En este sentido, la Ley 33/2010, de 5 de agosto, de modificación de la Ley 48/2003, de 26 de noviembre, de régimen económico y de prestación de servicios en los puertos de interés general, señala explícitamente, en su Preámbulo I, que: *la competitividad de nuestro sistema productivo viene condicionada por la eficacia y eficiencia de los puertos* y asimismo añade *la exigencia de adoptar en España medidas que mejoren la gestión de nuestros puertos y su eficiencia, impulsando su competitividad en una coyuntura de fuerte competencia internacional*.

El sistema portuario español se compone de 28 Autoridades Portuarias que gestionan 46 puertos de interés general en un régimen avanzado de autonomía de gestión. Las terminales se explotan mediante concesiones demaniales, pudiendo realizarse también mediante concesión de obra pública, lo que caracteriza la explotación de las terminales de contenedores con un sistema concesional mediante empresas privadas (González y Camarero [30], González y Camarero [27]). La realidad española de las terminales de contenedores es diversa y compleja. Por un lado están las terminales consolidadas con tráficos crecientes y con tipología diversa, de import export, transbordo, o atendiendo a ambos tráficos, operando con distintos equipos (reach stacker, RTG, straddle carrier...), y por otro, terminales en crecimiento que dan servicio a una zona específica. Además, las terminales de contenedores del sistema portuario español son muy heterogéneas debido a la diferente tipología de sus tráficos, la diversidad de sus equipos de manipulación y la variedad en su gestión y explotación (Camarero y González [7], Camarero y González [6]).

Ante esta situación, y con la necesidad de caracterizar las terminales de contenedores del sistema portuario español se ha investigado en la dirección de emplear modelos gráficos probabilísticos (redes bayesianas), con el objetivo de determinar los parámetros óptimos de explotación y planifi-

cación portuaria, a través de la clasificación de las terminales añadiendo la inferencia de escenarios virtuales. La tendencia general en el análisis de la explotación de una terminal de contenedores es comparar los ratios y parámetros de la bibliografía internacional sobre explotación y planificación portuaria. La bibliografía más empleada es (Consultants [19], Consultants [18], Consultants [17], Fossey et al. [24]) para terminales de contenedores. En este tipo de análisis se establecen los criterios de referencia con los que se deben explotar las terminales y se determina lo que una terminal se aleja o se acerca de la explotación de referencia. Este valor de referencia se representa con los valores de estos parámetros internacionales. Los puertos, al igual que sucede en otros modos de transporte donde la operación de los servicios se realiza a través de sistemas complejos, son infraestructuras para las cuales no resulta sencilla la definición de rendimiento (Monie [46], Monie [47]). Cada puerto está integrado por varios subsistemas interrelacionados que dan servicios a los barcos, y a los verdaderos usuarios finales que son las empresas que envían/reciben cargas a través del transporte marítimo. La planificación de las terminales, para una explotación eficiente, se realiza a medio y largo plazo y debe contener un estudio sistémico (Camarero y González [7]). En ninguno de los subsistemas que conforman la terminal deben producirse cuellos de botella que entorpezcan la operativa de la terminal. Es por ello necesario conocer la capacidad de cada uno de los subsistemas que conforman la terminal, así como el rendimiento que se produce en cada uno de ellos, para establecer cuál de ellos limita la capacidad de la misma (González y Camarero [28]). La capacidad de la terminal se encuentra condicionada por las infraestructuras, las instalaciones, los equipos y los recursos humanos participantes en cada una de las fases de la operación portuaria que se desarrollan en la terminal (Nombela [48]), es por tanto, necesario ahondar en el concepto de rendimiento portuario y más en detalle, de rendimiento de los elementos de la terminal. La capacidad de una terminal portuaria, o lo que es lo mismo su rendimiento anual en términos de material movido, es la de su sistema operativo de menor capacidad, por lo que la primera consideración que debe tenerse en cuenta a la hora de planificar y estructurar una terminal portuaria es la conjunción de sus subsistemas operativos (González [29]). La capacidad de cada subsistema operativo se encuentra condicionada por diversas variables, algunas de ellas endógenas, de competencia propia, y otras exógenas. Del amplio grupo de variables que afectan a la capacidad se pueden destacar como primordiales: variables exógenas (regularidad del trabajo y racionalidad del trabajo); variables endógenas (dimensión de la terminal y productividad que es la variable que se pretende determinar en el presente artículo). No se debe olvidar que el concepto de capacidad es un concepto teórico y el valor a adoptar es muy sensible a la gestión y explotación de la terminal. Por ello se debe analizar la capacidad que se obtiene de cada terminal estudiando al mismo tiempo la explotación que se desarrolla en la misma. En el ámbito español, la referencia monográfica más completa en materia de medición de la capacidad y niveles de servicio de terminales portuarias (Rodríguez-Pérez [57]) se remonta al año 1977, donde (Rodríguez-Pérez [57]) desarrolla un extenso trabajo sobre la materia. En estos trabajos se aborda el concepto genérico de capacidad, el de capacidad económica, las capacidades basadas en la espera y sus respectivos cálculos. También han intentado otros autores avanzar en este sentido (Pery y Camarero [54]), mediante un estudio de la línea de atraque de los puertos españoles, donde se establece el grado de ocupación de la misma, lo que es otro rendimiento de la explotación.

## 2. Estado del arte sobre eficiencia portuaria

El estudio sobre eficiencia y optimización de una terminal lleva inherente el estudio del rendimiento de la misma. Tradicionalmente el estudio de la eficiencia de un terminal se realiza mediante ratios de eficiencia obtenidos del estudio de la explotación de terminales. Estos ratios son tomados como parámetros de referencia para proyectar nuevas terminales o planificar desarrollos futuros de terminales ya existentes; por lo tanto, en el escenario de trabajo del rendimiento portuario se emplean,

habitualmente con escasa precisión o como falsos sinónimos diversos términos como tráfico, explotación óptima, capacidad, productividad, ocupación, eficiencia, etcétera. Es por tanto necesario determinar una clasificación del rendimiento portuario (Monie [47], Monie [46]), se plantean tres tipos o categorías de medición: mediciones de tráfico, mediciones de productividad y mediciones de utilización. Las mediciones de tráfico expresan volúmenes manipulados por unidad de tiempo, sin explicitar los recursos empleados. Las mediciones de productividad expresan volúmenes manipulados (producción) por unidad de recurso y por unidad de tiempo. Las mediciones de utilización son ratios (por ejemplo la tasa de ocupación de la línea de atraque), expresados en porcentaje, entre el uso de un determinado recurso y el máximo posible en un periodo temporal. Así, la capacidad de una infraestructura o equipo resulta del producto de una medición de utilización por otra de productividad para una unidad temporal, que será función del nivel de servicio requerido. Sobre el área de conocimiento del rendimiento portuario, desde mediados los años noventa se viene hablando de literatura sobre eficiencia (Clark et al. [15]), (Tongzon [68]), (Tongzon y Heng [67]) .

La eficiencia (Tongzon [68]), (Tongzon y Heng [67]) se entiende como la capacidad de alcanzar los objetivos programados con el mínimo de recursos disponible, logrando así su optimización o, alternativamente, puede expresarse como alcanzar el máximo objetivo para los recursos disponibles (González y Camarero [30]). En el contexto de este artículo y de la metodología desarrollada se trabajó sobre la definición de eficiencia de (Tongzon [68]), (Tongzon y Heng [67]).

