applied sciences

MDPI

Article

Exploring Boost Efficiency in Text Analysis by Using AI Techniques in Port Companies

Claudia Durán [1], Christian Fernández-Campusano [2,*], Leonardo Espinosa-Leal [3], Cristóbal Castañeda [4], Eduardo Carrillo [5], Marcelo Bastias [6] and Felipe Villagra [1]

[1] Departamento de Ingeniería Industrial, Universidad Tecnológica Metropolitana, Santiago 7800002, Chile; c.durans@utem.cl (C.D.); fvillagra@utem.cl (F.V.)
[2] Departamento de Ingeniería Eléctrica, Facultad de Ingeniería, Universidad de Santiago de Chile (USACH), Santiago 9170124, Chile
[3] Graduate School and Research, Arcada University of Applied Sciences, 00560 Helsinki, Finland; leonardo.espinosaleal@arcada.fi
[4] Multicaja S.A., Santiago 8340306, Chile; cristobal.castaneda@klap.cl
[5] Esmax SPA, Santiago 7561127, Chile; eduardo.carrillo@esmax.cl
[6] Salfa Mantenciones, Santiago 7561127, Chile; mfbastiasv@salfamantenciones.cl
* Correspondence: christian.fernandez@usach.cl

Abstract: This study presents how integrating natural language processing (NLP) and machine learning (ML) optimizes strategic management in the port sector. Using hybrid NLP-ML models, the accuracy of classification and prediction of strategic information is significantly improved by analyzing large sets of textual data, both unstructured and semi-structured. The methodological approach is developed in three phases: first, a strategic analysis of port systems is performed using NLP; then, ML is integrated with NLP for text classification using advanced tools such as BERT and Word2Vec; finally, advanced models, including Decision Trees and Recurrent Neural Networks are evaluated. Applied to 55 companies in three countries, this method extracts key strategic data such as mission, vision, values and corporate objectives from their websites to obtain strategic terms related to innovation and sustainability. The study improves the ability to interpret textual data, enabling more informed and agile decision-making, which is essential in a highly competitive and dynamic environment.

Keywords: natural language processing; machine learning; hybrid learning; sustainability; innovation

Academic Editor: Panagiotis Tsarouhas

Received: 17 March 2025
Revised: 10 April 2025
Accepted: 10 April 2025
Published: 21 April 2025

Citation: Durán, C.; Fernández-Campusano, C.; Espinosa-Leal, L.; Castañeda, C.; Carrillo, E.; Bastias, M.; Villagra, F. Exploring Boost Efficiency in Text Analysis by Using AI Techniques in Port Companies. *Appl. Sci.* **2025**, *15*, 4556. https://doi.org/10.3390/app15084556

1. Introduction

Ports play a critical role in national economic development, not only by facilitating the movement of goods, but also by shaping local economies, environmental policies and international relations [1,2]. In a globalized and dynamic environment, integrating innovation and sustainability into strategic decision-making is essential to ensure long-term competitiveness [3]. In this context, advanced techniques in natural language processing (NLP) and machine learning (ML) have emerged as key tools for analyzing strategic texts and optimizing risk and resource management [3,4].

The large volume of strategic textual data, often semi-structured or unstructured, presents significant processing challenges that are further complicated by the lack of specialized big data expertise and linguistic diversity [5–7]. In addition, enterprise databases often store inaccurate or incomplete text-based information, hindering effective decision-making [8]. Despite the growing recognition of the role of big data in management, the application of text analytics in strategic decision-making remains relatively underexplored [9].

Addressing the complexities of managing large strategic data requires effective structuring and categorization, with a particular emphasis on text semantics. This approach not only improves data interpretation, but also facilitates the identification of key strategies and stakeholders in the port sector [10]. However, these advances come at a high computational cost and pose challenges in ensuring the reproducibility of results, necessitating ongoing research and refinement in both theoretical and practical applications [11]. For example, ports in the Caspian Basin need to adopt big data into their business strategies [12]. Also, the ports of Valencia and Singapore need advanced systems to handle data diversity and optimize the integration of multiple sources [13,14]. In the Port of Thessaloniki, data security is vital to maintain safe operations [13], while the Port of Rotterdam focuses on real-time data processing to facilitate quick operational decisions [14].

To address these challenges, this study explores the integration of business intelligence technologies with advanced natural language processing (NLP) and machine learning (ML) techniques. These technologies are critical for efficiently processing large datasets; however, their effectiveness can be limited by semantic limitations and the need for expert supervision. Therefore, the development and adaptation of neural network-based models that minimize bias and enhance data interpretability are essential to ensure reliable and actionable insights for strategic decision-making [15,16].

This research directly responds to the need for a hybrid ML-NLP model capable of classifying companies based on strategic port-related information, analyzing strategic texts to identify sustainability and technological aspects, and assessing the challenges and opportunities of applying ML and NLP in business classification.

For executives to fully benefit from the insights generated by these models, it is essential to refine them with innovative techniques that reduce discrepancies between collected data and system structures. Hybrid models that combine ML and NLP have shown great potential for classifying companies based on strategic port information and identifying sustainability and technology-related factors. However, these models often face interpretability issues, as they are often perceived as "black boxes", making automated decisions difficult to understand and trust. Integrating ML with interpretative models significantly improves pattern recognition, fostering greater transparency and reliability in classification processes [15–17].

To address these challenges, this paper proposes the development of a hybrid model designed to effectively manage strategic port data and facilitate the extraction of valuable insights through advanced techniques such as web scraping and text mining. This innovative approach will provide stakeholders with a deeper understanding of the causal relationships behind strategic decisions, ultimately improving governance and competitiveness within port systems. However, significant challenges remain, primarily due to the diverse strategies of different stakeholders, which can impact governance and overall competitiveness. In addition, extracting relevant information is complicated by inconsistent terminology, lack of contextual clarity, fragmented content and incomplete statements, all of which contribute to ambiguity in strategic documents [18].

2. Background

2.1. Port System Overview

As critical nodes in the global logistics network, ports play a fundamental role in international competitiveness and regional development. These hubs handle vast amounts of information and communications essential to global trade, and their operational efficiency is key to ensuring the smooth transportation and distribution of goods. Investing in innovation and adopting sustainable practices are key strategies that drive socio-economic growth and support the transition to more integrated and environmentally friendly logistics

operations. Such initiatives strengthen relationships with local communities and position ports as catalysts for positive regional change, ensuring their long-term success [1,19].

Smart management and strategic collaboration across supply chains are essential in today's competitive landscape, where entire supply networks, rather than individual companies, vie for dominance [20]. As shown in Figure 1, port companies operate within a complex network of private and public stakeholders, each performing key operational and logistical functions. The port industry is a complex and interconnected ecosystem that facilitates synergies in the physical flow of cargo, as well as information and communication across economic, social, technological and environmental dimensions (Microenvironment and Macroenvironment) [4]. These synergies are critical to driving innovation and sustainability in the port industry.

Within this network, stevedores, freight forwarders and shipping lines work closely in coordination to improve cargo handling and global shipping operations. Container terminals serve as an effective warehouse management, ensuring intermodal transportation with reduced time. Logistics providers, warehouse operators and intermodal carriers work collaboratively to improve the supply chain by integrating sustainable technologies and practices that increase efficiency and reduce environmental impact. Port authority oversight ensures compliance with environmental and operational regulations. Meanwhile, local communities engage as key participants in sustainable port development to minimize environmental impacts on surrounding areas. Technology companies contribute innovations that improve information management and communication within the port, facilitating decision-making and increasing the port's responsiveness to global market dynamics. This collaborative approach and integration of advanced technology not only supports the global supply chain, but also promotes regional economic development and establishes the port as a center for sustainable and efficient economic activity.

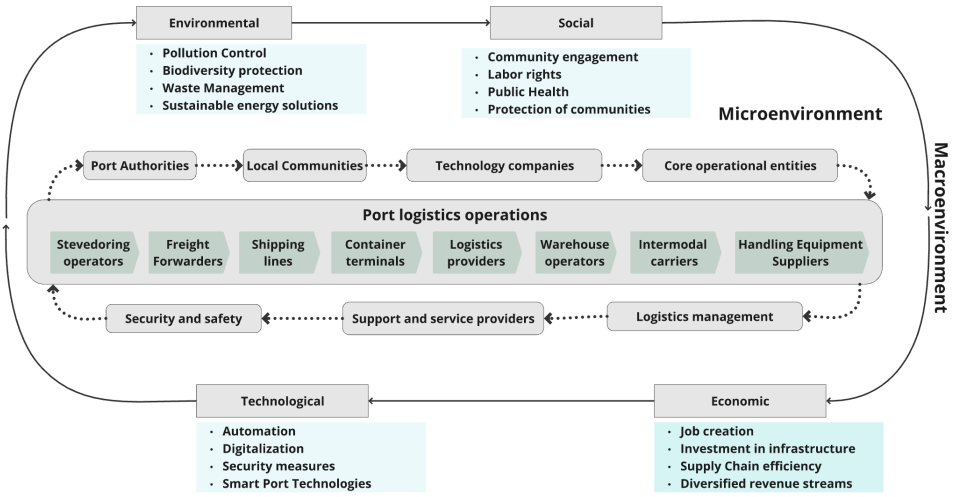

Figure 1. Factors influencing port logistics chains.

In addition, the ability to monitor and adapt organizational strategies in response to market dynamics is critical to maintaining a competitive edge. Strategic monitoring and competitive analysis enable companies to anticipate industry changes and adjust their operations accordingly to maintain or improve their market position [21,22].

Finally, a deep understanding of the external environment—its characteristics, challenges and emerging trends—is essential to business success. Managers with strong environmental scanning capabilities can gather, refine and adapt strategic insights to facilitate interactions with external stakeholders, minimize uncertainty and ensure that

ports continue to meet current and future demands while protecting human and natural ecosystems [23,24].

2.2. Literature Review

A review was conducted using the Web of Science (WoS) and SCOPUS databases to explore the techniques and methodologies used to analyze strategic and business texts related to logistics chains in different industry sectors. Keywords were selected to reflect advanced technologies that facilitate data analysis such as text mining and web scraping, innovative technologies that support strategic decision-making, logistics management, and terms related to sustainable practices. The following search query was used: TS = ((“data mining” OR “artificial intelligence” OR “machine learning” OR “data science”) AND (“smart manufacturing” OR “digital factory” OR “industrial automation” OR “big data in industry” OR “INDUSTRY 4.0” OR “smart industry”) AND (“supply chain” OR “logistics”) AND (“sustainability” OR “eco-friendly” OR “green manufacturing” OR “environmental impact” OR “sustainable practices”)).

