

# An AI-based conceptual framework for port ranking

Yanjie Zhou<sup>a</sup>, Xiaojin Wang<sup>a</sup>, Zehao Qian<sup>b</sup>, Tao Li<sup>a</sup>

<sup>a</sup>*School of Management, Zhengzhou University, China, 315100, China*

<sup>b</sup>*Department of Nature Sciences, Durham University, DH1 3LE, UK*

---

## Abstract

In this paper, i

*Keywords:*

---

## 1. Introduction

## 2. Port ranking

### 2.1. Container port

#### 2.1.1. Seaport

Seaports are often located on the coast, facing the ocean, with deep-water channels and large dock facilities, being used for maritime cargo transportation and ship docking. Seaports also provide cargo handling, warehousing and other related services, connecting the goods flow among countries and regions. Seaports can be connected to sea and air transport to work as a part of multimodal transportation. Seaports usually have great oceanic landscapes around, and could offer various amusement facilities, which will attract large visitors to come for vacations, ensuring them convenient and comfortable experiences. So, seaports could not only develop local economy in many aspects such as tourism, manufacturing and logistics, but also provide working opportunities. However, seaports may pollute the ocean by exhausting gas from ships, discharging wastewater, litter and so on. To ensure a sustainable development and protect marine ecosystem, seaports gradually pay attention to green development by lowering bad impacts on the environment around, such as encouraging ships to use clean and new energy, exhausting gas or wastewater after treatment, building shore power facilities to attracting ships to use, establishing protected areas, etc.

#### 2.1.2. Dry port

Dry ports are seated in inland areas, serving as cargo distribution centers, usually being far from seaports. It transports goods through railways or highways. Dry port provide similar logistics services as seaports, but it pay more attention to connecting railways and highways. But dry ports are lack in the aspects of tourism and entertainment. As for developing local economy, dry ports contribute the same as seaports and river ports, but in different way.

#### 2.1.3. River port

River ports are typically situated in inland areas, with shallow-water channels and smaller dock facilities, serving the transportation of small ships and light cargo. River ports, connecting different regions through rivers, can also serve as cargo distribution centers and tourist attractions and places for leisure. River ports facilitate trades respectively. Due to the inland locations, dry ports have smaller affection on the ecological environment, but still need to contribute to protect their environment around.

## 2.2. What is port ranking?

Port ranking refers to the evaluation and ranking of ports around the world or within a specific area based on a series of indicators and data to measure how important and competitive they are in the field of international trade and shipping. These indicators and data usually include cargo throughput, container throughput, ship calls, port facilities, service levels and so on. Port ranking offers stakeholders such as shipping companies, consignors of cargo and governments a reference to estimate the competitiveness and reliability of ports and make decisions correspondingly. The results of port ranking may vary because different methods and indicators might be used for assessing ports.

## 2.3. Who cares about port ranking?

### Benefits of port ranking

#### 2.3.1. Port operator

Ports at the top of port ranking of "Port cargo, container throughput" will attract more cargoes and containers. The higher the port ranking, the greater the operational capacity of the port is. So the port ranking helps more customers to send more cargoes to top ports.

Port ranking will encourage operators to allocate resources more rationally. The latest port ranking of national port throughput from January to July 2023 shows that Shenzhen port dropped to rank 20 (chart 1). The main reason is that a large number of Asia-North America routes are suspended/cancelled, while Shenzhen Port focuses on US shipping line. To go higher on the ranking and increase container throughput at Shenzhen port, the port operator could transfer more manpower and material resources to support other busier routes

Port ranking makes the top ports more famous, which will attract more professional talents to work for the top ports, which could in return develop the ports better. Port operators will also try their best to improve work efficiency and quality.

#### 2.3.2. Shipping liner

Port ranking is a comprehensive, authoritative and reliable reference list for shipping liners to select ports for shipping. For example, through port ranking based on container port performance index, the shipping liner will find ports that save voyage time for liners, reduce transportation costs, improve ship transportation capacity, and accelerate ship turnover.