El conjunto de categorías e indicadores tipo para medir el rendimiento portuario puede ser reclasificado atendiendo al uso u objetivo del correspondiente indicador tipo un ejemplo de ello se desarrolla en (González [29]). Así, cabe diferenciar entre mediciones de rendimiento a corto plazo, que son aquellas mediciones de rendimiento también denominadas básicas o individuales y corresponde a aquellas que caracterizan la calidad de servicio prestado a un buque o vehículo de transporte terrestre concreto, se caracterizan por qué su seguimiento por parte del control de la terminal se realiza en tiempo real, de modo que paralelamente, se pueden tomar decisiones para adecuar el servicio a las necesidades del cliente, son las inherentes al día a día de la explotación; y las mediciones de rendimiento a largo plazo también llamadas agregadas o colectivas y que son aquellas que caracterizan la calidad del servicio prestado en un periodo de tiempo (mes, año), su seguimiento y control por parte del gestor del puerto o terminal se realiza con el objetivo de la planificación de acciones a medio y largo plazo.

En el análisis de eficiencia y rendimiento portuario existen dos grandes grupos de técnicas que se han venido empleando para determinar el mismo:

- El análisis conocido como DEA (Data Envelopment Analysis), o Análisis Envoltante de Datos aplicado a tráfico portuario se trata en diversos estudios (Roll y Hayuth [58], Tongzon [66], Bonilla et al. [3]), este tipo de análisis ha sido tradicionalmente utilizado para la estimación de la eficiencia relativa de un conjunto de unidades productivas. El DEA es en origen un procedimiento no paramétrico que utiliza una técnica de programación lineal y que va a permitir la evaluación de la eficiencia relativa de un conjunto de unidades productivas homogéneas, es una técnica de medición de la eficiencia basada en la obtención de una frontera de eficiencia a partir de un conjunto de observaciones, sin necesidad de asumir ninguna forma funcional entre *input* o insumos y *ouputs* o productos. Es, en definitiva, una alternativa para extraer información de un conjunto de observaciones frente a los métodos paramétricos.. Más recientemente se ha publicado un estudio basado en DEA sobre las perspectivas de eficiencia del sistema portuario español.
- La estimación econométrica de funciones y distancias tratada para puertos en (Liu et al. [42], Tovar et al. [70], Jara-Díaz et al. [37], Tovar et al. [69], Tovar et al. [71]), es una estimación empírica de funciones de costes portuarias tiene su origen en la década de los años setenta con el trabajo de (Wanhill [73]) cuyo objetivo era diseñar un modelo que permitiera determinar el número de atraques óptimos que minimiza el coste total de uso del puerto, entendiendo como

tal la suma de dos componentes: el coste de suministrar la infraestructura (el atraque) y el coste del tiempo del barco en el puerto. En el trabajo de (Wanhill [73]) se considera que la inversión y planificación futuras han de hacerse teniendo en cuenta que los servicios portuarios no se pueden almacenar y, por tanto, hay un trade-off entre el coste de la capacidad portuaria y el coste de permanencia de los barcos en el puerto (tiempo de servicio más tiempo de espera), que es determinante y ha de ser tenido en cuenta en el proceso de planificación (Tovar et al. [69]).

La metodología de los modelos DEA supone la generalización del análisis tradicional de los ratios de actividad permitiendo considerar de forma simultánea varios inputs y/o outputs. El DEA y la estimación de funciones de frontera son alternativas para calcular la frontera de producción y, por tanto, medir la eficiencia en la producción y en los costes. A partir de estas dos técnicas se pueden obtener ratios de eficiencia relativa dentro de una muestra de unidades que se comparan con la frontera de eficiencia. La estimación de funciones frontera supone el uso de métodos econométricos (métodos paramétricos) mientras que el DEA es un método no paramétrico basado en el uso de la programación lineal.

La determinación de terminales previsiblemente más eficientes de manera que se alcancen los objetivos programados con el mínimo de recursos disponible, ofrece a los gestores y operadores de las terminales una referencia para gestión portuaria, así como un punto de referencia para una correcta planificación estratégica de las actuaciones a emprender. La jerarquización de las distintas terminales tiene interés tanto para las autoridades de coordinación, como para los clientes y operadores ya que permiten discriminar entre ellas, haciendo factible el poder priorizar aquellas con resultados mejores, es por ello que se pretende con este trabajo aplicar una nueva metodología, en este caso, inteligencia artificial, mediante el empleo de modelos gráficos probabilísticos (redes bayesianas) que permiten clasificar las diferentes terminales así como determinar los parámetros óptimos de explotación portuaria, añadiendo la inferencia de escenarios virtuales que permiten al usuario pueda interaccionar de forma efectiva. El objetivo de la inferencia probabilística será calcular la distribución de probabilidad a posteriori de un conjunto de variables dada la observación de un evento (valores observados para un subconjunto de variables)

En las últimas décadas se han desarrollado numerosas técnicas de análisis y modelización de datos en distintas áreas de la estadística y la inteligencia artificial (Duda y Hart [21], Bishop [2]). La Minería de Datos (MD) es un área moderna interdisciplinar que engloba a aquellas técnicas que operan de forma automática (requieren de la mínima intervención humana) y, además, son eficientes para trabajar con las grandes cantidades de información disponibles en las bases de datos de numerosos problemas prácticos. Estas técnicas permiten extraer conocimiento útil (asociaciones entre variables, reglas, patrones, etc.) a partir de la información cruda almacenada, permitiendo así un mejor análisis y comprensión del problema. En algunos casos, este conocimiento puede ser también post-procesado de forma automática permitiendo obtener conclusiones, e incluso tomar decisiones de forma casi automática, en situaciones prácticas concretas (sistemas inteligentes). La aplicación práctica de estas disciplinas se extiende a numerosos ámbitos comerciales y de investigación en problemas de predicción, clasificación o diagnosis (Cios et al. [14], Witten y Frank [74]). La aplicación de la minería de datos ha sido profusa en varias disciplinas, como en la biología (Witten y Frank [74]), economía (Vityaev y Kovalerchuk [72]), medicina (Cios y Kacprzyk [13]) etcétera.

Se pretende aplicar Minería de Datos (Felgaer et al. [23], Perichinsky et al. [53]) con el objeto de desarrollar un conjunto de técnicas y herramientas aplicadas al proceso no trivial de extraer y presentar conocimiento implícito, previamente desconocido, potencialmente útil y humanamente comprensible, a partir de grandes conjuntos de datos, con objeto de predecir de forma automatizada tendencias y comportamientos; y describir de forma automatizada modelos previamente desconocidos (Chen et al. [11], Fayyad et al. [22], Felgaer et al. [23], Mannila [43]). El término Minería de Datos Inteligente (Felgaer et al. [23], Han et al. [31], Michalski et al. [45]) se refiere es-

pecíficamente a la aplicación de métodos de aprendizaje automático (Felgaer et al. [23], Holsheimer et al. [35], Michalski [44]), para descubrir y enumerar patrones presentes en los datos, para estos, se desarrollaron un gran número de métodos de análisis de datos basados en la estadística (Felgaer et al. [23], Michalski [44]). En la medida en que se incrementaba la cantidad de información almacenada en las bases de datos, estos métodos empezaron a enfrentar problemas de eficiencia y escalabilidad y es aquí donde aparece el concepto de minería de datos. Una de las diferencias entre al análisis de datos tradicional y la minería de datos es que el primero supone que las hipótesis ya están construidas y validadas contra los datos, mientras que el segundo supone que los patrones e hipótesis son automáticamente extraídos de los datos (Felgaer et al. [23], Orallo et al. [50]). Las tareas de la minería de datos se pueden clasificar en dos categorías: minería de datos descriptiva y minería de datos predicativa (Fayyad et al. [22]).

Entre las diferentes técnicas disponibles en minería de datos las redes probabilísticas o redes bayesianas permiten modelizar de forma conjunta toda la información relevante para un problema dado, utilizando posteriormente mecanismos de inferencia probabilística para obtener conclusiones en base a la evidencia disponible (Pearl [52]), Castillo et al. [8]). Estos modelos se adaptan de forma natural a la concepción probabilística y local actual de la modelización de la eficiencia portuaria.