A comprehensive qualitative analysis of the retrieved studies was performed and the results are presented in Table 1, which highlights the studies that align with the research questions defined in Section 1. Note, where the symbol “✓” means that it complies with the category, and otherwise (“×”) it does not comply.

The articles listed in Table 1 illustrate the role of natural language processing (NLP) and text mining in improving supply chain management. These techniques have facilitated efficient and strategic data processing in various industries, including the port sector. However, most of these studies do not explicitly address sustainability, revealing an opportunity for further research.

Analysis of textual data suggests that methods such as BERTopic and Latent Dirichlet Allocation (LDA) are effective in identifying patterns and trends within large datasets. These techniques play a critical role in supporting data-driven decision-making, particularly in dynamic environments such as port operations. The results highlight both the challenges and opportunities of using machine learning (ML) and NLP to classify companies based on strategic port-related information.

To broaden the scope, an extensive literature review was also conducted using Web of Science (WoS), SCOPUS and Google Scholar. The following search terms were used to identify additional relevant studies: “Hybrid model” OR “Machine learning” OR “Deep Learning” OR “Natural language processing” OR “Text mining” OR “Text classification” OR “Semantic text classification” OR “clustering” AND “management” OR “strategic” OR “industry” AND “innovation” OR “sustainability”.

The aim of this review was to investigate the application of semantic text classification using hybrid approaches in machine learning and NLP to support decision-making and strategic planning in organizations focused on innovation and sustainability. This method aims to provide advanced analytical tools that facilitate informed decision-making and strategic planning in contexts where innovation and sustainability are critical.

Table 2 summarizes the implementation of hybrid methodologies that integrate ML and NLP to analyze and categorize strategic texts across industries. These technologies improve strategic data management by providing automated, scalable and accurate analytical tools. However, the integration of sustainability considerations into these methodologies remains a significant challenge. Therefore, further research is needed to adapt text mining and ML-based approaches to drive sustainable innovations in port management, ultimately improving both operational efficiency and environmental responsibility.

Table 1. Critical aspects of innovation and sustainability with web scraping and text mining [1].

Major Contribution	Industry	Method/Tech.	NLP	TM [2]	WS [3]	S [4]
A sentiment analysis algorithm was developed to classify user reviews based on word-level emotion. The responses are structured according to rating criteria to help reference, compare and select medical centers [25].	Healthcare	Sentiment analysis	✓	×	✓	×
Analyzes large amounts of data to identify prevalent industries and locations cited in different contexts. Regression models are used to explore patterns in NLP findings and track shifts in attitudes toward big data over time [26].	Media and communications	Literature review	✓	✓	×	×
Presents NLP4Scoping, an innovative tool designed to support scope reviews. The study details the requirements, design and implementation of the tool and illustrates its functionality through a scenario analysis focused on innovation management in digital ecosystems [27].	Information technology and services	Literature review	✓	×	✓	×
Examines the alignment between public perception and academic perspectives on environmental, social and governance (ESG) factors by analyzing global news and academic papers through text mining. It shows that media coverage often mirrors academic findings, enabling companies to better align their strategies with market and societal expectations [28].	Financial services	BERTopic and topic modeling	✓	×	×	✓
Analyzes university slogans from 61 countries within the Belt and Road Initiative and identifies five educational themes. The study highlights the impact of China's BRI and how COVID-19 has fostered innovation in higher education while emphasizing the need for global cooperation in sustainable development [29].	Education	NLP and text mining	✓	✓	✓	×
Proposes a framework that integrates text mining and machine learning to predict technology trends and identify opportunities for innovation in refrigerated containers. The approach includes creating a technology roadmap and leveraging expert opinion to explore advances such as lighter, greener compressors [30].	Transport and logistics	Text mining and the Latent Dirichlet Allocation (LDA) topic model	✓	✓	✓	×
Examines challenges in the German wind energy sector, including regulatory barriers and declining investment. Proposes the use of NLP techniques to assess the credibility of the sector amid political controversy and declining validation. It shows that integrating big data governance into business strategies can improve environmental sustainability and reduce environmental impacts [31].	Electricity	Sentiment analysis and topic modeling	✓	×	✓	✓
Integrates text mining techniques into transportation infrastructure research, with an emphasis on practical applications and methodological choices. The study acknowledges its limited international scope [32].	Transport infrastructure	Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA)	✓	✓	×	×
Examines how text analytics can be used to optimize employer branding decisions and reduce training costs during economic downturns. The study highlights how some companies are prioritizing human capital over financial considerations, aligning with Indian leadership culture values of morality and ethics [33].	Human resources services	NLP and text mining	✓	✓	×	×
Validates a Word2Vec model on a large dataset of English Wikipedia articles, enabling the transformation of words into vectors to measure semantic similarities. This technique enhances innovative data management and semantic analysis applications [34].	Technology and information services	Cosine similarity and Pearson correlation	✓	✓	✓	×
Explores thematic modeling as a means to improve business management by generating new theories and concepts from textual data. The study identifies key constructs that enhance understanding of online audiences, consumer behavior, and sociocultural movements [35].	Financial services	NLP and topic modeling	✓	×	×	×
Develops a semi-supervised text mining technique for analyzing accident reports. By using domain-specific keywords and topic modeling with minimal expert input, the study demonstrates reduced manual intervention and improved keyword organization into topics [36].	Aviation and process	Text mining and topic modeling	✓	✓	×	×
Enhances pandemic-related text analysis by examining public sentiment and emerging patterns. The study aims to support frontline workers and healthcare professionals by providing rapid analysis of large datasets on public attitudes and trending topics [37].	Healthcare	Automated LDA for topic modeling, WordCloud and Word2Vec	✓	✓	×	×

[1] Accessed on 13 November 2024, https://www.webofscience.com. [2] Text mining. [3] Web scraping. [4] Sustainability.

Table 2. Hybrid ML and NLP analysis of management text on innovation or sustainability [1].

Description	Hybrid Apch.	Data Types	C [2]	
			I [3]	S [4]
Combines business intelligence and NLP for healthcare knowledge management. A case study from a Turkish hospital shows how to classify and analyze digital data [38].	Bag of Words (BoW) and K-means	Log records detailing digital activities and communications	✓	×
Examines wind energy patents and identifies key terms related to towers, shafts and turbines. Highlights innovations that improve shaft and motor efficiency and streamline assembly processes [39].	Text mining (TM) and K-means	Documents structural and mechanical component innovations	✓	×
Uses text mining and patent analysis to forecast technological advances in refrigerated container technology. Develops a technology roadmap focusing on compressors, power systems and refrigerants [30].	Term Frequency–Inverse Document Frequency (TF-IDF), LDA	Patents and scientific articles	✓	×
Introduces a novel text mining framework for patent analysis in Building Information Modeling (BIM) that identifies key applications and trends to drive technological advances and innovation in BIM construction [40].	TM, Social Network Analysis (SNA) and LDA	Patent documents	✓	×
Utilizes interdisciplinary methods, including clustering analysis and text similarity, to understand contributions to sustainable development goals (SDGs) [41].	TM and clustering in knowledge maps	Academic papers	✓	✓
Uses automated classification and keyword extraction to improve tourism analysis. It uses web diagrams for thematic differentiation and integrates the 7P marketing strategy into co-word analysis [42].	TM and hierarchical cluster analysis	Publicly communicated	✓	✓
Uses semi-supervised ML and NLP to extract and classify sustainability insights from Twitter data, demonstrating applicability to sustainable product innovation [43].	Support Vector Machines (SVMs), transductive SVM	Tweet ideas and opinions	✓	✓
Analyzes large supply chain management datasets to assess the impact of digital transformation. The methodology categorizes strategic information to improve decision-making and promote sustainable practices [44].	TM techniques, clustering and topic modeling	Journal articles	✓	✓
Employs large datasets for energy management and decision-making under uncertainty, using a hybrid approach to categorize and interpret data to support strategic decisions in energy policy and sustainability [45].	TM, clustering and topic modeling techniques	Journal articles	✓	✓
Uses text mining on patent data to analyze wind energy innovations, focusing on critical keywords of turbine components such as towers, shafts and assembly methods. This approach improves strategic management and oversight in the renewable energy sector [39].	TM and K-means	Patent documents.	✓	×
Develops an automated classification system for tourism to verify consistency, identify keywords, visualize thematic differences with web diagrams, integrate the 7P marketing strategy into the co-word analysis and highlight expert involvement for thematic consistency [42].	TM and hierarchical K-means clustering analysis	Journal articles	✓	✓
Incorporates ethical, demographic and legal dimensions into Iberostar's environmental management strategy to improve community involvement and meet hospitality standards [46].	TM and hierarchical cluster analysis	Strategic web texts	×	✓
Examines the circular economy as an alternative to traditional economic models, highlighting its role in environmental sustainability. The study uses modeling techniques to analyze 4488 research articles (2005–2023) to identify trends and research gaps [47].	TM and LDA	Journal articles	✓	✓
Integrates advanced natural language processing, noise-free topic modeling and multidimensional bibliometrics to identify emerging topics in nanomedicine and highlight their transformative impact on diagnostics, therapeutics and regenerative medicine [48]	BERT and Noiseless Latent Dirichlet Allocation (NLDA)	Journal articles	✓	✓

[1] Accessed on 20 November 2024, https://www.webofscience.com. [2] Contributions. [3] Innovation. [4] Sustainability.

3. Methods

The implementation of advanced machine learning techniques and analytical metrics is critical in the port industry. These tools improve management transparency and efficiency by enabling managers to identify trends, anticipate operational needs and make proactive adjustments. This section provides an overview of supervised and unsupervised machine learning methods that are essential for classifying and analyzing textual data. These approaches are particularly useful for identifying patterns related to innovation and sustainability in the port sector, as well as for developing the hybrid model proposed in Section 3.3.

3.1. Machine Learning Techniques

This study uses supervised learning techniques to develop predictive models for classifying strategic text within port-related datasets. These models are designed to detect key elements such as sustainability and innovation. In addition, unsupervised learning methods are used to analyze data without predefined categories, enabling the discovery of emerging patterns and trends. The integration of these methods enhances the processing of large, unstructured textual datasets and effectively addresses the challenges associated with them.

The following subsections describe the supervised and unsupervised learning algorithms used in this research.

