Acadlore Transactions on AI and Machine Learning
https://www.acadlore.com/journals/ATAIML

A Comparative Study on AI-Based Algorithms for Cost Prediction in Pharmaceutical Transport Logistics

Fadwa Farchi[1*], Chayma Farchi[1], Badr Touzi[2], Charif Mabrouki[1]

[1] Department of Mechanical and Electrical Engineering, Faculty of Sciences and Technologies, Hassan 1st University, 26000 Settat, Morocco
[2] Department of Economics and Management Sciences, Faculty of Legal, Economic and Social Sciences – Souissi, Mohammed V University, 10000 Rabat, Morocco

[*] Correspondence: Fadwa Farchi (F.farchi@uhp.ac.ma)

Received: 06-20-2023 **Revised:** 08-13-2023 **Accepted:** 08-20-2023

Citation: F. Farchi, C. Farchi, B. Touzi, and C. Mabrouki, "A comparative study on AI-based algorithms for cost prediction in pharmaceutical transport logistics," *Acadlore Trans. Mach. Learn.*, vol. 2, no. 3, pp. 129–141, 2023. https://doi.org/10.56578/ataiml020302.

Abstract: Pharmaceutical transport logistics, especially in humanitarian and hospital contexts, is becoming increasingly essential with a growing need to monitor associated costs. In Morocco, however, studies focusing on the cost implications of pharmaceutical delivery conditions are conspicuously absent. This creates a high-dimensional classification framework, where the selection of variables becomes challenging in the face of correlated distribution predictors. The integration of Artificial Intelligence (AI) in cost prediction has emerged as a vital necessity amidst escalating complexities and cost considerations. Cost prediction, being inherently correlated with almost all variables and inputs, offers an interpretable value in performance management, financial planning, and contract negotiation. This study undertakes a comparative analysis of a broad spectrum of prediction algorithms applied to the same, albeit reduced, database. A dozen such algorithms are put into practical use, with variable selection implemented through importance measures. The primary objective of this comparative evaluation is to determine the superior performing algorithm — one that delivers optimal adaptation to the context within a fixed environment. The prediction algorithm incorporates a myriad of inputs and constraints derived from data collection systems. AI's application facilitates the inclusion of diverse variables such as transportation routes, congestion, distances, freight weight, and environmental factors, thereby enhancing the accuracy and efficiency of cost estimation. The Orthogonal Matching Pursuit model emerged as the most successful, boasting an R^2 value nearing unity. Accurate cost prediction in transport can yield valuable insights into budgeting, estimation, customer service, managerial risk, environmental considerations, and strategic deployment for a company. Improved decision-making and resource allocation can thereby be achieved, leading to enhanced profitability and sustainability.

Keywords: Artificial intelligence; Comparative study; Metrics; Orthogonal Matching Pursuit; Pharmaceutical transportation

1 Introduction

The persistent issues facing urban logistics remain challenging due to conflicting priorities around cost, time, and developing suitable technologies. Scientific literature demonstrates artificial intelligence's (AI) potential through critical reviews and monitoring of road transport infrastructure surveillance utilizing ground penetrating radar [1], as well as recommended energy transport and CO_2 emissions modeling [2], or evidenced in studies of emerging smart city technologies applying geospatial data, advanced analytics and machine learning approaches in internet of things frameworks [3]. Machine learning (ML) and deep learning (DL) methods have enhanced predictive, planning and uncertainty models across diverse aspects of urban development [4]. Traditional frameworks still play an important role in managing transport flows across most industrial sectors, though targeted publications addressing pharmaceutical or medical logistics appear limited.

The pharmaceutical industry stands out as a major player in urban freight distribution and transport [5], especially where hospitals and pharmacies in close proximity receive deliveries multiple times daily from warehouses and wholesalers located nearby. Morocco notably hosts over 12,000 pharmacies and 768 hospitals [International Trade Administration], forming fertile ground for research in this sector.

Pharmaceutical transport in Morocco, as elsewhere, confronts several product- and process-specific hurdles relating to sensitivities and quality/safety prerequisites [6]. A core challenge lies in cold chain and temperature control givens drugs' and biologics' temperature sensitivity [7]. Distribution thus adheres to strict regulation of quality, safety and compliance. Moreover, transportation costs proving high especially for storage-conditioned deliveries stem from cold chain and security needs. Further, studies along these lines remain somewhat dated and sparse in literature [8].