Se puede decir que una Red Bayesiana es un conjunto de nodos y arcos. Cada nodo corresponde a una variable, que a su vez representa una entidad del mundo real, y los arcos que unen los nodos indican relaciones de influencia causal entre las variables. Una Red Bayesiana es un grafo acíclico dirigido en el que cada nodo representa una variable y cada arco una dependencia probabilística, en la cual se especifica la probabilidad condicional de cada variable dados sus padres (Castillo et al. [8]). La variable a la que apunta el arco es dependiente (causa-efecto) de la que está en el origen de éste. La topología o estructura de la red ofrece información sobre las dependencias probabilísticas entre las variables y sus dependencias condicionales dada otra(s) variable(s). Dichas dependencias, simplifican la representación del conocimiento (menos parámetros) y el razonamiento (propagación de las probabilidades).

La potencia de las Redes Bayesianas radica en que, una vez especificada la estructura de la red es posible realizar cualquier tipo de inferencia dada la información disponible. Es decir, se puede hacer inferencias predictivas (si la terminal de transporte tiene una superficie X, ¿cuál es la probabilidad de que tenga Z grúas de patio en el almacenamiento?) o abductivas (si la terminal tiene menos de X grúas de patio en el almacenamiento, ¿cuál es la probabilidad de que conozca el concepto superficie de almacenamiento?). De esta forma, un mismo nodo puede ser tanto fuente de información como objeto de predicción. Dichas inferencias se realizan aplicando algoritmos de propagación de probabilidades que se han desarrollado específicamente para tal fin. El primer método de propagación para redes bayesianas que se desarrolló es el algoritmo de propagación en árboles (Pearl [51]). La idea consiste en que cuando se modifica la información asociada a un nodo, éste traspasa la información a sus nodos vecinos mediante un conjunto de mensajes; estos nodos, a su vez, procesan la información recibida junto con la que ellos poseen y la pasan a sus nodos vecinos (aún no modificados) y así sucesivamente hasta que todos los nodos han actualizado su información.

Para utilizar una Red Bayesiana se deben identificar las variables y las relaciones de influencia causal entre ellas, y cuantificar estas relaciones de influencia causal asignando las probabilidades condicionadas y a priori. Una vez definida la red se puede utilizar alguno de los paquetes de software comerciales o de libre disposición como HUGIN [36], Norsys [49], Sotolongo y Valdivia [62], Chavez et al. [10] para realizar las inferencias que interesen. En el modelo que se propone se utiliza el software Consortium [16].

Las redes bayesianas son una representación compacta de una distribución de probabilidad multivariante. Formalmente, una red bayesiana es un grafo dirigido acíclico donde cada nodo representa una variable aleatoria y las dependencias entre las variables quedan codificadas en la propia estructura del grafo según el criterio de d-separación. Asociada a cada nodo de la red hay una distribución de probabilidad condicionada a los padres de ese nodo, de manera que la distribución conjunta factoriza como el producto de las distribuciones condicionadas asociadas a los nodos de la red. Es decir, para una red con  $n$  variables  $X_1, X_2, \dots, X_n$  (ecuación 1).

$$\prod_{i=1}^n p(x_i | pa(x_i)) \quad (1)$$

Las redes probabilísticas automatizan el proceso de modelización probabilística utilizando la expresividad de los grafos. Los modelos resultantes combinan resultados de la teoría de grafos (para representar las relaciones de dependencia e independencia del conjunto de variables) y de la probabilidad (para cuantificar estas relaciones). Esta unión permite realizar de forma eficiente tanto el aprendizaje automático del modelo, como la inferencia a partir de la evidencia disponible. La base de conocimiento de estos sistemas es una estimación de la función de probabilidad conjunta de todas las variables del modelo, mientras que el módulo de razonamiento es donde se hace el cálculo de probabilidades condicionadas.

El estudio de esta técnica proporciona una buena perspectiva global del problema del aprendizaje estadístico y la minería de datos, y permite entender mejor otras técnicas alternativas (Hastie et al. [33], Duda y Hart [21]).

Las redes bayesianas están siendo empleadas dentro de los sistemas de transporte, preferentemente, para desarrollar actuaciones en carreteras, como (Sun et al. [63]) que emplea las redes bayesianas para predecir flujos de vehículos, así como (Tebaldi y West [64]). Dentro de la planificación se puede encontrar (Bromley et al. [4]) donde se desarrolla una gestión integrada de recursos hídricos (Cain [5]) o mejoras en la planificación en la gestión de los recursos naturales.

Las redes bayesianas proveen una forma compacta de representar el conocimiento y métodos flexibles de razonamiento, basados en las teorías probabilísticas, capaces de predecir el valor de variables no observadas y explicar las observadas. Entre las características que poseen las redes bayesianas, se puede destacar que permiten aprender sobre relaciones de dependencia y causalidad, permiten combinar conocimiento con datos (Díaz y Corchado [20], Heckerman [34], Orallo et al. [50]) y pueden manejar bases de datos incompletas (Chickering y Heckerman [12], Heckerman [34], Ramoni y Sebastiani [55]).

### 3. Caso de estudio

Para la consecución del objetivo de determinar la eficiente en las terminales de contenedores del Sistema Portuario Español mediante redes bayesianas se ha desarrollado la siguiente metodología que se divide en dos grandes tareas: una para determinar el escenario de trabajo y la segunda para desarrollar el modelo de inteligencia artificial.

#### **Determinación del escenario de trabajo**

##### *Diagnosis y estado del arte*

Consiste en la revisión del estado del arte para identificar el conjunto de variables de medida de la explotación en terminales marítimas de contenedores, mediante el empleo de buscadores especializados y gestores de aplicaciones. Se desarrolla en dos etapas:

1. Determinación de las variables de la terminal y selección de las mismas: se realiza un estudio de todas las variables susceptibles de investigación.
2. Obtención del valor de las variables para las distintas terminales de estudio: una vez seleccionadas las terminales de estudio que se quieren obtener se utilizan diferentes fuentes de información para obtener los valores de las mismas.

Las variables (rasgos) que intervienen en el cálculo de las probabilidades tanto a priori como las condicionales son las variables. El número de variables (rasgos) conforma una tabla de  $2n$  combinaciones, donde  $n$  es un número natural que puede ser considerablemente grande; lo cual constituye una dificultad, cuya solución pudiera consistir en reducir el espacio de representación inicial, de forma tal que si existen variables superfluas, se analice si se mantienen o no, según su importancia desde el punto de vista metodológico.

Una alternativa de solución al problema de la selección de variables, es el uso del conjunto de testores típicos Ruiz-Shulcloper y Lazo [59]. Los testores típicos son el número de variables (atributos o rasgos) con los cuales se deben describir los objetos (cuestionarios) que inciden de manera determinante en un problema. En el modelo propuesto se aplica el algoritmo LEX Alganza y Porrata [1] para obtener los testores típicos.

El estudio se ha realizado para las terminales de contenedores del sistema portuario español. La mayoría de las terminales de contenedores del sistema portuario español son terminales pequeñas, es decir, mueven menos de un millón de TEUs, tan sólo el 12% de las mismas mueven volumen superior a un millón. Destaca el hecho de que el 64% de las terminales mueva menos de 500.000 TEUs. Con el escenario de crisis actual esta tendencia sigue siendo representativa de las terminales de contenedores del sistema portuario español. Los datos de partida representan el volumen de tráfico movido por tipologías. El parámetro de tráfico fundamental se mide en TEUs anuales movidos por la terminal, pero también se necesita conocer más desagregado el número contenedores de 20 y 40 pies, el número de contenedores frigoríficos y de mercancías peligrosas, el tráfico import/export y trasbordo y de contenedores llenos y vacíos.

Las variables empleadas para la investigación han sido las que se recogen en el cuadro 1.

Cuadro 1. Variables de estudio.