What's more, small ports do not provide enough containers and cargoes. Ship companies could easily choose ports with greater cargo and container throughput as docking points to strive for supply of goods and improve space utilization.

#### 2.3.3. Government

The port ranking is a reference for the government to do Infrastructure investments and resource allocation. The larger ports ranked in the top have greater driving capacity for the surrounding economy and higher demand for supporting facilities. The government can intuitively understand the local economic development by ports ranking and increase investments in infrastructure accordingly.

Port ranking can reflect the current situation of equipment and safety management in the port. The equipment in ports with large throughput has a large load. So, the government should supervise and strengthen the maintenance and repair of equipment in ports.

Port ranking can be a reference tool for formulating strategic planning and policies for ports. The top-ranked ports' operation, management, and construction modes are more representative and standardized. The government learned from top-ranked ports to formulate and improve the corresponding rules and systems.

### 3. The conceptual model

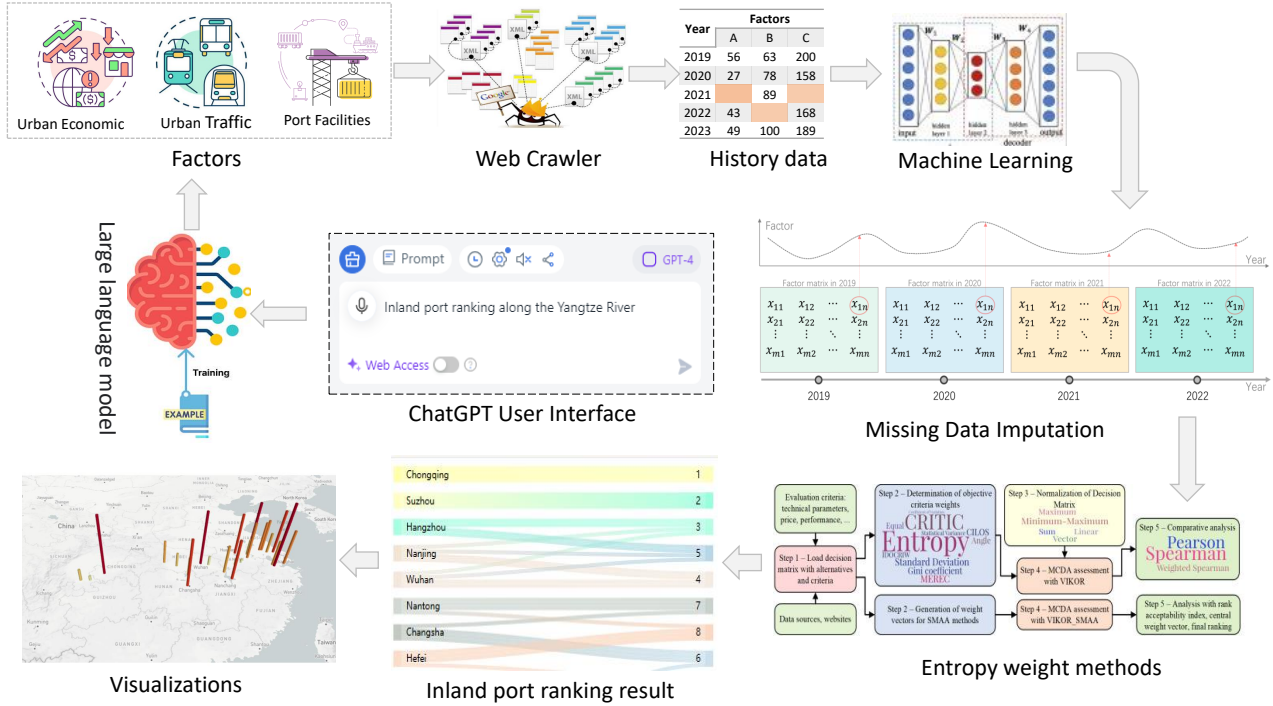


Figure 1: Caption

#### 3.1. Factors

Port ranking results that vary differently depend on network scales and centrality, and that the closer a port is to a main road or highway, the better the port development will be [O Dinu1 et al.,2016; Qing Liu,2022]. Three types of centrality measures, degree centrality, betweenness centrality and closeness centrality, are indicators to comprehensively evaluating the port importance [Chengpeng Wan et al., 2021]. The factors which affect port ranking result to different extent also include logistics cost, port service, hinterland condition and convenience level [Gitae Yeo et al., 2011]. Lu Shengrong and others build a port competitiveness evaluation index system of 6 first-class indicators: port natural environment, port infrastructures, Port service-level, port operating condition, port modernization management level, and port finance, development and image. When choosing ports to deploy the floating wind with lower cost, H. Díaz and C. Guedes Soares find ports that near the wind farms and logistic centers is more efficiently exploit, and deep-water ports with large capacity is better [2023]. Adequate infrastructures, frequency of ship visits and port location draw high attention from shipper when they select a port [CHINONYE UGBOMA et al., 2006]. Dan Zhao and others introduce green performance, port finance, market share and on-time delivery rate to evaluate a port [2021].

First-class indicators	Second-class indicators	calculation
Port inland network scale performance index	Rail network	Length of rail connection
	Road network	distance to the nearest highway in km
	Transshipment facilities for intermodal transport	YES/NO
Port centrality performance index	Number of central places in the hinterland	
	Level of localization analysis for central places	Local
		Regional
		International
	"scale" of central places	Population
	Distance (by road or rail)	km