3.1.1. Supervised Learning

(i) **Recurrent Neural Networks (RNNs).**

RNNs are a class of supervised learning algorithms designed to process sequential data by capturing temporal dependencies. They achieve this by maintaining information from previous states in hidden layers, which is essential for understanding dynamic patterns over time [49] (Equation (1)):

$$h_t = \sigma(W_h x_t + U_h h_{t-1} + b_h) \quad (1)$$

Here, h_t represents the hidden state at time t, capturing information from the current input x_t and the previous hidden state h_{t-1}. This mechanism allows the RNN to integrate prior data, with weight matrices W_h and U_h influencing the input and the transition from one state to the next, respectively, while b_h serves as the bias.

The output at time step t is given by

$$y_t = \phi(W_y h_t + b_y) \quad (2)$$

In this equation, y_t is the output, W_y is the weight matrix linking the hidden state to the output, b_y is the bias and ϕ represents a task-specific activation function, such as softmax for classification or identity for regression. The process demonstrates how an RNN uses the present inputs, x_t, and previous states, h_{t-1}, to generate a new hidden state, h_t, which serves as a "memory" and influences the subsequent output.

In the present study, the selection of RNN over alternative sequential models, such as LSTM, is substantiated by its efficacy in capturing short-term dependencies, a critical aspect of the dataset under consideration, wherein the sequences are not extensive. While LSTMs are regarded as superior in terms of learning long-term dependencies and circumventing issues such as gradient disappearance, in this study, RNNs are preferred due to their simplicity and lower computational cost. This choice enables faster processing and reduced resource utilization.

(ii) **Multi-Layer Perceptron (MLP)**.

The Multi-Layer Perceptron (MLP) is a neural network architecture designed for both classification and regression tasks. It consists of multiple layers of interconnected neurons that systematically process input data. The transformation of an input vector $\mathbf{x}$ through each hidden layer is defined as follows [50]:

$$\mathbf{z}^{(l)} = f\left(\mathbf{W}^{(l)}\mathbf{z}^{(l-1)} + \mathbf{b}^{(l)}\right) \quad (3)$$

Here, $\mathbf{z}^{(0)} = \mathbf{x}$ is the input, $\mathbf{z}^{(l)}$ is the output of the layer l (where $l = 1, 2, \ldots, L$), and $\mathbf{W}^{(l)}$ and $\mathbf{b}^{(l)}$ are the weight matrix and the bias vector, respectively, while f is a nonlinear activation function. The output layer generates the final output $\mathbf{y}$ as follows:

$$\mathbf{y} = f_{\text{out}}\left(\mathbf{W}^{(L)}\mathbf{z}^{(L-1)} + \mathbf{b}^{(L)}\right) \tag{4}$$

where f_{out} is an activation function chosen based on the task, such as softmax for classification or identity for regression.

(iii) **Support Vector Machine (SVM)**.
The Support Vector Machine (SVM) is a supervised learning algorithm that constructs a hyperplane to maximize the separation between two classes of data. The model is trained on a dataset represented as $(\mathbf{x}_i, y_i)$, with $i = 1, 2, \ldots, n$, where each $\mathbf{x}_i \in \mathbb{R}^d$ is a feature vector and $y_i \in \{-1, +1\}$ indicates the class label, defined by the following [51]:

$$f(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + b = 0 \tag{5}$$

where $\mathbf{w}$ is the weight vector orthogonal to the hyperplane and b is the bias term. To ensure accurate classification, the model enforces the following:

$$y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1 \quad \forall i \tag{6}$$

This condition ensures that all data points are correctly classified while maintaining a margin between the hyperplane and the nearest data points. The classification of new instances is determined by

$$\hat{y} = \text{sign}(\mathbf{w}^T\mathbf{x} + b) \tag{7}$$

(iv) **Decision Tree for ML (DT)**.
Decision trees (DTs) use a hierarchical structure to make decisions by recursively splitting data into branches based on selected features. This model is widely used for both classification and regression tasks due to its interpretability. The Gini index is often used to measure inequality within a node [52]:

$$\text{Gini}(t) = 1 - \sum_{i=1}^{C} p_i^2 \tag{8}$$

where t is the current node, C is the number of classes and p_i is the proportion of examples within the node that belong to class i. In addition, the information gain based on the entropy is calculated as

$$\text{Information gain} = \text{Entropy}(t) - \sum_{k=1}^{K} \frac{N_k}{N} \cdot \text{Entropy}(t_k) \tag{9}$$

where N is the total number of examples in the node t, N_k is the number of examples in the subnode t_k and K is the number of subnodes after splitting.

3.1.2. Unsupervised Learning

(i) **Random Forest Clustering (RFC)**
RFC adapts traditional random forest methods, typically used for classification and regression, to clustering tasks. This method constructs a forest of decision trees, each trained on different random subsets of the dataset, using an isolation forest approach to effectively handle unlabeled data [53]. Unlike supervised learning, RFC does not rely on predefined categories for data segmentation.

In RFC, the similarity between instances is quantified by the frequency with which two points land on the same leaf, which is captured in a similarity matrix S. The matrix element $S(i, j)$ represents the proportion of trees where two points share the same leaf, with values ranging from 0 to 1—higher values indicate greater similarity. This model supports the use of traditional clustering algorithms, such as k-means or hierarchical clustering, by providing a similarity matrix that can be interpreted as distances in spectral clustering approaches. RFC's ability to use random forest structures to identify intrinsic data similarities makes it particularly useful for addressing complex clustering challenges in studies involving unlabeled data.

3.2. Metrics

In this study, the predictive accuracy of machine learning models applied to strategic text analysis in port companies is evaluated. A set of established metrics is used to measure model performance in data classification and analysis. These metrics include root mean square error (RMSE), mean absolute error (MAE) and median mean absolute error (MDAE). Each metric serves to quantify different facets of prediction error, facilitating the assessment of model precision under different scenarios [54].

In addition, other validation metrics are used for a nuanced assessment of model performance in classification tasks, specifically accuracy, recall, F-measure and precision [54]. These metrics provide a detailed perspective on the model's ability to accurately classify data, which is essential for refining the models to meet the specific needs of port-related applications.

3.3. Hybrid Method

This study presents a hybrid methodology that combines advanced machine learning (ML) and natural language processing (NLP) techniques to address key challenges in strategic port management. The method aims to address three primary research questions:

1. How can a hybrid ML and NLP model be constructed to classify companies based on strategic port information?
2. How can strategic texts be analyzed to identify sustainable and technological aspects?
3. What are the potential challenges and opportunities associated with using ML and NLP to classify companies based on port information?

The hybrid method proposed in this study emphasizes innovation, which involves the application of new technologies and concepts that significantly enhance the analysis of large textual datasets, facilitating the discovery of previously hidden patterns and relationships. This approach prioritizes both analytical effectiveness and sustainability, aiming to ensure efficient and responsible use of technological resources while minimizing environmental impact.

This method was developed through an extensive literature review and practical experimentation with strategic port data collected from port entities such as terminals and logistics companies in Chile, Argentina and Spain. The data include missions, visions, values and strategic goals, among other key strategic information. The integration of these different data sources enables a comprehensive understanding of the operational context and supports the development of a robust and effective analytical framework.

In this study, the potential synergy between business intelligence, natural language processing (NLP) and machine learning (ML) to enhance strategic text analysis is also explored. The use of NLP, in particular, facilitates the automated analysis of text data to identify relevant trends and patterns. The application of ML uses these insights to predict behaviors and trends, improving the accuracy of analysis and providing actionable information for fast and effective strategic decision-making. This hybrid model flexibly

adapts to the specific needs of each situation and better aligns technological capabilities with strategic objectives, strengthening decision-making in port operations.

The developed hybrid method for strategic text analysis in port systems is outlined below:

3.3.1. Phase 1: NLP-Driven Strategic Analysis in Port Systems

1. **System overview and data collection:**
    - Port system description: The structure of the port system is reviewed to identify key players and their strategic roles, thereby enhancing the understanding of systemic operations. This phase also includes the selection of critical strategic texts that describe the decision-making processes within port companies.
    - Strategic data collection: Key information is systematically collected from multiple websites using both manual and automated web scraping techniques to ensure comprehensive data collection. The data collected include company names, roles within the port, geographic locations and strategic variables such as mission, vision, values, goals and other corporate text.
2. **Data processing and text analysis:**
    - Automated text extraction: Text is automatically extracted from digital documents and websites, streamlining the data collection process.
    - Text preprocessing techniques:
        - Normalization and tokenization: These processes transform raw text into a structured format suitable for further analysis.
        - Stop word removal and lemmatization: These steps refine the dataset by removing extraneous words and applying lemmatization to improve the quality of text analysis.
    - NLP and text characterization:
        - Text Mining: Statistical and machine learning algorithms are used to identify patterns and extract valuable insights from large text datasets.
        - Web scraping: This technique systematically extracts data from websites, providing a robust database for subsequent text mining.
        - NLP techniques: Advanced NLP techniques are used to deeply characterize textual information, enabling entity identification, text classification, and sentiment analysis to reveal strategic patterns and trends.

3.3.2. Phase 2: Advanced ML and NLP Methods for Strategic Text Classification

This phase consists of two main options, each of which uses a different analytical framework:

1. **Option 1: BERT-based text analysis and classification**
    - Preprocessing and vectorization: Texts are transformed into numerical vectors using BERT to achieve a deep contextual representation.
    - Sentiment analysis: This process evaluates and classifies emotions and opinions in the text, focusing on content relevant to port systems.
    - Dimensionality reduction and clustering: This step is critical for simplifying models and identifying significant data groups or patterns without compromising essential information.
    - Model training and evaluation: Fine-tunes the parameters and performance of BERT-based models.
    - Network diagram integration: Classification models are developed to incorporate network diagrams that help illustrate data connections and dependencies.

2. **Option 2: Word2Vec-based text analysis and classification**
    - Construction of representative dictionaries: Dictionaries are created to identify key terms in the areas of innovation, sustainability and technology.
    - Compilation of strategic texts: Texts representative of each category are organized for further training and analysis.
    - Vectorization: Word2Vec is used to convert text to vector form, capturing its semantic essence.
    - Training and evaluation: Machine learning classification models are trained and evaluated using the vector representations as input.
    - Integrate additional vectors: By integrating additional vectors into the models, the analysis is extended and the classification accuracy is improved.