Tipo de variable	Definición
Tipología de tráfico	Porcentaje de contenedores import/export
	Porcentaje de contenedores 20-40 pies
	Porcentaje de contenedores llenos/vacíos
Subsistema atraque-carga/descarga	Metros de línea de atraque
	Calado en metros
	Número total de grúas de muelle
Subsistema almacenamiento	Superficie de la zona de operación en m <sup>2</sup>
	Estancia media del contenedor en días
	Número de huellas del área de almacenamiento
Subsistema entrega y recepción	Superficie de la zona de almacenamiento en m <sup>2</sup>
	Superficie de la zona de servicios en m <sup>2</sup>

### Construcción del modelo de inteligencia artificial

Obtener una red bayesiana a partir de datos es un proceso de aprendizaje que se divide en dos etapas: el aprendizaje estructural y el aprendizaje paramétrico Pearl [52]. La primera de ellas,

consiste en obtener la estructura de la red bayesiana, es decir, las relaciones de dependencia e independencia entre las variables involucradas. La segunda etapa, tiene como finalidad obtener las probabilidades a priori y condicionales requeridas a partir de una estructura dada. A continuación se describe la discretización de las variables, la construcción del modelo y la inferencia y clasificación.

#### *Discretización de variables*

Una vez seleccionadas las variables de estudio es necesario, para el proceso de construcción de los modelos, la discretización de las variables. Normalmente las redes bayesianas consideran variables discretas o nominales, por lo que si no lo son, hay que discretizarlas antes de construir el modelo. Aunque existen modelos de redes bayesianas con variables continuas, éstos están limitados a variables gaussianas y relaciones lineales. Los métodos de discretización se dividen en dos tipos principales: (i) no supervisados y (ii) supervisados, así se estudian diferentes tipos de discretizaciones y será una opción del software desarrollado. Para la aplicación de redes bayesianas al estudio, se va a trabajar con las variables obtenidas al determinar el escenario de trabajo, dichas variables son variables discretas para lo cual las variables tomadas que son continuas se han discretizado en los intervalos determinados por los percentiles 25,50 y 75 cuyo valores están representado en el cuadro 2 separados consecutivamente por ':':.

Cuadro 2. Discretización de las variables de estudio.

Variable	Discretización
Tráfico total en TEU	95760 : 240587 : 713875
Porcentaje de contenedores import/export	62 : 93 : 100
Porcentaje de contenedores 20/40 pies	47 : 49 : 56
Porcentaje de contenedores llenos/vacíos	56 : 68 : 71
Metros de línea de atraque	360 : 713.5 : 1325
Calado en metros	12 : 14 : 16
Número total de grúas de muelle	3:4:9
Superficie de la zona de operación en m2	13500 : 35749 : 54860
Estancia media del contenedor en días	7 : 8.5 : 10
Número de huellas del área de almacenamiento	1224 : 1960 : 5016
Superficie de la zona de almacenamiento en m2	55235 : 240587 : 293125
Superficie de la zona de servicios en m2	0 : 2256 : 33223

#### *Construcción del modelo*

En esta parte del proceso, el aprendizaje estructural consiste en encontrar las relaciones de dependencia entre las variables, de forma que se pueda determinar la topología o estructura de la red bayesiana. De acuerdo al tipo de estructura, se aplican diferentes métodos de aprendizaje estructural: aprendizaje de árboles, aprendizaje de poli-árboles, aprendizaje de redes multiconectadas, métodos basados en medidas y búsqueda, métodos basados en relaciones de dependencia.

Para la construcción del modelo, en esta investigación se ha empleado el clasificador Naive-Bayes. este clasificador es uno de los más efectivos, en el sentido de que su capacidad predictiva es competitiva con el resto de clasificadores existentes, es el llamado *Naive-Bayes*, descrito, por ejemplo, por Duda y Hart en 1973 [15] y por Langley, Iba and Thompson en 1992 [26]. Este clasificador (??) aprende de un conjunto de entrenamiento la probabilidad condicional de cada atributo  $X_i$  dada la clase  $C$ . La clasificación se hace entonces aplicando la regla de Bayes para

calcular la probabilidad de  $C$  dadas las instancias de  $X_1, X_2, \dots, X_n$  y la tomando como clase predicha la de mayor probabilidad a posteriori. Estos cálculos se basan en una fuerte suposición de independencia: todos los atributos  $X_i$  son condicionalmente independientes dado el valor de la clase  $C$ . El paradigma clasificatorio en el que se utiliza el teorema de Bayes en conjunción con la hipótesis de independencia condicional de las variables predictoras dada la clase que se conoce bajo diversos nombres que incluye los de idiota Bayes ( Hand [32]), (Larrañaga e Inza [40]), naïve Bayes (Kononenko [39]), simple Bayes (Gammerman y Thatcher [25]) y Bayes independiente (Todd et al. [65]).

A pesar una larga tradición en la comunidad de reconocimiento de patrones (Duda y Hart [21]) el clasificador Naïve Bayes aparece por primera vez en la literatura del aprendizaje automático a finales de los ochenta (Cestnik et al. [9]) con el objetivo de comparar su capacidad predictiva con la de métodos más sofisticados. De manera gradual los investigadores de esta comunidad de aprendizaje automático se han dado cuenta de su potencialidad y robustez en problemas de clasificación supervisada. El paradigma naïve Bayes, el cual debe su nombre a las hipótesis tan simplificadoras: independencia condicional de las variables predictoras dada la variable clase, sobre las que se construye dicho clasificador.

Dado un ejemplo  $x$  representado por  $k$  valores el clasificador naïve Bayes se basa en encontrar la hipótesis más probable que describa a ese ejemplo. Si la descripción de ese ejemplo viene dada por los valores  $a_1, a_2, \dots, a_n$ , la hipótesis más probable será aquella que cumpla (ecuación 1):

$$v_{\text{MAP}} = \operatorname{argmax}_{v_j \in V} P(v_j | a_1, \dots, a_n) \quad (2)$$

es decir, la probabilidad de que conocidos los valores que describen a ese ejemplo, éste pertenezcan a la clase  $v_j$  (donde  $v_j$  es el valor de la función de clasificación  $f(x)$  en el conjunto finito  $V$ ).

Por el teorema de Bayes (ecuación 3):

$$v_{\text{MAP}} = \operatorname{argmax}_{v_j \in V} \frac{P(v_j | a_1, \dots, a_n) p(v_j)}{P(a_1, \dots, a_n)} = \operatorname{argmax}_{v_j \in V} P(v_j | a_1, \dots, a_n) p(v_j) \quad (3)$$

Podemos estimar  $P(v_j)$  contando las veces que aparece el ejemplo  $v_j$  en el conjunto de entrenamiento y dividiéndolo por el número total de ejemplos que forman este conjunto. Para estimar el término  $P(a_1, \dots, a_n | v_j)$ , es decir, las veces en que para cada categoría aparecen los valores del ejemplo  $x$ , debe recorrer todo el conjunto de entrenamiento. Este cálculo resulta impracticable para un número suficientemente grande de ejemplos por lo que se hace necesario simplificar la expresión. Para ello se recurre a la hipótesis de independencia condicional con el objeto de poder factorizar la probabilidad. Esta hipótesis dice lo siguiente:

Los valores  $a_j$  que describen un atributo de un ejemplo cualquiera  $x$  son independientes entre sí conocido el valor de la categoría a la que pertenecen. Así la probabilidad de observar la conjunción de atributos  $a_j$  dada una categoría a la que pertenecen es justamente el producto de las probabilidades de cada valor por separado (ecuación 4):

$$P(v_j | a_1, \dots, a_n) p(v_j) = \prod_i P(a_i | v_j) \quad (4)$$

Las redes Bayesianas llevan tiempo incorporándose a tareas de clasificación supervisada pero no al caso de planificación en lo relacionado con los puertos, en base a ideas expuestas por Michalski [44], Orallo et al. [50], y ampliadas en Pearl [52] según las cuales se pueden utilizar las factorizaciones de probabilidad representadas por las redes Bayesianas para realizar clasificaciones, considerando para ello la existencia de una variable especial, la variable a clasificar, que viene a

ser predicha por un grupo de variables, el resto, y de forma que la estructura de la red obtenida puede ser utilizada para la predicción del valor de la clase de esta variable a clasificar, mediante la asignación de valores a las predictoras, y la posterior propagación de la evidencia introducida en la red, esto es, mediante el cálculo de la probabilidad a posteriori del nodo asociado a la variable especial dados los valores del resto.