Artificial intelligence (AI) can play a significant role in enhancing the transport of pharmaceutical products by providing innovative solutions and operational benefits, such as optimization of supply chains, route planning, inventory management and distribution of pharmaceutical products. It can take into account various constraints including temperature, warehouse availability, cold chain monitoring, cost and demand prediction by analyzing historical and real-time data to predict fluctuations in transportation costs and demand. In more complex cases, AI can be leveraged to customize delivery solutions based on individual customer needs, optimizing lead times and routes.

The distribution of pharmaceuticals is a topic worthy of thorough analysis and ongoing monitoring. In this context, controlling several parameters is crucial to managing the total cost of transport, which depends on factors like departure and arrival coordinates, tire pressure, transport temperatures, freight weight, fuel consumption, and carbon dioxide emissions – the latter being highly correlated with total costs. Fuel used by vehicles transporting goods in urban logistics is widely recognized as one of the primary causes of air pollution in urban areas [9], particularly regarding pharmaceutical distribution that requires multiple daily trips largely accomplished by fleets emitting carbon dioxide emissions in need of investigation and mitigation.

Several studies described in literature and conducted across different cities concerning urban distribution of pharmaceutical products have highlighted the most significant aspects requiring improvement in this sector, such as reducing lead time, carbon dioxide emissions, costs, vehicle volumes as well as daily trips [10].

The purpose of this study is to evaluate and compare the performance of multiple algorithms within the context of pharmaceutical transportation. By testing and analyzing diverse algorithms, we aim to identify the most effective and efficient approach for forecasting total costs of these complex logistical operations. This evaluation will support decision-makers, logistics managers, and transportation providers regarding algorithm selection and implementation for pharmaceutical transport needs.

The testing process will involve collecting relevant data, including characteristics of the transportation network, shipment volumes, distances, delivery time windows, and any constraints specific to pharmaceutical products. Several algorithms, such as heuristics, metaheuristics, mathematical optimization models, and simulation techniques, will be selected based on their suitability for optimizing pharmaceutical transport.

Accurately predicting costs is crucially important for planning and allocating resources within an organization. By anticipating future costs associated with various projects and operations, the company can develop accurate budgets and efficiently allocate the necessary resources. This strategic approach avoids overspending, minimizes waste and ensures optimal use of financial and material resources. In addition, cost prediction plays a critical role in performance management by enabling objective and continuous performance evaluation. By comparing actual costs to forecasts, discrepancies can be identified, underlying causes understood, and strategies adjusted accordingly to improve operational efficiency. Transparency is also enhanced through clear and transparent communication of cost predictions to internal and external stakeholders. This transparency fosters trust, facilitates informed decision-making and strengthens relationships with customers, investors and business partners, thereby contributing to more responsible and results-oriented management. Paradoxically, optimization aims to determine the inputs required to attain a given cost, whereas this study seeks to forecast costs as an output through predictive analysis.

2 Background on Pharmaceutical Transport

At this juncture, a review of the background context both globally and within Morocco is merited.

2.1 General Background

The pharmaceutical industry relies on efficient and secure transportation to ensure the timely delivery of life-saving medications and healthcare products. With advancements in artificial intelligence (AI), there is growing interest in leveraging AI technologies to optimize various aspects of pharmaceutical transport. AI-driven solutions have the potential to enhance the efficiency, accuracy, and safety of pharmaceutical logistics operations.

Logistics management, or the value chain, represents one of the most crucial steps in the pharmaceutical sales sector given transport and storage costs accounting for over 40% of total logistics expenses. According to the study [11], in such systems generally solutions like RFID (radio frequency identification), IoT (Internet of Things), or blockchain are employed for traceability, data storage, and processing [12]. However, within this specific domain predictive studies of transport costs utilizing machine learning have yet to be found.