Leverage cross-data training: Data from each option are used over the other to increase validation and robustness. Specifically, when option 1 is selected, data from option 2 are used for training and vice versa, ensuring that each model generalizes effectively across different datasets.

3.3.3. Phase 3: Advanced Predictive Analytics and Model Evaluation

1. **Neural network (NN) and hybrid model development**
    - NN training: Neural network models are trained to analyze operations within the port environment, enhancing the predictive capabilities of the system.
    - Hybrid model implementation: These models integrate neural networks with vectors generated by natural language processing (NLP), creating robust hybrid models that leverage both textual and numerical data.
2. **NLP processing and vectorization:**
    - Vectorization for prediction: Textual data are transformed into vector formats using NLP techniques, enabling these data to be used as input for predictive modeling.
    - Application of hybrid models: The hybrid models are applied to both classification and prediction tasks in the port context to improve accuracy and reliability.
    - Cluster classification of port companies: Classify port companies into clusters based on operational and strategic characteristics to identify patterns and improve decision-making.
3. **Hybrid model: NN and ML**
    - ML algorithms: Implement and train machine learning models, including decision trees, recurrent NN, multi-layer perceptrons, random forests and support vector machines.
    - Performance evaluation: Evaluate models using accuracy, recall, F1 score and overall precision metrics.
4. **Cluster classification of port companies:**
    - Perform cluster classification: Categorizes port companies into clusters based on their operational and strategic characteristics. This classification helps identify patterns that can significantly streamline decision-making processes.
5. **Analysis and real-time data integration:**
    - Results Analysis: Analyzes the effectiveness of models to determine their performance and identify areas for improvement.
    - Predictive Analytics: Explore real-time data integration coupled with predictive analytics. This approach is designed to predict trends and behaviors in port operations, enabling proactive management and optimization.

Figure 2 illustrates the hybrid model developed.

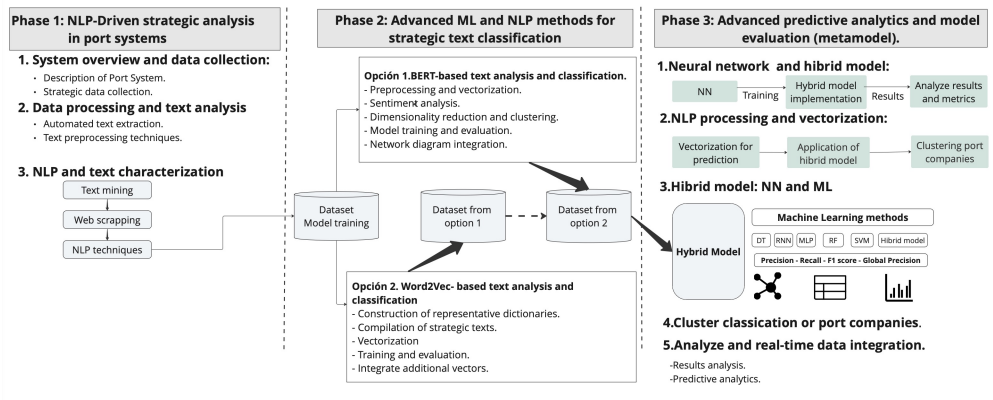

Figure 2. Hybrid method proposal.

Our research uses advanced natural language processing (NLP) and machine learning (ML) techniques to overcome the limitations of traditional statistical methods. Table 3, [55], provides a summary of the advantages of neural networks (NNs) over traditional methods applied to port management:

Table 3. Comparison between traditional methods and neural networks.

Criteria	Traditional Methods	Neural Network Methods
Efficiency	Suitable for simple problems that require fewer resources.	They require more resources due to their ability to handle large amounts of data and complex relationships.
Accuracy	Limited due to reliance on linear relationships and few predictors.	High due to ability to capture complex and non-linear dynamics.
Flexibility	Designed for specific data and relationships with limited adaptability.	Highly adaptable to different types of data and patterns.
Interpretability	Clear and easy to understand, with well-defined relationships between variables.	Less interpretable due to complexity and opaque operations.

This comparison underscores the superiority of AI methods, particularly in terms of efficiency, accuracy, and flexibility, which are critical for effective decision-making in complex environments.

4. Implementation and Evaluation of the Hybrid Method

In the implementation and evaluation of the hybrid method, as shown in Figure 2, the study meticulously applies and scrutinizes the hybrid model in different phases of analysis.

4.1. Phase 1: NLP-Driven Strategic Analysis in Port Systems

This phase starts with a manual exploratory review of websites of various international ports, specifically selecting those with content in Spanish and that contain all the strategic variables under study: mission, vision, values and strategic objectives. The research identifies 55 companies that display strategic content on their websites and have operational relationships with ports in Chile, Spain and Argentina. This cohort includes port managers, terminals and logistics companies. Data such as company names, their generic classification based on their role within the port, URLs and corresponding countries are systematically organized and archived in a CSV file.

After data collection, the preprocessing and text analysis phase begins with automated text extraction from the websites of these port companies, using libraries such as 'requests'

and 'BeautifulSoup' to capture HTML content. The text is normalized and tokenized to effectively structure it for further analysis. Natural language processing (NLP) techniques such as stripping and lemmatization are then used to refine the text and improve the quality of the analysis.

As shown in Figure 3, the database generated by web scraping does not fully capture the strategic information—missions, visions, values and strategic goals—of the 55 companies analyzed. Manual extraction significantly increases retrieval rates to 97% for missions, 46% for visions, 39% for values, 33% for strategic goals and 34% for other strategic information. The study shows that mission statements are often published because they provide a definitive view of the company's purpose, facilitate public understanding and internal alignment, and comply with regulatory standards and industry norms. Conversely, strategic visions and goals are often excluded due to their competitive sensitivity and the likelihood of frequent changes. Cultural differences also influence the prioritization of these elements, potentially affecting transparency and alignment with community values, and thus the company's social license to operate.

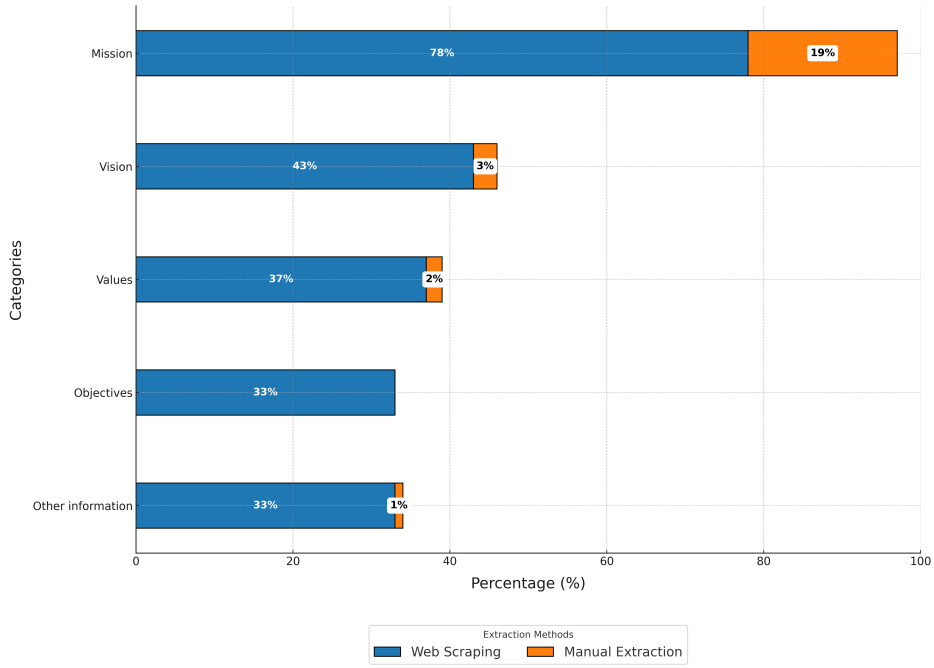

Figure 3. Efficiency of web scraping versus manual extraction.

4.2. Phase 2: Advanced ML and NLP Methods for Strategic Text Classification

4.2.1. K-Means Algorithm

In this phase, the K-means clustering algorithm plays a key role in the strategic text classification of port companies, as shown in Figure 4. The elbow method is used to determine the optimal number of clusters, an essential step in effectively grouping companies based on their strategic and operational characteristics [56]. This segmentation is critical for facilitating refined decision-making processes.

The inertia metric evaluates the internal cohesion of the clusters, indicating the compactness of the grouped data. The discovery of four optimal clusters confirms the algorithm's ability to detect significant strategic patterns in the dataset. Such insights are essential for performing text analysis aimed at influencing industry-specific strategies.

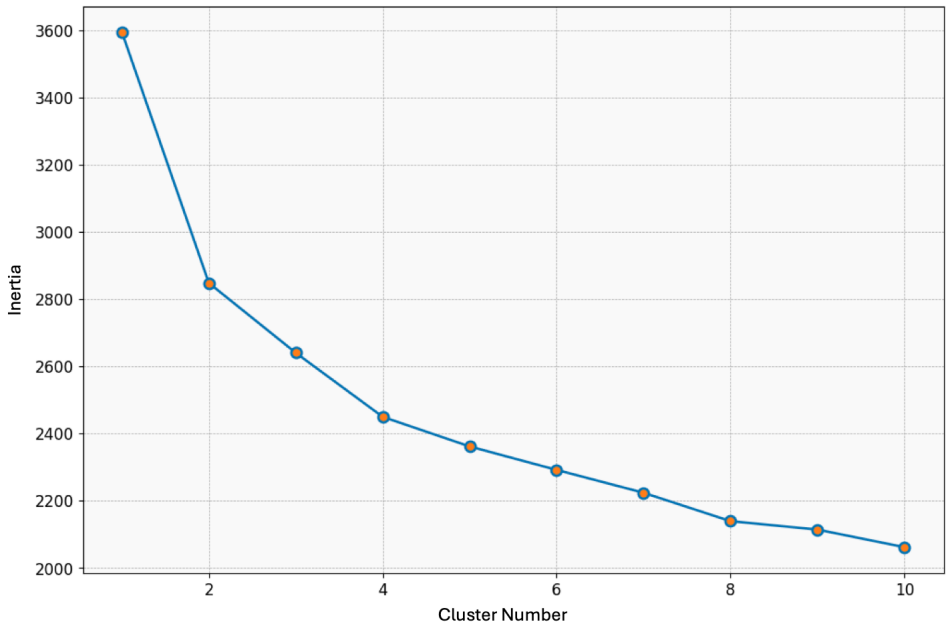

Figure 4. K-means elbow method.