Port Logistics System	waterway port to central place	Travel time/h
	information system of port logistics	score
hinterland condition	handling system	score
	the size and application level of the free trade zone in the area behind the port	score
convenience	the total container throughput transaction	USD
	Port natural environment (main channel depth)	km
	the level and utilization of the port information system	score
	the stability of the port labor	average working ages
logistics cost	Inland transportation freight, ship/cargo entry and exit related charges and free installation period	USD
	Terminal operating expenses	
Port facilities	numbers of loading and unloading equipments	
	Storage area	m <sup>2</sup>
Port service level	Instant service, 24/7 service, instant docking loading/unloading service	score
	The professionalism and proficiency of port operators	score
	Efficiency of cargo-handling	10,000 tons/month
	Customs clearance efficiency	
	Average layover time	
	numbers of damage cargo per year	
Port operating condition	Lane coverage rate	lanes per month
	Collection and transmission system	score
	Terminal charges level	score
	Vessel availability and extent of detention	score
Port modernization management level	EDI system	score
	Safety monitoring system	score
	GPS system	score
Port finance and development image	Average throughput growth rate	%
	Earnings per share	USD
	Whole image	score

Figure 2: Enter Caption

### 3.2. Data collection

To collect accurate data, a survey with is conducted for about 200 practitioners in the field of port operating and related. We also get data of factors from port management instituions, international trading organizations, shipping companies, port database and so on.

Type	Advantages	Disadvantages
Traditional Web Crawlers	<ul style="list-style-type: none"> <li>- Fast speed</li> <li>- Low resource consumption</li> <li>- Simple and easy to use</li> </ul>	<ul style="list-style-type: none"> <li>- Poor flexibility</li> <li>- Easily blocked by anti-crawling mechanisms</li> </ul>
Selenium-based Crawlers	<ul style="list-style-type: none"> <li>- Highly flexible</li> <li>- Can handle complex scenarios</li> </ul>	<ul style="list-style-type: none"> <li>- Slower speed</li> <li>- High resource consumption</li> <li>- High complexity</li> </ul>
LLM-based Intelligent Crawlers	<ul style="list-style-type: none"> <li>- Strong understanding ability</li> <li>- Strong adaptability</li> <li>- Capable of complex interactions</li> </ul>	<ul style="list-style-type: none"> <li>- Potential speed limitations</li> <li>- Resource-intensive</li> <li>- Challenges in accuracy</li> </ul>

### 3.3. Missing Data imputation

In modern port management and Operation Research, the integrity and accuracy of data is very important. Effective data analysis can not only reveal the current state of port operation, but also predict future trends and potential problems. However, data loss is a common problem in data collection, which can be caused by a number of factors, including technical failures, recording errors or delayed information updates. The purpose of this paper is to explore and solve this challenge in order to enhance the understanding and application of port data.