Para la construcción de la red se ha utilizado el software Elvira específico para trabajar con redes bayesianas Consortium [16]. El programa Elvira cuenta con un formato propio para la codificación de los modelos, un lector-intérprete para los modelos codificados, una interfaz gráfica para la construcción de redes, con opciones específicas para modelos canónicos, algoritmos exactos y aproximados (estocásticos) de razonamiento tanto para variables discretas como continuas, métodos de explicación del razonamiento, algoritmos de toma de decisiones, aprendizaje de modelos a partir de bases de datos, fusión de redes, etc.

### *Inferencia y clasificación*

Una vez construidos los modelos, en esta fase se estudia la capacidad de inferencia. Una red bayesiana proporciona un sistema de inferencia, donde una vez encontradas nuevas evidencias sobre el estado de ciertos nodos, se modifican sus tablas de probabilidad; y a su vez, las nuevas probabilidades son propagadas al resto de los nodos. La propagación de probabilidades o inferencia probabilística es la probabilidad de algunas variables de ser calculadas dadas evidencias en otras variables. Las probabilidades antes de introducir evidencias se conocen como probabilidades a priori; una vez introducidas las evidencias, las nuevas evidencias propagadas se llaman probabilidades a posteriori.

En esta fase se aprovechan las características que poseen los métodos bayesianos en tareas de aprendizaje. Cada ejemplo observado va a modificar la probabilidad de que la hipótesis formulada sea correcta (aumentándola o disminuyéndola). Es decir, una hipótesis que no concuerda con un conjunto de ejemplos más o menos grande no es desechada por completo sino que lo que harán será disminuir esa probabilidad estimada para la hipótesis. Estos métodos son robustos al posible ruido presente en los ejemplos de entrenamiento y a la posibilidad de tener entre esos ejemplos de entrenamiento datos incompletos o posiblemente erróneos. Los métodos bayesianos nos permitirán tener en cuenta en la predicción de la hipótesis el conocimiento a priori o conocimiento del dominio en forma de probabilidades.

La red construida con el clasificador Naive Bayes es la que aparece en la figura 1, en la que se incluye la probabilidad a priori.

A partir de la red construida mediante inferencia se pueden obtener las probabilidades que pueden adoptar las discretizaciones de las diferentes variables, ello permite obtener escenarios eficientes para las terminales de contenedores de manera que si la eficiencia (Tongzon [68]), (Tongzon y Heng [67]) se entiende como la capacidad de alcanzar los objetivos programados con el mínimo de recursos disponible, logrando así su optimización, se obtienen diferentes escenarios para los diferentes volúmenes de tráfico a mover (objetivo a alcanzar) minimizando los recursos disponibles: grúas, superficies, huellas, ...

Los diferentes escenarios se encuentran descritos en la figura 3.

Cada uno de los escenarios con las diferentes probabilidades que adopta cada variable se describen en las figuras siguientes ( figura2, figura3, figura4, figura5). Las discretizaciones se han realizado en intervalos determinados por los percentiles 25,50 y 75 cuyo valores están representado en el cuadro 2 separados consecutivamente por ':', se representan por  $s1$ , para la primera discretización percentil 25;  $s2$ , para la discretización percentil 50;  $s3$ , para la discretización percentil 75.

## **Discusión de los resultados**

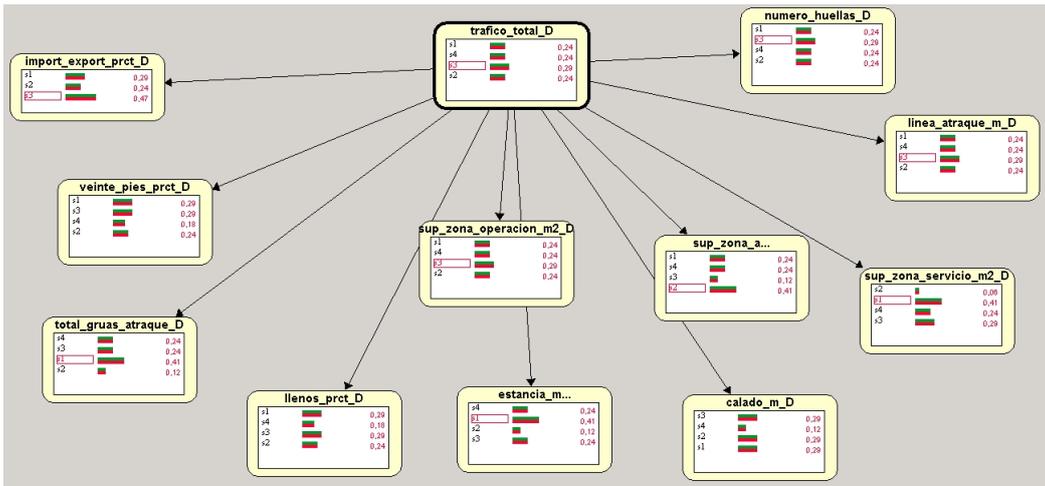


Figura 1. Red bayesiana, clasificador Naive Bayes. Probabilidad a priori.

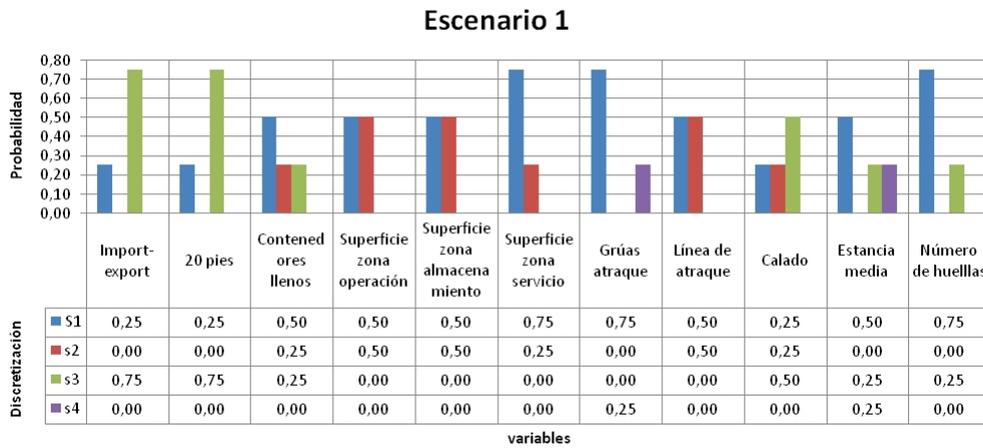


Figura 2. Escenario 1, clasificador Naive Bayes. Inferencia.

Respecto a la tipología de los tráficos, en el caso de operativa con contenedores import/export, ésta se produce en los Escenarios 1, 2 y 3; el Escenario 4 tiene mayor probabilidad de operar contenedores de trasbordo, los contenedores de trasbordo modifican la explotación dado que el contenedor llega en un buque y se va en otro, lo que modifica la operativa de la terminal. Respecto a la relación entre contenedores de 20 o 40 pies, la realidad es que la discretización ha dado como resultado horquillas muy similares, por lo que en los cuatro escenarios se puede hablar de un reparto casi al 50%. La relación entre contenedores llenos y vacíos si presenta diferencias a destacar. El transporte marítimo movió el 33,5% del total de la mercancía contenerizada de 59,3 millones de UTIs en la UE-27, el 70,4% eran contenedores de 40 pies, el 26,4% contenedores de 20 pies y el 3,9% otro tipo de contenedores (Jiménez [38]). El Escenario 1, con menor volumen de tráfico tiene un porcentaje de contenedores llenos inferior, por lo tanto superior de vacíos, lo interesante para una terminal es operar contenedores llenos, los vacíos tienen el destino más incierto por lo que hasta que se les asigna nuevo destino pasan mayor tiempo en la terminal con la correspondiente ralentización de la operativa general.