4.2.2. Integration of TF-IDF and Clustering Models

The integration of the Term Frequency–Inverse Document Frequency (TF-IDF) technique and clustering models plays a key role in this research. The TF-IDF method is used to identify the 15 most relevant words in each cluster by analyzing their frequency both within the specific cluster and across the entire dataset. The configuration and results of the TF-IDF model are presented in Table 4. This table organizes the data with rows corresponding to each cluster and columns corresponding to specific terms. The values in the matrix represent the relative importance of each word within its cluster, helping to identify key terms that characterize the strategic focus of each group.

Table 4. Key terms by topic with Integration of the Term Frequency-Inverse Document Frequency (TF-IDF) (%).

Keywords	Topic 1	Topic 2	Topic 3	Topic 4
quality	0.00%	39.30%	23.90%	23.60%
customers	0.00%	38.00%	61.60%	49.70%
company	0.00%	22.50%	20.10%	41.90%
companies	0.00%	38.00%	3.10%	17.60%
team	0.00%	30.90%	11.90%	11.50%
experience	0.00%	12.60%	13.10%	21.20%
management	0.00%	28.10%	10.10%	18.80%
group	0.00%	29.60%	8.70%	13.30%
needs	0.00%	9.80%	22.60%	12.70%
offering	100.00%	13.80%	9.80%	11.90%
security	0.00%	19.70%	6.20%	18.80%
service	0.00%	26.70%	35.70%	33.40%
services	0.00%	22.50%	43.90%	31.50%
solutions	0.00%	19.70%	20.10%	14.60%
transportation	0.00%	4.20%	22.60%	28.50%

The analysis based on this technique, as shown in Table 4, highlights the dominant terms that define the strategic orientations within the clusters. For example, Theme 1 is characterized by the prominent use of the term “offer”, indicating a focus on service and resource provision. Themes 2 and 3 are characterized by terms such as “quality” and

"service", reflecting a commitment to improving customer experience and operational excellence. Theme 4 emphasizes "interaction", underscoring the importance of customer–company engagement, which is essential to strengthening business relationships.

In addition, the word cloud shown in Figure 5 visually represents the most common terms across the clusters, providing a concise overview of the key concepts within the dataset. This visualization highlights the strategic importance of elements such as customer service, quality, management efficiency and team collaboration—fundamental components for success in port operations.

Figure 5. Global word cloud for all clusters.

4.2.3. Pattern and Strategic Alliance Visualization

It discusses the use of advanced machine learning (ML) and natural language processing (NLP) techniques, particularly BERT and Word2Vec, which are instrumental in improving dimensionality reduction and clustering capabilities. These methods facilitate the simplification of models and enable the identification of relevant records without compromising the integrity of critical information. The implementation of these techniques is critical to effectively locate text that discusses sustainability and innovation, key issues in the port sector.

In addition, the t-Distributed Stochastic Neighbor Embedding (t-SNE) technique is applied in both two- and three-dimensional formats to visualize and analyze the spatial distribution of high-dimensional data. This visualization, as shown in Figures 6 and 7, is essential for distinguishing between clusters and identifying patterns and anomalies. Such insights are critical for understanding the dynamics of clustering and the spatial coherence between clusters. In these visualizations, the t-SNE axes are labeled according to the reduced dimensions, with distinct colors to enhance the visual distinction of the clusters.

In Figure 8, a comparative visualization of the results of three different clustering models is shown, represented by unique colors and shapes: Model 1 with blue circles, Model 2 with orange squares, and Model 3 with green diamonds. To enhance the clarity of this visualization and to mitigate point overlap, random jitter is applied and the opacity of the markers is adjusted, with the size of the points reflecting the density of overlap at specific locations. This comparative analysis is crucial for identifying the concordances and discrepancies between the models, thereby facilitating a more robust understanding of the clustering results.

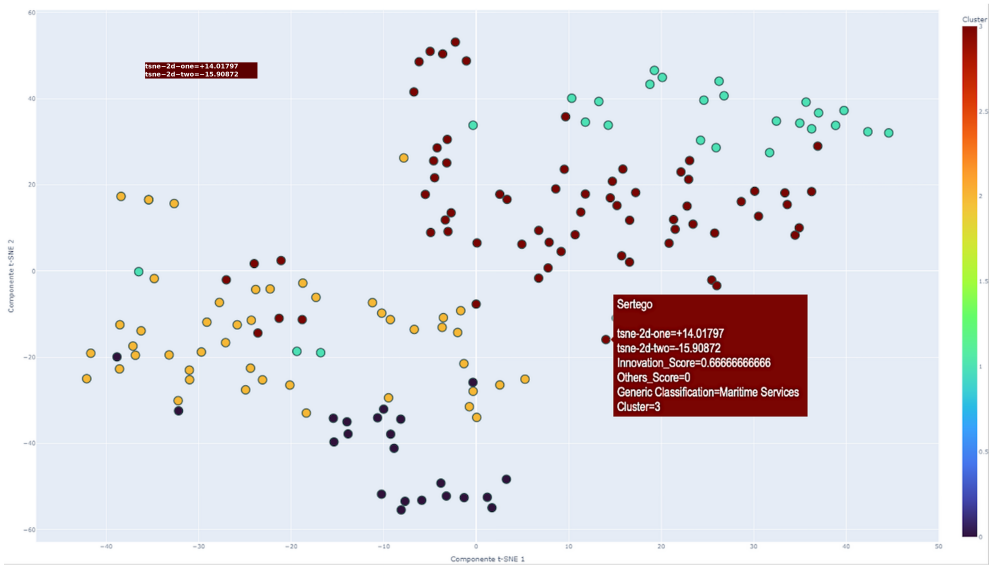

Figure 6. Full sampling clusters.

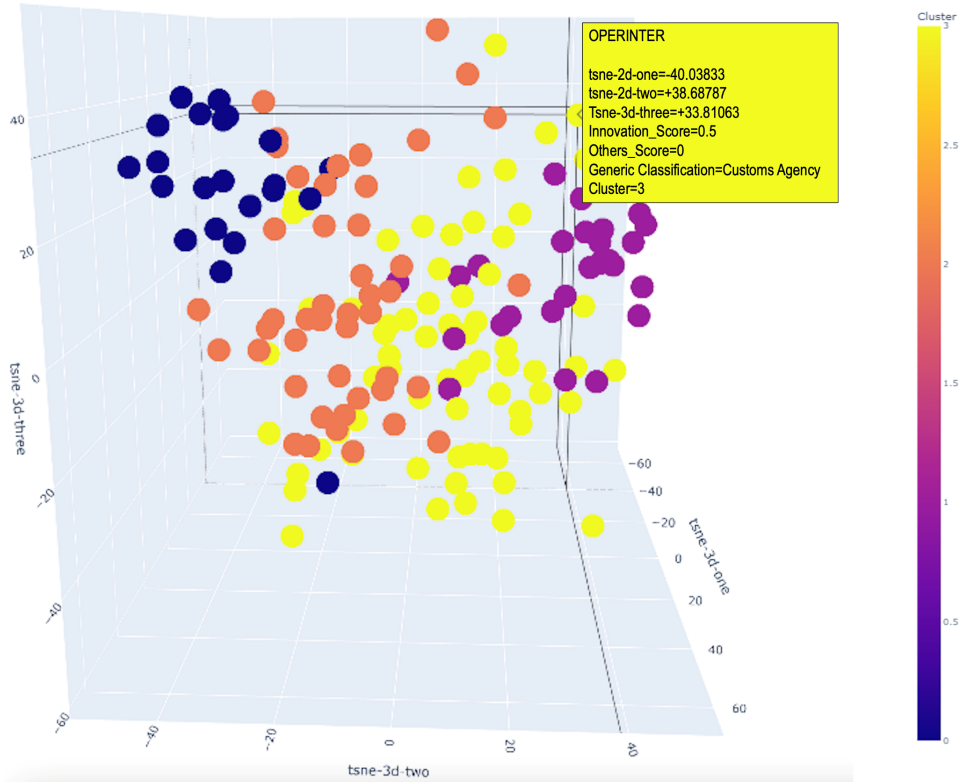

Figure 7. Three-dimensional clusters.

In regions where the three models classify points similarly, confidence in the analytical results is enhanced. Conversely, areas where points are assigned to different clusters reveal ambiguities in data characteristics, improving understanding of the underlying structure and supporting informed strategic decisions.

Figure 9 shows a similarity matrix based on cosine similarity to evaluate the relationships between ports and port companies. This visualization highlights both matches and mismatches in strategic data, helping to identify potential strategic alliances. Such capabilities are critical to the effective planning and execution of collaborative strategies in the port sector, allowing managers to make decisions based on clearly defined relationship patterns.

Figure 8. Clustering Models 1, 2 and 3.

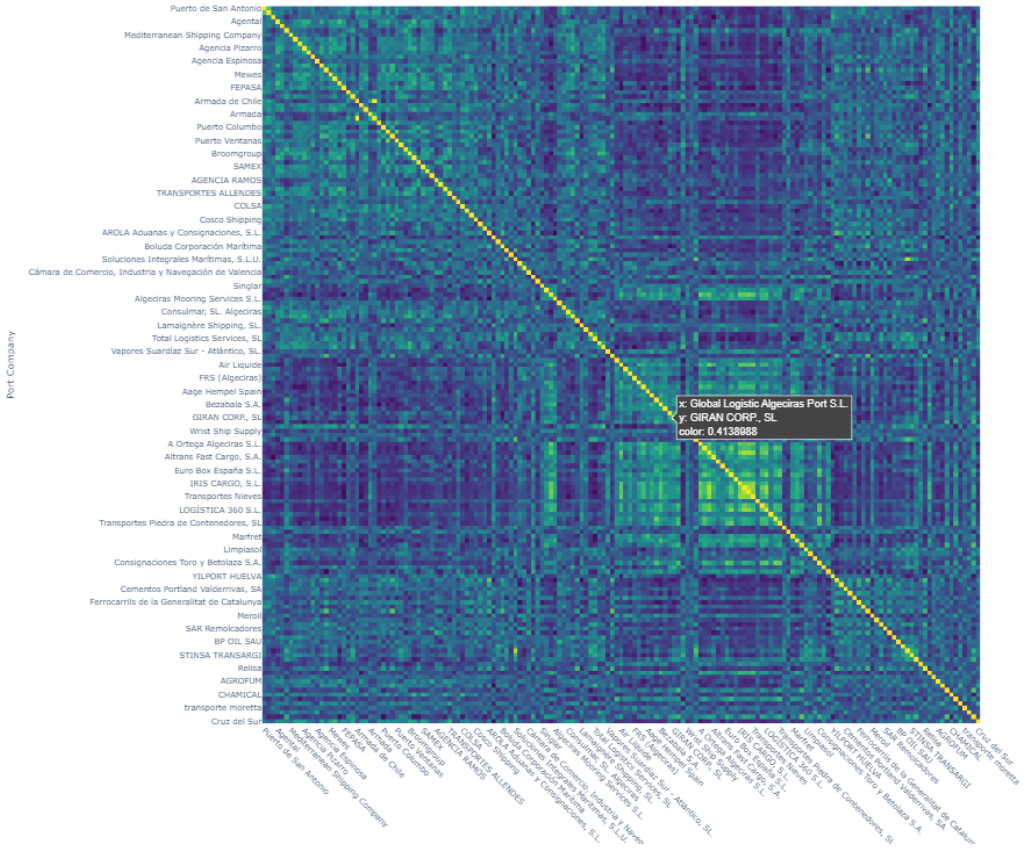

Figure 9. Correlation matrix.