The successful integration of technologies such as RFID, Internet of Things (IoT) and blockchain in pharmaceutical transportation has profoundly and enduringly transformed how medical products are managed and delivered. RFID enables real-time location tracking and careful monitoring of pharmaceutical shipments, heightening visibility and traceability across every stage of the supply chain. IoT, meanwhile, broadens this scope by providing real-time data on critical parameters like temperature, humidity and vibration, ensuring sensitive products remain in optimal condition. Blockchain, through ensuring the immutability and transparency of records, builds trust by confirming track and trace data cannot be tampered with. Collectively, these technologies cultivate an environment where safety, quality and compliance are enhanced, elevating pharmaceutical transportation to new heights of efficiency, reliability and trust.

AI algorithms and techniques can be applied to address various challenges in pharmaceutical transport. Such challenges include route optimization, real-time tracking and monitoring, temperature control, regulatory adherence, risk management, and supply chain visibility. By harnessing the power of AI, pharmaceutical companies, logistics providers, and healthcare organizations can enhance overall performance of their transportation operations and better serve patients' needs. As noted above, the focus will lie in cost predictions to furnish an estimate approximating reality closely.

What distinguishes our study versus others is the utilization of other researchers' outcomes in the form of means to examine transport costs. Namely, we leverage both traceability modalities and intelligent and manual data gathering to profile an output as well as data processing and cleaning methodologies while ensuring secure archiving of data.

2.2 Moroccan Background

In Morocco, adoption of AI in the pharmaceutical industry is gaining momentum. The country is witnessing advancements in technology and digital transformation, creating opportunities for integrating AI across various sectors, including pharmaceutical transport. Implementing AI-based solutions in pharmaceutical logistics can significantly benefit Morocco's healthcare system and enhance medical product delivery nationwide. In doing so, specific challenges faced by the industry may be addressed, such as optimizing routes for efficient and timely deliveries, monitoring temperature-sensitive pharmaceuticals during transportation, and ensuring regulatory compliance [6].

AI-powered predictive analytics and machine learning algorithms can aid in demand forecasting, inventory management, and supply chain optimization. By analyzing historical data, AI models can identify patterns and trends, enabling better decision-making regarding stock levels, distribution strategies, and demand planning.

As adoption of AI technologies in pharmaceutical transport expands in Morocco, essential considerations include potential challenges relating to data privacy, regulatory compliance, and infrastructure development. Collaboration between industry stakeholders, technology providers, and regulatory bodies is critical to ensuring the successful integration and deployment of AI solutions in pharmaceutical logistics, contributing to Morocco's healthcare ecosystem advancement [8].

Researchers' statements and the literature highlight growing adoption of AI in Morocco's pharmaceutical industry, driven by technological progress and ongoing digital transformation. Integrating AI across various sectors, including pharmaceutical transportation, presents significant opportunities to improve healthcare services. This integration has potential to address specific industry challenges.

It merits noting that AI contributions remain relatively nascent in Morocco across all fields, including pharmaceutical - where it functions more as a logistics management tool. However, like other researchers, we perceive its usefulness.

3 Description of the State of Play

The nature of pharmaceutical product transport and delivery in Morocco is quite specialized as it is generally subcontracted to dedicated logistics providers. Working collaboratively with these transport partners, we aim to build a robust model by drawing on data collected from the pharmaceutical distribution process in the city of Marrakech (located in central Morocco). We applied our forecasting model to travel management within the pharmaceutical field, more specifically on transporting pharmaceutical goods. These include products from pharmaceutical laboratories or biotechnology companies.

We chose this focus due to the sensitive nature of transported products susceptibility to damage if transport fails to meet specific conditions, and the importance of the pharmaceutical sector overall. This product transport closely mirrors transport of edible or chemical liquids or fresh/dry agricultural goods as they are heat-sensitive and can rapidly deteriorate due to various factors.

Upcoming sections will delve into transport service nature, data collection, and metric selection. Understanding logistics intricacies is crucial for optimizing operations and ensuring efficient goods movement. Additionally, we will discuss the importance of data collection in transport logistics, highlighting various data sources and significance for performance, tracking and satisfaction insights. Furthermore, we will address the critical task of choosing appropriate metrics to measure and evaluate transport service performance, through which we will identify the most efficient program and define it as a predictive model.

Key performance indicators (KPIs) will define algorithms' performance assessments. Potential KPIs include metrics like total transport time, on-time delivery rate, cost-efficiency, resource utilization, delivery route reliability, and flexibility to accommodate dynamic demand patterns. Algorithms will be implemented and executed using collected data, with performance evaluated against established KPIs.