Firstly, the data collected from the port were examined to identify the patterns and possible causes of missing data. On this basis, we will explore various strategies to deal with missing data to ensure the accuracy and reliability of the analysis results. We will focus in particular on two major approaches: Imputation and Deletion. The interpolation method aims at estimating missing values and filling data gaps with these estimates, while the deletion method is to remove data records containing missing values.

**Delete or Impute?** When deciding whether to delete or impute missing data, consider the proportion and pattern of missingness, the nature of the data, the requirements of the analysis method, and the purpose of the study. Deletion is suitable for small proportions of randomly missing data and when it won't introduce bias, while imputation is preferred for large proportions of missing data, non-random missingness, to maintain dataset size and integrity, and to retain critical information. The decision should balance the characteristics of the data, the reasons for missingness, and the objectives of the analysis. [Jakobsen et al. \(2017\)](#)

#### 3.3.1. Traditional Missing Data Imputation Methods

**Mean/Median/Mode Imputation** [Kleinbort \(2020\)](#) is a common and simple method where all missing values are replaced with the mean, median, or mode of the column. Although this approach is easy and fast, it has its drawbacks. It can skew the statistical nature of the data, underestimate variance, and distort histograms. This method is generally advisable only when data is missing completely at random (MCAR) or missing at random (MAR), but it is not suitable if data is missing not at random (MNAR)

**Multiple Imputation** Sterne et al. (2009) Khan and Hoque (2020) involves creating multiple copies of the dataset, where missing values are replaced with imputed values based on their predictive distribution from observed data. This method uses standard statistical methods to fit models to each imputed dataset, and then averages the results to provide overall estimated associations. Multiple imputation, based on a Bayesian approach, accounts for all uncertainty in predicting missing values by introducing appropriate variability into the imputed values. It is essential in multiple imputation to model the distribution of each variable with missing values accurately. However, pitfalls can occur, such as omitting the outcome variable from the imputation process or dealing incorrectly with non-normally distributed variables. The validity of results from multiple imputation depends on careful and appropriate modeling.

**KNN and linear regression imputing** Aljrees (2024) Zhang et al. (2017) Murti et al. (2019) Emmanuel et al. (2021)

### 3.3.2. Imputing Missing Data with Deep Learning and LLMs

The great potential of deep learning in imputation of missing values in data. In the field of port management, these methods can be used to process and analyze large amounts of complex data, such as cargo flows, vessel dynamics, weather conditions, and port operation efficiency. Data imputation using deep learning not only improves data completeness and accuracy, but also helps predict and optimize port operations, thereby improving efficiency and safety. Mei et al. (2021) Wang et al. (2022) Park et al. (2022) Camino et al. (2019)

In order to overcome the challenge of missing values in port data, this study proposes a method to predict and interpolate these missing values using a pre-trained model Xi machine learning. We first use existing port data, including multi-dimensional information such as cargo throughput, vessel arrival frequency, weather conditions, and port facility usage, to train our machine Xi model. This data has been cleaned and preprocessed to ensure the quality of the data fed into the model. We chose a machine Xi model suitable for working with time series data, such as long short-term memory networks (LSTMs) or gated recurrent units (GRUs), because port data often have significant time-dependent and seasonal characteristics. The training of the model was performed on a rich historical dataset containing complete and missing instances of data collected from normal operations. In this way, the model learns Xi complex patterns and associations in the data, which can be used to predict missing values.