The matrix shown in Figure 9 illustrates a moderately positive relationship between the Port of Valparaíso and the Empresa Portuaria de Iquique, indicating a strategic alignment in their operations. This observation not only validates the effectiveness of K-means clustering, but also highlights the matrix's ability to support informed strategic decisions. When combined with sentiment analysis, this tool becomes instrumental in identifying different strategic patterns across ports and companies, thereby enhancing the collaborative decision-making process and identifying potential partnerships based on shared goals.

Figure 10 shows a scatterplot that quantifies the alignment of 36 companies within the port sector with specific categories. Each bubble in the scatterplot corresponds to a company, with the size of the bubble representing the probability that the company fits into a particular category. This probability is derived from the congruence between the descriptive text of the company and the criteria of the designated category. This graphical representation not only confirms the results of the K-Means clustering algorithm, but also helps to explore and analyze strategic patterns in the industry.

Figure 10. Probability of category membership and corresponding keywords.

The companies are grouped into different categories, such as port managers, port operators, shipping agencies and government organizations, reflecting the functional diversity of the sector. The probabilities range widely, from 0.47 to 1.0, illustrating the extent to which the textual content of each company is consistent with its assigned category. For example, "Corporación de Prácticos del Puerto Bahía de Algeciras" and "Puerto Mejillones" have a perfect alignment score of 1, while "Puerto Ventanas" has the lowest score of 0.47, indicating a possible misalignment or lack of clarity in its strategic communication. The strategic keywords associated with each company, such as "sustainability", "quality", "service" and "safety", underscore priorities such as sustainability and safety. In-depth analysis of these probabilities along with the associated keywords is critical for identifying potential strategic alliances and collaborations, suggesting that companies with high alignment scores and similar strategic orientations are prime candidates for partnerships, while those with lower scores may need to improve their strategic communications to better align with their defined roles.

4.3. Phase 3: Advanced Predictive Analytics and Model Evaluation

The Phase 3 experiments were run on a system equipped with a 64-bit Windows 11 operating system, a 2.30 GHz Intel® Core™ i7-12700 processor, a 6 GB GeForce RTX 3060 GPU, and 16 GB of RAM. This configuration provided the necessary computational efficiency and processing power to perform a thorough evaluation of the models.

4.3.1. Evaluation of Classification Models Using Error Metrics

When evaluating classification models using error metrics, several tests were performed to determine the effectiveness of these models compared to traditional text classification methods. According to the results in Table 5, the Multi-Layer Perceptron (MLP) demonstrated superior performance in several error metrics, such as root mean square error (RMSE) and mean absolute error (MAE), achieving the lowest values, indicating high precision and stability. On the other hand, decision tree (DT) and Recurrent Neural Networks (RNNs) showed lower performance as reflected in a higher RMSE, indicating lower accuracy.

The hybrid method, which combines supervised and unsupervised learning techniques, showed moderate performance in RMSE but significant improvements in MAE. This advance provides a balance between accuracy and the ability to handle variation and noise in the data, making it suitable for applications that require efficient adaptation to complex and diverse datasets while optimizing the use of computational resources.

Finally, the median-based error, commonly used in regression problems, is advantageous when outliers are expected in the data or when the error distribution is asymmetric. This median absolute error (MDAE) is robust, meaning that a few extreme values do not distort the metric.

Table 5. Error metrics.

Errors	RNN	MLP	RF	SVM	DT	Hybrid
RMSE	1.605	0.174	0.179	0.311	1.606	0.596
MAE	0.758	0.030	0.032	0.097	0.387	0.161
MDAE	0.00	0.000	0.0	0.0	0.0	0.0

4.3.2. Performance Analysis of Classification Models

Table 6 analyzes the performance of different classification models using several key metrics. The Recurrent Neural Network (RNN) shows moderate performance, characterized by limitations in accuracy and ability to accurately identify positive cases, as indicated by its low F1 score. In contrast, the Multi-Layer Perceptron (MLP) demonstrates superior efficiency, achieving the highest levels of accuracy and recall, and an almost perfect balance of F1 scores, establishing it as the most reliable model for applications requiring high precision and sensitivity.

The random forest (RF) and Support Vector Machine (SVM) models show the worst performance across all evaluated metrics, indicating potential problems with configuration or dataset compatibility. These models require significant parameter adjustments to improve performance. Meanwhile, the deterministic model shows consistent, albeit intermediate, results in accuracy and recall, but does not match the effectiveness of the MLP or the hybrid approach.

The hybrid method stands out as a robust option with high accuracy, excellent recall and a balanced F1 score. This method provides a viable alternative to the MLP that is particularly beneficial in scenarios where the integration of multiple modeling techniques provides additional benefits such as robustness and interpretability. This analysis highlights the need for careful selection and calibration of classification models to tailor their

performance to specific tasks, and illustrates the benefits of a comprehensive approach to solving complex text classification challenges.

Table 6. Performance metrics.

Technique	RNN	MLP	RF	SVM	DT	Hybrid
Precision	0.807	0.977	0.0359	0.0361	0.7717	0.922
Recall	0.727	0.967	0.1895	0.1895	0.7712	0.903
F1 Score	0.697	0.971	0.0604	0.0607	0.7694	0.901
Global Precision	0.727	0.970	0.1895	0.1895	0.771	0.903

5. Discussion

5.1. Challenges of the Hybrid Model

The implementation and evaluation of the hybrid model have demonstrated its effectiveness in improving text classification and strategic analysis in the port sector. By integrating advanced machine learning (ML) and natural language processing (NLP) techniques, the model competently processes strategic information in Spanish that is critical for supporting decisions on sustainability and innovation in seaports. This approach has refined strategic decision-making and facilitated the identification of opportunities for collaboration and strategic alliances among port entities, which are essential for maintaining global competitiveness.

The integration of the K-means algorithm with TF-IDF analysis has revealed significant linguistic patterns that reflect the strategic orientations of port companies, highlighting key terms such as “quality”, “service” and “security”. In addition, the use of data visualization techniques such as t-SNE has significantly improved the interpretation of large datasets, allowing for clearer distinctions between clusters and helping to identify outliers and trends.

The study identifies significant challenges due to the linguistic variability and inconsistency of terms used in strategic documents, which complicates the standardization of text analysis processes. Such variability requires ongoing expert monitoring and continuous model adjustments to account for evolving language use and changes in business context, requiring a meticulous and persistent approach [6,7].

From an applied perspective, the findings highlight the importance of contextually adapting natural language processing (NLP) tools to improve governance and decision-making in port systems. This adaptation takes into account economic, social, technological and environmental factors [57]. Strategic divergences between companies pose significant challenges to the governance and competitiveness of port systems [4].

Neural networks (NNs) have significantly improved computational efficiency and predictive accuracy, which are critical for strategic analysis in sectors where fast and accurate decision-making is critical for maintaining competitive advantages [21]. These technologies support an end-to-end training architecture that eliminates the need for task-specific feature engineering. This simplification helps manage large amounts of training data and facilitates the implementation of advanced models [58]. However, the complexity inherent in NNs can complicate the interpretation of results. Therefore, careful data preparation and continuous monitoring are essential to ensure that predictions are aligned with business needs [59]. High computational costs and challenges in the reproducibility of results highlight the need for sustained focus in both research and practical applications [11].

Future research should focus on improving the transparency and explainability of ML and NLP models in strategic contexts. With the increasing reliance on automation and artificial intelligence in strategic management, balancing operational efficiency with ethical and social responsibility becomes paramount. Furthermore, extending these technologies to

different areas of text analysis could improve the adaptability of hybrid models to different linguistic and cultural contexts.

5.2. Strategic and Technological Analysis in Ports

The integration of hybrid models using machine learning (ML) and natural language processing (NLP) improves port management and strategic decision-making. These models facilitate better planning and resource utilization, optimize security and compliance through automated document analysis, and enable accurate, data-driven decision-making. In addition, they foster improved communication within the global port ecosystem and drive service innovation, thereby increasing port competitiveness.

As detailed in Section 4, the strategic text analysis of ports and terminals in Chile and Spain highlights regional differences in the integration of sustainability and innovation—key elements for evolving into smart and sustainable ports. For example, the Port of San Antonio and the Port of Valparaiso in Chile are noted for their sustainability initiatives and effective management, demonstrating a proactive stance in adopting advanced technologies to improve logistics chain management. These ports are committed to sustainability and innovation, although actual implementation remains limited. The implementation of intelligent traffic management systems and real-time analytics has significantly improved logistics dynamics and reduced waiting times, thereby increasing port operational efficiency.

In Spain, companies such as APM Terminals Algeciras and CSP Iberian Bilbao Terminal S.L. demonstrate a strong commitment to social responsibility and environmental sustainability. APM Terminals Algeciras has implemented technologies to reduce energy consumption and prevent pollution, while CSP Iberian Bilbao has become a SMART logistics center, adopting practices that promote sustainable development and improve operational efficiency. These cases illustrate different regional approaches to sustainability and innovation: Chilean ports prioritize operational efficiency and logistics integration, while Spanish ports emphasize social responsibility and environmental impact.