Cuadro 3. Previsión de escenarios eficientes para diferentes volúmenes de tráfico a mover.

Variable	Escenario 1: tráfico total menor de 95760 TEU	Escenario 2: tráfico total entre 95760 y 240587 TEU	Escenario 3: tráfico total entre 240587 y 713875 TEU	Escenario 4: tráfico total mayor de 713875
Porcentaje de contenedores import/export	entre 93-100	entre 93-100	entre 62-93	menor de 62
Porcentaje de contenedores 20/40 pies	entre 49 y 56	menor de 49	mayor de 56	menor de 47
Porcentaje de contenedores llenos/vacíos	menor de 56	menor de 68	entre 68-71	mayor de 71
Metros de línea de atraque	menor de 713.5	entre 360 y 713.5	entre 713.5 y 1325	mayor de 1325
Calado en metros	entre 14 y 16	menor de 12	menor de 12	entre 12 y 14
Número total de grúas de muelle	menor de 3	menor de 3	menor de 9	mayor de 9
Superficie de la zona de operación en m2	menor de 35749	menor de 13500	entre 35749 y 54860	mayor de 54860
Estancia media del contenedor en días	menor de 7	menor de 7	menor de 7 y entre 8.7 y 10	-
Número de huellas del área de almacenamiento	menor de 1224	entre 1224 y 1960	entre 1960 y 5016	entre 1960 y 5016
Superficie de la zona de almacenamiento en m2	menor de 240587	menor de 240587	entre 55235 y 240587	mayor de 293125
Superficie de la zona de servicios en m2	sin zona de servicio	entre 2256 y 33223	mayor de 2256	sin zona de servicio o mayor de 33223

Si se analizan las distintas superficies de la terminal, en el caso de la superficie de la zona de operación se observa que los escenarios que mueven un volumen de tráfico menor necesitan menos superficie de operación y los que mueven un volumen mayor necesitan superficies mayores como se prevé a priori. Respecto a la superficie de almacenamiento, el Escenario 1 y el Escenario 2 tienen la misma probabilidad para la variable superficie almacenamiento, a mayor volumen más superficie y en el caso del Escenario 4, la probabilidad es 1 para el caso de la discretización de mayor superficie. En el caso de la zona de servicio los resultados son diferentes debido a que muchas terminales que se encuentran saturadas en la zona de almacenamiento han ido ocupando la zona de servicios para cubrir estas carencias.

De los resultados obtenidos se puede concluir que respecto a las variables del subsistema

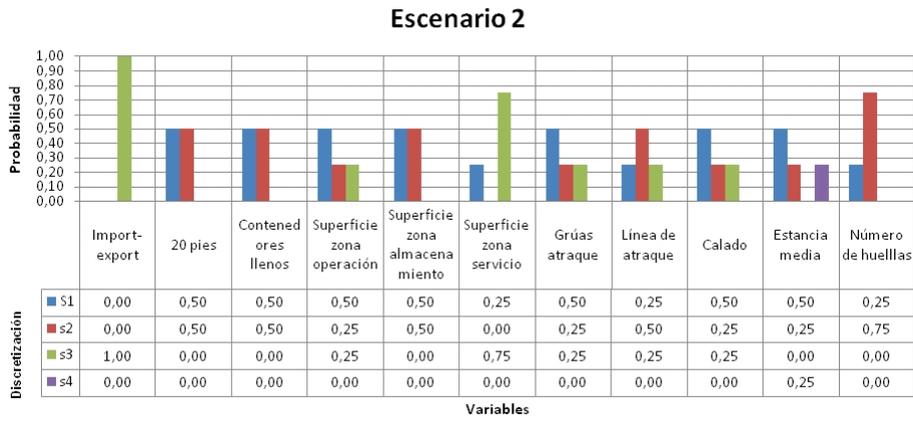


Figura 3. Escenario 2, clasificador Naive Bayes. Inferencia.

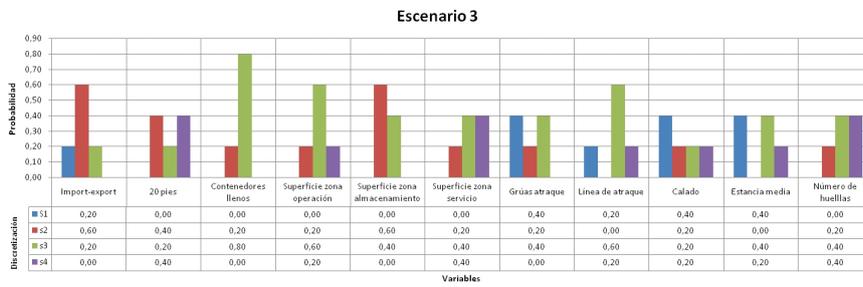


Figura 4. Escenario 3, clasificador Naive Bayes. Inferencia.

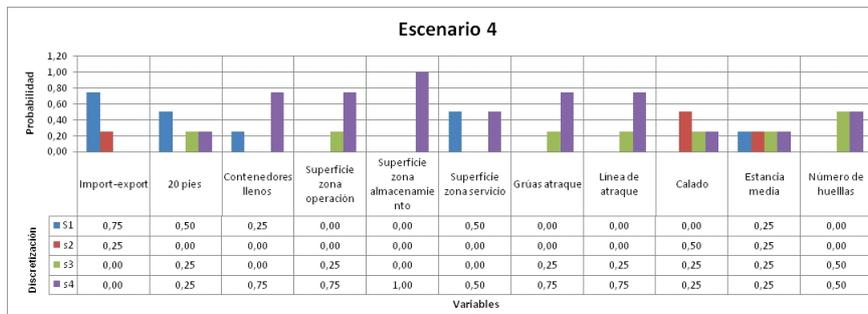


Figura 5. Escenario 4, clasificador Naive Bayes. Inferencia.

atraque-carga/descarga, en primer lugar respecto al calado, un volumen de tráfico menor no se traduce en un calado menor, si el calado del puerto es elevado, mayor de 16 metros, y es un calado natural, es óptimo pues no se ha necesitado de obras de dragado, si la obtención del calado es con obras de ingeniería, el coste es muy elevado y no es óptimo. Para anticiparse a los futuros buques portacontenedores denominados megabuques, como corresponde a la Clase E Maersk Line con capacidad para 18000 TEU (González y Camarero [30]), se necesitan calados de 15,5 metros (González y Camarero [26]), por lo que el Escenario 1, presenta mayor probabilidad de poder

atender a dichos buques. Una vez realizada la inferencia en la red bayesiana se puede destacar que el Escenario 4 tiene una probabilidad del 25 % de poder atender a dichos buques, si bien el 50 % de la probabilidad es de poder atender buques con necesidades de calado entre 12 y 14 metros. Las limitaciones de calado no ralentizan la operativa en el atraque, o se puede operar o no se puede. Respecto a la línea de atraque a mayor volumen de tráfico es necesaria mayor línea de atraque, pudiéndose observar que para el Escenario 4, con un 75 % de probabilidad la línea de atraque debe ser mayor de 1.325 metros, el Escenario 3 contempla probabilidades en tres estratos de discretización por lo que parece que determinadas terminales al ser capaces de optimizar la operativa y mover el mismo volumen con menores recursos.

Respecto al número de grúas en el atraque, tanto el Escenario 1, como el Escenario 2 y el Escenario 3 tienen un 50 % de probabilidad de operar con menos de 3 grúas en el atraque, por lo que son capaces de mover tráficos con necesidad de menores recursos. El Escenario 4 presenta una probabilidad del 75 % de operar con más de 9 grúas.