### 3.4. Entropy weight method

### 3.5. Ranking result analysis

### 3.6. Visualization of results

## 4. Conclusions

## References

- Aljrees, T., 2024. Improving prediction of cervical cancer using knn imputer and multi-model ensemble learning. PLOS ONE 19, 1–24. URL: <https://doi.org/10.1371/journal.pone.0295632>, doi:10.1371/journal.pone.0295632.
- Camino, R.D., Hammerschmidt, C.A., State, R., 2019. Improving missing data imputation with deep generative models. [arXiv:1902.10666](https://arxiv.org/abs/1902.10666).

157 Emmanuel, T., Maupong, T., Mpoeleng, D., Semong, T., Mphago, B., Tabona, O., 2021. A survey  
 158 on missing data in machine learning. *Journal of Big Data* 8, 140. URL: <https://doi.org/10.1186/s40537-021-00516-9>, doi:10.1186/s40537-021-00516-9. iD: Emmanuel2021.

160 Jakobsen, J.C., Gluud, C., Wetterslev, J., Winkel, P., 2017. When and how should multiple imputa-  
 161 tion be used for handling missing data in randomised clinical trials - a practical guide with flowcharts.  
 162 *BMC Medical Research Methodology* 17, 162. URL: <https://doi.org/10.1186/s12874-017-0442-1>,  
 163 doi:10.1186/s12874-017-0442-1. iD: Jakobsen2017.

164 Khan, S.I., Hoque, A.S.M.L., 2020. Sice: an improved missing data imputation technique. *Journal of Big*  
 165 *Data* 7, 37. URL: <https://doi.org/10.1186/s40537-020-00313-w>, doi:10.1186/s40537-020-00313-w.  
 166 iD: Khan2020.

167 Kleinbort, A., 2020. Data imputation: Beyond mean, median, and mode. *Open Data Science* URL: <https://opendatascience.com/data-imputation-beyond-mean-median-and-mode/>.

169 Mei, Y., Song, S., Fang, C., Yang, H., Fang, J., Long, J., 2021. Capturing semantics for imputation with  
 170 pre-trained language models, in: *2021 IEEE 37th International Conference on Data Engineering (ICDE)*, pp.  
 171 61–72. doi:10.1109/ICDE51399.2021.00013.

172 Murti, D., Pujianto, U., Wibawa, A., Akbar, M., 2019. K-nearest neighbor (k-nn) based missing data  
 173 imputation, in: *2019 5th International Conference on Science in Information Technology*, pp. 83–88.  
 174 doi:10.1109/ICSITech46713.2019.8987530.

175 Park, J., Muller, J., Arora, B., Faybishenko, B., Pastorello, G., Varadharajan, C., Sahu, R., Agarwal, D., 2022.  
 176 Long-term missing value imputation for time series data using deep neural networks. [arXiv:2202.12441](https://arxiv.org/abs/2202.12441).

177 Sterne, J.A.C., White, I.R., Carlin, J.B., Spratt, M., Royston, P., Kenward, M.G., Wood, A.M., Carpen-  
 178 ter, J.R., 2009. Multiple imputation for missing data in epidemiological and clinical research: potential  
 179 and pitfalls. *BMJ* 338. URL: <https://www.bmj.com/content/338/bmj.b2393>, doi:10.1136/bmj.b2393,  
 180 [arXiv:https://www.bmj.com/content](https://arxiv.org/abs/https://www.bmj.com/content).

181 Wang, Z., Akande, O., Poulos, J., Li, F., 2022. Are deep learning models superior for missing data imputation  
 182 in large surveys? evidence from an empirical comparison. [arXiv:2103.09316](https://arxiv.org/abs/2103.09316).

183 Zhang, S., Li, X., Zong, M., Zhu, X., Cheng, D., 2017. Learning k for knn classification. *ACM Transactions on*  
 184 *Intelligent Systems and Technology* 8, 1–19. doi:10.1145/2990508.