Despite the progress made, there remains an essential need to integrate advanced technologies such as machine learning (ML) and natural language processing (NLP) into the operational strategies of all ports under study. The ability of these models to analyze and classify strategic information could greatly enhance sustainability management and innovation, offering significant opportunities to improve strategic decision-making and strengthen global competitiveness.

Future research should focus on optimizing the use of these hybrid models to address critical challenges such as safety, operational efficiency and environmental sustainability. In addition, it is imperative to further explore the challenges and opportunities associated with using ML and NLP for business classification and strategic information management in the port context. Identifying how these technologies can be tailored to the specific needs of each port will be critical to maximizing their impact and ensuring successful implementation.

5.3. Related Work

The literature review, as detailed in Section 2.2, reveals the limited application of machine learning (ML) and natural language processing (NLP) techniques in automating the analysis of business texts. The review identifies key decisions supported by these technologies and elucidates the keywords and sources of the analyzed texts, with a particular focus on innovation and sustainability. Although NLP and ML are increasingly used in various fields, their integration into business strategy and sustainability management is still in its infancy.

This study advances the application of NLP and ML by tailoring these technologies for direct strategic decision-making, setting it apart from previous research that focuses

on specific operational applications. Unlike traditional methods that merely optimize operational processes without a strategic foundation, this research proposes a model designed to predict trends and dynamically adjust strategies in real time. This capability is critical for maintaining competitiveness in fluctuating markets. Notably, this research uses strategic data extracted directly from corporate websites, providing a direct and modern approach to strategic analysis. As a result, it enhances the practical applications of NLP and ML and provides a foundational framework for future research on their integration into corporate strategic planning and management.

Table 7 presents a synthesis of key strategies related to innovation and sustainability identified in the reviewed studies.

Table 7. Key strategies for innovation and sustainability.

Strategic Decision	Keywords	Text Sources
It supports recommendation systems and information platforms that enhance medical center selection by providing reliable and up-to-date data [25].	Charge, service, bill, price, hospital.	User reviews of medical services on online platforms.
It explores how interpreting market dynamics and consumer trends through comprehensive data analytics can inform business strategies [26].	Big data, executives, computer software, business analytics.	News accessible through LexisNexis® Academic.
Organizing and visualizing large amounts of academic information is critical for quickly identifying emerging areas of research and addressing knowledge gaps [27].	Business model, digital platforms, firm, management, government.	Journal articles and conference proceedings from SCOPUS databases.
Facilitates corporate and investor access to critical information on how environmental, social and governance (ESG) issues affect long-term shareholder value and sustainability [28].	Industry, emission, trade, business, governance.	News and articles from LexisNexis® and Web of Science.
Supports proactive decision-making processes aimed at improving safety, service quality and customer satisfaction [36].	Traffic control, flight en-route, turbulence, level flight, cruise climb.	Accident reports, manuals, glossaries and Wikipedia.

Explainability and trust in AI are essential because while AI improves decision accuracy by modeling complex relationships, its lack of transparency can be a challenge. Adopting explainable AI that details decision-making processes increases transparency and facilitates stakeholder acceptance, ensuring that AI models are aligned with strategic needs and are understandable, thereby building trust and driving adoption.

6. Conclusions

This research demonstrates that the integration of natural language processing (NLP) and machine learning (ML) improves strategic management in the port sector. Hybrid NLP-ML models significantly increase the accuracy of classifying and predicting strategic information by analyzing large volumes of textual data. This capability enables more informed and agile decision-making in a highly competitive and dynamic environment. Despite technical and regulatory challenges, the benefits of using these advanced technologies are clear, particularly in identifying key trends in innovation and sustainability that are essential for the future growth of the sector. These models optimize operational management and promote sustainable business practices by effectively processing and analyzing data. Continued development of these technologies is essential, adapting them to specific industrial applications and minimizing their reliance on expert intervention. Such advancements will not only broaden the application of NLP and ML in various strategic

areas, but also promote their widespread adoption, ensuring that port operations and other sectors benefit fully and efficiently from artificial intelligence.

Author Contributions: Conceptualization, C.D. and C.F.-C.; formal analysis, C.D., C.F.-C. and L.E.-L.; funding acquisition, C.D.; investigation, C.D., C.F.-C. and L.E.-L.; methodology, C.D., C.F.-C., C.C., E.C., M.B. and F.V.; project administration, C.D. and C.F.-C.; software, C.D., C.F.-C., C.C., E.C., M.B. and F.V.; supervision, C.D., C.F.-C. and L.E.-L.; validation C.D., C.F.-C. and L.E.-L.; visualization, C.C., E.C., M.B. and F.V.; writing—original draft, C.D., C.F.-C. and L.E.-L.; writing—review and editing, C.D., C.F.-C. and L.E.-L. All authors have read and agreed to the published version of the manuscript.

Funding: Project funded by the Research Continuity Project Fund, year 2022, code LCLI22-05, Universidad Tecnológica Metropolitana.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets used or analyzed in the current study are available from the corresponding author on reasonable request.

Conflicts of Interest: Author Cristóbal Castañeda was employed by the company Multicaja S.A.; Eduardo Carrillo was employed by the company Esmax SPA; Marcelo Bastias was employed by the company Salfa Mantenciones. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

NLP	Natural Language Processing
ML	Machine Learning
LDA	Latent Dirichlet Allocation
WoS	Web of Science
RNN	Recurrent Neural Network
MLP	Multi-Layer Perceptron
RF	Random Forest
SVM	Support Vector Machine
DT	Decision Tree
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MDAE	Median Absolute Error
MSE	Mean Square Error
RMSE	Root Mean Square Error

References

1. Park, J.S.; Seo, Y.J. The impact of seaports on the regional economies in South Korea: Panel evidence from the augmented Solow model. *Transp. Res. Part Logist. Transp. Rev.* **2016**, *85*, 107–119. [CrossRef]
2. Hossain, T.; Adams, M.; Walker, T. Role of sustainability in global seaports. *Ocean Coast. Manag.* **2021**, *202*, 105435. [CrossRef]
3. Adom, A.Y.; Nyarko, I.K.; Som, G.N.K. Competitor Analysis in Strategic Management: Is it a Worthwhile Managerial Practice in Contemporary Times? *J. Resour. Dev. Manag.* **2016**, *24*, 116–127.
4. Durán, C.; Córdova, F. ScienceDirect Information Technology and Quantitative Management (ITQM 2016) Conceptual model to identify technological synergic relationships of strategic level in a medium-sized Chilean port. *Procedia Comput. Sci.* **2016**, *91*, 382–391. [CrossRef]
5. Menon, A.; Choi, J.; Tabakovic, H. What You Say Your Strategy Is and Why It Matters: Natural Language Processing of Unstructured Text. *Acad. Manag. Proc.* **2018**, *2018*, 18319. [CrossRef]
6. Sharoff, S. What neural networks know about linguistic complexity. *Russ. J. Linguist.* **2022**, *26*, 371–390. [CrossRef]

7. Ranjan Jayanthi, F.C. Big Data Analytics in Building the Competitive Intelligence of Organizations. *Int. J. Inf. Manag.* **2021**, *56*, 102231. [CrossRef]
8. Zhecheva, D.; Nenkov, N. Business demands for processing unstructured textual data – text mining techniques for companies to implement. *Access J. Access Sci. Busin. Innov. Digit. Econ.* **2022**, *3*, 107–120. [CrossRef]
9. Mouratidis, I.; Kamariotou, M.I.; Kitsios, F.C. Big Data Strategy and Business Analytics: A Literature Review. In *Operational Research in the Era of Digital Transformation and Business Analytics*; Matsatsinis, N.F., Kitsios, F.C., Madas, M.A., Kamariotou, M.I., Eds.; Springer: Cham, Switzerland, 2023; pp. 171–178.
10. Evangelopoulos, N.; Zhang, X.; Prybutok, V. Latent Semantic Analysis: Five Methodological Recommendations. *Eur. J. Inf. Syst.* **2012**, *21*, 70–86. [CrossRef]
11. Sobrie, O.; Mousseau, V.; Pirlot, M.; Fortemps, P.; Mahmoudi, S.; De Smet, Y.; Labreuche, C.; Gillis, N.; Ouerdane, W.; de Mons, U.; et al. *Learning Preferences with Multiple-Criteria Models*; Université Paris Saclay (COmUE); Université de Mons, 2016. Available online: https://theses.hal.science/tel-01370555 (accessed on 9 April 2025).
12. Alekberli, R.Z.; Haussmann, R.E. Integrating Big Data Governance and Corporate Strategies in Small and Medium Caspian Basin Seaports. In Proceedings of the 2024 IEEE Global Conference on Artificial Intelligence and Internet of Things (GCAIoT), Dubai, United Arab Emirates, 19–21 November 2024; pp. 1–6. [CrossRef]
13. Nikolakopoulos, A.; Julian Segui, M.; Pellicer, A.B.; Kefalogiannis, M.; Gizelis, C.A.; Marinakis, A.; Nestorakis, K.; Varvarigou, T. BigDaM: Efficient Big Data Management and Interoperability Middleware for Seaports as Critical Infrastructures. *Computers* **2023**, *12*, 218. [CrossRef]
14. Durlik, I.; Miller, T.; Cembrowska-Lech, D.; Krzemińska, A.; Złoczowska, E.; Nowak, A. Navigating the Sea of Data: A Comprehensive Review on Data Analysis in Maritime IoT Applications. *Appl. Sci.* **2023**, *13*, 9742. [CrossRef]
15. Robert, A.; Frank, L.; Potter, K. Explainable AI: Interpreting and Understanding Machine Learning Models. *Artif. Intell.* **2024**. [CrossRef]
16. Barredo Arrieta, A.; Díaz-Rodríguez, N.; Del Ser, J.; Bennetot, A.; Tabik, S.; Barbado, A.; Garcia, S.; Gil-Lopez, S.; Molina, D.; Benjamins, R.; et al. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Inf. Fusion* **2020**, *58*, 82–115. [CrossRef]
17. Gutierrez-Bustamante, M.; Espinosa-Leal, L. Natural language processing methods for scoring sustainability reports—A study of Nordic listed companies. *Sustainability* **2022**, *14*, 9165. [CrossRef]
18. Berry, D.M. Ambiguity in Natural Language Requirements Documents. In *Innovations for Requirement Analysis. From Stakeholders' Needs to Formal Designs*; Paech, B., Martell, C., Eds.; Springer: Berlin/Heidelberg, Germany, 2008; pp. 1–7.
19. Durán, C.; Palominos, F.; Carrasco, R.; Carrillo, E. Influence of Strategic Interrelationships and Decision-Making in Chilean Port Networks on Their Degree of Sustainability. *Sustainability* **2021**, *13*, 3959. [CrossRef]
20. Rodriguez Estevez, D.; González-Cancelas, N.; Camarero, A.; Vaca Cabrero, J. Development of a "Smart Dry Port" Indicator and Ranking Calculation for Spanish Dry Ports. *Future Transp.* **2023**, *3*, 1272–1291. [CrossRef]
21. Feng, J.; Han, P.; Zheng, W. Identifying the factors affecting strategic decision-making ability to boost the entrepreneurial performance: A hybrid structural equation modeling – artificial neural network approach. *Front. Psychol.* **2022**, *13*, 1038604. [CrossRef]
22. Amoako, G.; Omari, P.; Kumi, D.; Agbemabiase, G.; Asamoah, G. Conceptual Framework—Artificial Intelligence and Better Entrepreneurial Decision-Making: The Influence of Customer Preference, Industry Benchmark, and Employee Involvement in an Emerging Market. *J. Risk Financ. Manag.* **2021**, *14*, 604. [CrossRef]
23. Lienert, J.; Linkov, I. Editorial featured papers on environmental decisions. *EURO J. Decis. Process.* **2019**, *7*, 151–157. [CrossRef]
24. Denktas-Sakar, G.; Karatas-Cetin, C. Port Sustainability and Stakeholder Management in Supply Chains: A Framework on Resource Dependence Theory. *Asian J. Shipp. Logist.* **2012**, *28*, 301–319. [CrossRef]
25. Leong, K.H.; Dahnil, D.P. Classification of Healthcare Service Reviews with Sentiment Analysis to Refine User Satisfaction. *Int. J. Electr. Comput. Eng. Syst.* **2022**, *13*, 323–330. [CrossRef]
26. Haider, M.; Gandomi, A. When big data made the headlines: Mining the text of big data coverage in the news media. *Int. J. Serv. Technol. Manag.* **2021**, *27*, 23. [CrossRef]
27. Chiarello, F.; Gastaldi, L.; Martini, A. Design and implementation of a text mining-based tool to support scoping reviews. *Int. J. Technol. Manag.* **2023**, *91*, 147. [CrossRef]
28. Lee, H.; Lee, S.H.; Lee, K.R.; Kim, J.H. ESG Discourse Analysis Through BERTopic: Comparing News Articles and Academic Papers. *Comput. Mater. Contin.* **2023**, *75*, 6023–6037. [CrossRef]
29. Wang, H.; Lu, Q. Understanding Philosophies of Higher Education between Countries in China's Belt and Road Initiative: Analysis of University Mottos Based on Natural Language Processing Technology. *Sage Open* **2022**, *12*, 21582440221. [CrossRef]
30. Wang, Y.; Feng, L.; Wang, J.; Zhao, H.; Liu, P. Technology Trend Forecasting and Technology Opportunity Discovery Based on Text Mining: The Case of Refrigerated Container Technology. *Processes* **2022**, *10*, 551. [CrossRef]