Respecto al número de huellas para disponer los contenedores en la zona de almacenamiento, el Escenario 1 y el Escenario 2 tienen una probabilidad de 75 % de disponer de menos de 1.224 huella y entre 1.224 y 1.960 respectivamente, a mayor volumen, mayor espacio de almacenamiento y mayor número de huellas.

Las terminales de contenedores del sistema portuario español parece que siguen el modelo de optimización, de manera que se alcanzan los objetivos programados con el mínimo de recursos disponible, se cumple en el caso de mover mayor tráfico minimizando los recursos de explotación, para las cuatro escenarios propuestos parece ser que las terminales estén llegando a un cierto nivel de colapso en la operativa, al emplear los recursos tan al límite.

#### 4. Conclusiones

El enfoque estocástico de las redes bayesianas en contraposición con el determinista del análisis DEA permite contrastar hipótesis estadísticas y proponer diferentes escenarios para los casos de estudio. Al contrario que el análisis DEA, las redes bayesianas proporcionan una distribución de probabilidad que permite realizar supuestos. Otra ventaja de las redes es su capacidad para trabajar con un número grande de variables y ser capaz de tratar correctamente los outliers.

Las redes Bayesianas tienen un número de características que hacen que sean apropiadas para el apoyo a los gestores de infraestructuras portuarias. Entre las más destacadas se recogen las siguientes:

Por su naturaleza, las redes Bayesianas proveen una representación gráfica de las relaciones explícitas de dependencia del dominio. Generalmente las variables en la explotación portuaria están influenciados por muchos factores. Las redes Bayesianas permiten modelar estos sistemas complejos permitiendo a los gestores entender las relaciones causales visualizándolas por medio del grafo.

Las redes Bayesianas están formadas por un componente cualitativo, un grafo, y una parte cuantitativa, las tablas de probabilidades, que permiten utilizar criterios objetivos y subjetivos, por ejemplo, utilizando creencias de expertos en explotación y gestión portuaria.

Las redes Bayesianas pueden hacer inferencia en ambos sentidos, es decir, las variables de entrada pueden ser usadas para predecir las variables de salida y viceversa. Fijando las variables de salida con los valores deseados, es posible predecir qué valores de las variables de entrada permiten dicha salida. Por ejemplo, usando inferencia hacia adelante, se puede predecir el 'número grúas' basándose en variables como 'tráfico movido'.

Dado un conjunto de evidencias, las redes Bayesianas permiten fácilmente calcular la sensibilidad de ciertas variables, simplemente modificando las evidencias.

Las redes bayesianas pueden modelar grados de certidumbre, en vez de valores exactos. Por

tanto, permiten modelar la incertidumbre de manera efectiva y explícitamente, por lo que pueden realizar buenas predicciones con información incompleta, las estimaciones son una evaluación probabilística de un suceso futuro, lo que permite a los gestores trabajar con información incompleta de la explotación portuaria.

La salida de una red Bayesiana es una distribución de probabilidad en vez de valores únicos. Este tipo de información se puede usar para medir la confianza que se puede depositar en la salida de la red Bayesiana, lo cual es esencial si el modelo va a ser usado en la toma de decisiones. Por ejemplo, en una variable con estados bajo, medio, y alto, como es el tráfico de la terminal de contenedores, la red Bayesiana estima la probabilidad de cada uno de los estados.

### Referencias bibliográficas

1. Alganza, Y. S., Porrata. A. P. (2003). Lex: Un nuevo algoritmo para el cálculo de los testores típicos. *REVISTA CIENCIAS MATEMATICAS, Santiago de Cuba, CUBA*, 21(1): 2–3.
2. Bishop, C. M. (2006). *Pattern recognition and machine learning*, volume 4. Springer New York, SpringerLink (Online service).
3. Bonilla, M., Casasús, T., Medal, A., Sala, R. (2004). An efficiency analysis with tolerances of the spanish ports system. *International Journal of Transport Economics*, 32(3):380–400.
4. Bromley, J., Jackson, N. A., Clymer, O. J., Giacomello, A. M., Jensen, F. V. (2005). The use of hugin to develop bayesian networks as an aid to integrated water resource planning\* 1. *Environmental Modelling and Software*, 20(2):231–242.
5. Cain, J. (2001). Planning improvements in natural resource management. guidelines for using bayesian networks to support the planning and management of development programmes in the water sector and beyond.
6. Camarero, A., González, N. (2005). *Cadenas integradas de transporte*. Fundación Agustín de Betancourt. Ministerio de Fomento, Madrid, España.
7. Camarero, A., González, N. (2007). *Logística y transporte de contenedores*. Fundación Agustín de Betancourt. Ministerio de Fomento, Madrid, España.
8. Castillo, E., Gutiérrez, J. M., Hadi, A. S. (1997). *Expert systems and probabilistic network models*. Springer Verlag.
9. Cestnik, B., Kononenko, I., Bratko, I. (1987). Assistant 86: A knowledge elicitation tool for sophisticated users. *Progress in machine learning*, 31:45.
10. Chavez, M., Grau, R., García, M. (1999). Un método para construir redes bayesianas. *Revista de Ingeniería de la Universidad de Antioquia*.
11. Chen, M. S., Han, J., Yu, P. S. (1996). Data mining: an overview from a database perspective. *Knowledge and data Engineering, IEEE Transactions on*, 8(6):866–883.
12. Chickering, D., Heckerman, D. (1996). Efficient approximations for the marginal likelihood of incomplete data given a bayesian network. In *Proc. Twelfth Conference on Uncertainty in Artificial Intelligence (UAI96)*, pages 158–168.
13. Cios, K. J. and Kacprzyk, J. (2001). *Medical data mining and knowledge discovery*. Physica-Verlag.
14. Cios, K. J., Pedrycz, W., Swiniarski, R. W. (2007). *Data mining: a knowledge discovery approach*. Springer Verlag.
15. Clark, X., Dollar, D., Micco, A. (2004). *Port efficiency, maritime transport costs and bilateral trade*.
16. Consortium, E. (2002). Elvira: An environment for probabilistic graphical models. proceedings of the first european workshop on probabilistic graphical models. Proceedings of the First European Workshop on Probabilistic Graphical Models (PGM'02), pages 222–222–230.
17. Consultants, D. S. (1998). World container terminals: Global growth and private profit.