31. Dehler-Holland, J.; Okoh, M.; Keles, D. Assessing technology legitimacy with topic models and sentiment analysis – The case of wind power in Germany. *Technol. Forecast. Soc. Chang.* **2022**, *175*, 121354. [CrossRef]
32. Chowdhury, S.; Alzarrad, A. Applications of Text Mining in the Transportation Infrastructure Sector: A Review. *Information* **2023**, *14*, 201. [CrossRef]
33. Karkhanis, G.V.; Chandnani, S.U.; Chakraborti, S. Analysis of employee perception of employer brand: A comparative study across business cycles using structural topic modelling. *J. Bus. Anal.* **2023**, *6*, 95–111. [CrossRef]
34. Jatnika, D.; Bijaksana, M.A.; Suryani, A.A. Word2Vec Model Analysis for Semantic Similarities in English Words. *Procedia Comput. Sci.* **2019**, *157*, 160–167. [CrossRef]
35. Hannigan, T.; Haans, R.F.; Vakili, K.; Tchalian, H.; Glaser, V.L.; Wang, M.; Kaplan, S.; Jennings, P.D. Topic modeling in management research: Rendering new theory from textual data. *Acad. Manag. Ann.* **2019**, *13*, 586–632. [CrossRef]
36. Ahadh, A.; Binish, G.; Srinivasan, R. Text mining of accident reports using semi-supervised keyword extraction and topic modeling. *Process Saf. Environ. Prot.* **2020**, *155*, 455–465. [CrossRef]
37. Batool, A.; Byun, Y.C. Enhanced Sentiment Analysis and Topic Modeling During the Pandemic Using Automated Latent Dirichlet Allocation. *IEEE Access* **2024**, *12*, 81206–81220. [CrossRef]
38. Turkeli, S.; Ozaydin, F. A Novel Framework for Extracting Knowledge Management from Business Intelligence Log Files in Hospitals. *Appl. Sci.* **2022**, *12*, 5621. [CrossRef]
39. Altuntas, F.; Gok, M.S. A data-driven analysis of renewable energy management: A case study of wind energy technology. *Clust. Comput.* **2023**, *26*, 4133–4152. [CrossRef]
40. Pan, X.; Zhong, B.; Wang, X.; Xiang, R. Text mining-based patent analysis of BIM application in construction. *J. Civ. Eng. Manag.* **2021**, *27*, 303–315. [CrossRef]
41. Vinayavekhin, S.; Li, F.; Banerjee, A.; Caputo, A. The academic landscape of sustainability in management literature: Towards a more interdisciplinary research agenda. *Bus. Strategy Environ.* **2022**, *107*, 5748–5784. [CrossRef]
42. Chang, I.C.; Horng, J.S.; Liu, C.H.; Chou, S.F.; Yu, T.Y. Exploration of Topic Classification in the Tourism Field with Text Mining Technology—A Case Study of the Academic Journal Papers. *Sustainability* **2022**, *14*, 4053. [CrossRef]
43. Ozcan, S.; Suloglu, M.; Sakar, C.O.; Chatufale, S. Social media mining for ideation: Identification of sustainable solutions and opinions. *Technovation* **2021**, *107*, 102322. [CrossRef]
44. Tavana, M.; Shaabani, A.; Vanani, I.R.; Gangadhari, R.K. A Review of Digital Transformation on Supply Chain Process Management Using Text Mining. *Processes* **2022**, *10*, 842. [CrossRef]
45. Tavana, M.; Shaabani, A.; Santos-Arteaga, F.J.; Vanani, I.R. A Review of Uncertain Decision-Making Methods in Energy Management Using Text Mining and Data Analytics. *Energies* **2020**, *13*, 3947. [CrossRef]
46. Eddy Soria Leyva, D.P.P. Environmental approach in the hotel industry: Riding the wave of change. *Sustain. Future* **2021**, *3*, 100050. [CrossRef]
47. Mishra, M.K.; Sharma, C.; Sharma, S.; Kumar, S.; Srivastav, A.L. Exploring Antecedents, Consequences, Research Constituents and Future Directions of Circular Economy: A Predictive Analysis in the Preview of Text Mining. *J. Knowl. Econ.* **2024**, *2024*, 1–25. [CrossRef]
48. Wang, Y.; Liu, X.; Zhu, X.L. Enhancing emerging technology discovery in nanomedicine by integrating innovative sentences using BERT and NLDA. *J. Data Inf. Sci.* **2024**, *9*, 155–195. [CrossRef]
49. Al-Smadi, M.; Qawasmeh, O.; Al-Ayyoub, M.; Jararweh, Y.; Gupta, B. Deep Recurrent neural network vs. support vector machine for aspect-based sentiment analysis of Arabic hotels' reviews. *J. Comput. Sci.* **2018**, *27*, 386–393. [CrossRef]
50. Naskath, J.; Sivakamasundari, G.; Begum, A.A.S. A Study on Different Deep Learning Algorithms Used in Deep Neural Nets: MLP SOM and DBN. *Wirel. Pers. Commun.* **2022**, *128*, 2913–2936. [CrossRef] [PubMed]
51. Mammone, A.; Turchi, M.; Cristianini, N. Support vector machines. *Wiley Interdiscip. Rev. Comput. Stat.* **2009**, *1*, 283–289. [CrossRef]
52. Almunirawi, K.M.; Maghari, A.Y.A. A Comparative Study on Serial Decision Tree Classification Algorithms in Text Mining. *Int. J. Intell. Comput. Res.* **2016**, *7*, 754–760. [CrossRef]
53. Yuan, D.; Huang, J.; Yang, X.; Cui, J. Improved random forest classification approach based on hybrid clustering selection. In Proceedings of the 2020 Chinese Automation Congress (CAC), Shanghai, China, 6–8 November 2020.
54. Shani, G.; Gunawardana, A. Evaluating Recommendation Systems. In *Recommender Systems Handbook*; Springer: Boston, MA, USA, 2011; Chapter 8, pp. 257–297.
55. Grebovic, M.; Filipovic, L.; Katnic, I.; Vukotic, M.; Popovic, T. Overcoming Limitations of Statistical Methods with Artificial Neural Networks. In Proceedings of the 2022 International Arab Conference on Information Technology (ACIT), Abu Dhabi, United Arab Emirates, 22–24 November 2022; pp. 1–6. [CrossRef]
56. Shi, C.; Wei, B.; Wei, S.; Wang, W.; Liu, H.; Liu, J. A quantitative discriminant method of elbow point for the optimal number of clusters in clustering algorithm. *EURASIP J. Wirel. Commun. Netw.* **2021**, *2021*, 31. [CrossRef]

57. Rezaei, J.; van Wulfften Palthe, L.; Tavasszy, L.; Wiegmans, B.; van der Laan, F. Port performance measurement in the context of port choice: An MCDA approach. *Manag. Decis.* **2018**, *57*, 396–417. [CrossRef]
58. Lauriola, I.; Lavelli, A.; Aiolli, F. An introduction to Deep Learning in Natural Language Processing: Models, techniques, and tools. *Neurocomputing* **2022**, *470*, 443–456. [CrossRef]
59. Truong Ngoc, C.; Le Ngoc, L.; Kim, H.S.; You, S.S. Data analytics and throughput forecasting in port management systems against disruptions: A case study of Busan Port. *Marit. Econ. Logist.* **2022**, *25*, 61. [CrossRef]