18. Consultants, D. S. (2006a). The drewry container market quarterly, vol. 7. *March, London*.
19. Consultants, D. S. (2006b). The drewry container shipper insight fourth quarter 2006. *Drewry Shipping Consultants: London. 83pp*.
20. Diaz, F., Corchado, J. (1999). Rough sets bases learning for bayesian networks. In *International workshop on objective bayesian methodology, Valencia, Spain*.
21. Duda, R. O., Hart, P. E. (1973). Patten classification and scene analysis. *John Wiley and Sons, I, 973*.
22. Fayyad, U., Piatetsky-Shapiro, G., Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI magazine, 17(3):37*.
23. Felgaer, P., Britos, P., Sicre, J., Servetto, A., García-Martínez, R., Perichinsky, G. (2003). Optimización de redes bayesianas basada en técnicas de aprendizaje por instrucción. In *Proceedings del VIII Congreso Argentino de Ciencias de la Computacion. Pag, volume 1687*.
24. Fossey, J., Busch, E. M., Consultants, D. S. (2004). *The Drewry Annual Container Market Review and Forecast 2004-05: Incorporating the Drewry Container Market Quaterly-September 2004*. Drewry Shipping Consultants.
25. Gammerman, A., Thatcher, A. (1991). Bayesian diagnostic probabilities without assuming independence of symptoms. *Methods of information in medicine, 30(1):15-22*.
26. González, M., Camarero, A. (2009a). Caracterización de los parámetros de diseño de terminales portuarias de contenedores. *Revista de Obras Públicas, (3501):39-50*.
27. González, M., Camarero, A. (2009b). Caracterización de parámetros físicos de las terminales de contenedores del sistema portuario español. *Ingeniería y ciencia, (10):49*.
28. González, M., Camarero, A. (2010). Explotacion de terminales de graneles solidos. *Litoral, ordenación y modelos de futuro: IV Congreso de Ingenieria Civil, Territorio y Medio Ambiente, pages 37-38*.
29. González, M. N. (2007). Metodología para la determinación de los parámetros de diseño de las terminales portuarias de contenedores a partir de datos de tráfico marítimo.
30. González, M. N., Camarero, A. (2009c). Caracterización de parámetros físicos de las terminales de contenedores del sistema portuario español. *Ingeniería y ciencia, (10):49*.
31. Han, J., Kamber, M., Pei, J. (2011). *Data mining: concepts and techniques*. Morgan Kaufmann.
32. Hand, D. J., Yu, K. (2001). Idiot's bayes.not so stupid after all? *International Statistical Review, 69(3):385-398*.
33. Hastie, T., Tibshirani, R., Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction*. Springer Verlag.
34. Heckerman, D. (1998). A tutorial on learning with bayesian networks. *Nato Asi Series D Behavioural And Social Sciences, 89:301-354*.
35. Holsheimer, M., Siebes, A., voor Wiskunde en Informatica, C. (1994). *Data mining: The search for knowledge in databases*. Centrum voor Wiskunde en Informatica.
36. HUGIN, E. A. (2012). Hugin expert a/s.
37. Jara-Díaz, S., Martínez-Budria, E., Cortes, C., Vargas, A. (1997). Marginal costs and scale economies in spanish ports. 25th european transport forum. In *Proceedings Seminar L, PTRC, London, pages 137-147*.
38. Jiménez, M. M. (2011). La influencia de la invención y desarrollo del contenedor en el comercio internacional y en ámbitos ajenos al mismo: (construcción naval y urbanismo en ciudades portuarias).
39. Kononenko, I. (2001). Machine learning for medical diagnosis: history, state of the art and perspective. *Artificial Intelligence in Medicine, 23(1):89-109*.
40. Larrañaga, P. e Inza, I. (2004). Clasificadores Bayesianos; Departamento de Ciencias de la Computación e Inteligencia Artificial; Universidad del País Vasco- Euskal Herreiko Unibertsiti-

- tatea. <http://www.sc.ehu.es/ccwbayes/docencia/mmcc/docs/t6bayesianos.pdf>
41. Lehavy, R., Li, F., Merkley, K. (2011). The effect of annual report readability on analyst following and the properties of their earnings forecasts. *The Accounting Review*, 86:1087.
  42. Liu, C. I., Jula, H., Ioannou, P. A. (2002). Design, simulation, and evaluation of automated container terminals. *Intelligent Transportation Systems, IEEE Transactions on*, 3(1):12–26.
  43. Mannila, H. (1997). Methods and problems in data mining. *Database Theory ICDT'97*, pages 41–55.
  44. Michalski, R. S. (1983). A theory and methodology of inductive learning. *Artificial Intelligence*, 20(2):111–161.
  45. Michalski, R. S., Bratko, I., Bratko, A. (1998). *Machine Learning and Data Mining; Methods and Applications*. John Wiley and Sons, Inc.
  46. Monie, G. D. (1987). *Measuring and evaluating port performance and productivity*. United Nations.
  47. Monie, G. D. (1989). Medición y evaluación del rendimiento y de la productividad de los puertos. *Monografías de la UNCTAD sobre Gestión de Puertos*.
  48. Nombela, G. (2009). Modelos de capacidad de infraestructuras de transporte. *Documento de trabajo. CEDEX. Ministerio de Fomento, accesible en www.evaluaciondeproyectos.es*.
  49. Norsys (2012). Netica.
  50. Orallo, J. H., Quintana, M. J.R., Ramírez, C.F. (2004). Introducción a la minería de datos. *Editorial Pearson Educación SA, Madrid*.
  51. Pearl, J. (1982). The solution for the branching factor of the alpha-beta pruning algorithm and its optimality. *Communications of the ACM*, 25(8):559–564.
  52. Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann.
  53. Perichinsky, G., García-Martínez, R., Proto, A. (2000). Knowledge discovery based on computational taxonomy and intelligent data mining. In *VI Argentinean Congress of Computer Science, CACIC, CD. National University of San Juan Bosco. Seat of Ushuaia. Argentina*. Citeseer.
  54. Pery, P., Camarero, A. (2003). *Determinación de la línea de atraque en los puertos españoles*. Universidad Politécnica de Madrid.
  55. Ramoni, M., Sebastiani, P. (1997). *Learning Bayesian networks from incomplete databases*. City Univ., Dep. of Actuarial Science and Statistics.
  56. Rodríguez-Dapena, A. (2009). *Prospectiva económica de Interés Portuario*.
  57. Rodríguez-Pérez, F. (1985). *Dirección y Explotación de Puertos*. Puerto Autónomo de Bilbao, Bilbao, España.
  58. Roll, Y., Hayuth, Y. (1993). Port performance comparison applying data envelopment analysis (dea). *Maritime Policy and Management*, 20(2):153–161.
  59. Ruiz-Shulcloper, J., Lazo, M. (1990). Modelos matemáticos para el reconocimiento de patrones. *Editorial UCLV, Santa Clara, Cuba*.
  60. Serrano, M. M. G., Castellano, L. T. (2005). La medición de la eficiencia en el sector portuario: revisión de la evidencia empírica. *Documentos de trabajo conjuntos: Facultades de Ciencias Económicas y Empresariales*, (6):1.
  61. Song, D., Zhang, J., Carter, J., Field, T., Marshall, J., Polak, J., Schumacher, K., Sinha-Ray, P., Woods, J. (2005). On cost-efficiency of the global container shipping network. *Maritime Policy and Management*, 32(1):15–30.
  62. Sotolongo, D. M., Valdivia, Z. Z. G. (2011). Red bayesiana con aspecto de mapa conceptual: Una perspectiva para los sistemas de enseñanza/aprendizaje inteligentes.
  63. Sun, S., Zhang, C., Yu, G. (2006). A bayesian network approach to traffic flow forecasting.

- Intelligent Transportation Systems, IEEE Transactions on*, 7(1):124–132.
64. Tebaldi, C., West, M. (1998). Bayesian inference on network traffic using link count data. *Journal of the American Statistical Association*, 93(442):557–573.
  65. Todd, B. S., Stamper, R., Macpherson, P. (1994). The design and construction of a medical simulation model. *Computer methods and programs in biomedicine*, 42(2):77–91.
  66. Tongzon, J. (2001). Efficiency measurement of selected australian and other international ports using data envelopment analysis. *Transportation Research Part A*, 35(2):107–122.
  67. Tongzon, J., Heng, W. (2005). Port privatization, efficiency and competitiveness: Some empirical evidence from container ports (terminals). *Transportation Research Part A: Policy and Practice*, 39(5):405–424.
  68. Tongzon, J. L. (1995). Determinants of port performance and efficiency. *Transportation Research Part A: Policy and Practice*, 29(3):245–252.
  69. Tovar, B., Jara-Díaz, S., Trujillo, L. (2003). Production and cost functions and their application to the port sector aliterature survey. *World*.
  70. Tovar, B., Jara-Díaz, S., Trujillo, L. (2004a). Funciones de producción y costes y su aplicación al sector portuario. una revisión de la literatura. *DOCUMENTO DE TRABAJO*, 6.
  71. Tovar, B., Trujillo, L., Jara-Díaz, S. (2004b). Organization and regulation of the port industry: Europe and Spain. *Essays on microeconomics and industrial organization. 2nd edn. Springer-Verlag, Heidelberg*, 189–207.
  72. Vityaev, E., Kovalerchuk, B. (2004). Empirical theories discovery based on the measurement theory. *Mind and Machine*, 14(4):551–573.
  73. Wanhill, S. R. C. (1974). A study in port planning: The example of mina zayed. *Maritime Policy and Management*, 2(1):48–55.
  74. Witten, I. H., Frank, E. (2005). *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann Pub.