Evaluation results will provide insights into each algorithm's strengths/weaknesses, highlighting suitability for scenarios/requirements in pharmaceutical transport. Factors like scalability, robustness, adaptability to changing conditions and computational efficiency will also factor into the evaluation process.

3.1 An Outsourced Service

The transport of medical products imported into Morocco is commonly subcontracted. Outsourcing transport has emerged as a prominent strategy within the healthcare sector. With globalization and expanding pharmaceutical markets, companies increasingly rely on external partners for product transportation. This approach offers multiple advantages, including enhanced efficiency, cost savings, and access to specialized expertise.

External logistics providers possess the necessary infrastructure, resources, and regulatory acumen to ensure pharmaceuticals' safe, secure international transit. They adhere to stringent quality standards, employing temperature-controlled solutions to preserve sensitive products' integrity and efficacy. By outsourcing transport, pharmaceutical corporations can prioritize core competencies like research and development, entrusting logistics to reliable partners for streamlined operations and improved customer satisfaction [13]. There are several reasons why Morocco may choose to outsource the domestic transportation of pharmaceutical products:

(1) Expertise and Infrastructure: External logistics providers often possess specialized knowledge and experience in safely handling pharmaceutical transportation. They have robust infrastructure including temperature-controlled storage, vehicles, and tracking systems to ensure product integrity throughout the supply chain.

(2) Regulatory Compliance: Transporting pharmaceuticals involves stringent adherence to regulations and standards like Good Distribution Practices and temperature monitoring. External providers have deep familiarity with these compliance requirements and can navigate them deftly.

(3) Cost Efficiency: Establishing and maintaining an in-house transportation network for pharmaceuticals can be costly. Outsourcing to external providers allows optimized resource allocation, reducing overhead expenses associated with vehicle fleets, storage facilities, and personnel. Providers may also have negotiated competitive shipping rates, enabling potential cost savings.

(4) Focus on Core Activities: Outsourcing transportation frees internal resources to concentrate on primary objectives like manufacturing, research and development, and healthcare delivery. Entrusting logistics to providers promotes innovation, enhanced products and patient care.

(5) Scalability and Flexibility: Morocco's pharmaceutical industry may face fluctuating demand or require market expansion. Outsourcing permits agile adaptation to changing needs through leveraging a provider's pre-existing capabilities and network without sizable upfront investment.

3.2 Data Collection

Transportation problems have persisted in Moroccan cities as urbanization advanced, mirroring challenges faced by other nations for many years. Issues addressed in this paper include road congestion, traffic, air pollution, traffic accidents, and high costs of transportation. Specifically, we choose to model unforeseen costs.

After selecting the applicable field and mode of transport, data consolidation begins.

Data was collected over four calendar weeks. Sources include statistics provided by drivers and distribution handlers, vehicle dashboard figures, GPS, various detectors and sensors, and estimated motion environment factors. Additional client-related information will be detailed below.

Utilizing a modest dataset of approximately 300 observations presents advantages and challenges. One benefit lies in significantly reduced computational requirements and processing time for testing multiple algorithms versus larger datasets, enabling more efficient experimentation and evaluation. A smaller set also allows more comprehensive exploration of algorithmic capabilities by permitting evaluation of a broader range of algorithms within reasonable time frames.

However, a small dataset also presents challenges. The limited size may increase risk of overfitting, where algorithms perform well on training data but fail to generalize to new, unseen instances. Carefully assessing robustness and generalization abilities becomes crucial, as performance could heavily depend on the dataset's characteristics and idiosyncrasies.

We exercised great care in data collection to represent actual normal process operation and behavior. We note the duration represents a typical month's pattern. This carrier repetitively follows the same monthly transport routine: delivering preset monthly quantities. Rarely, exceptional overdemand aligns supply and transforms quantities in line with needs. Thus, despite its modesty, the number of observations accurately portrays reality.

Below, Table 1 consolidates all variables and constraints considered in this study.

Table 1. Variables' definitions

Variables' Designation	Definitions
Vehicle per fleet	The company's transport fleet consists of internal vehicles responsible for transporting pharmaceutical products within each city.
Fleet number	The number of fleets means of transport each comprising n refrigerated vehicles.
Driver ID	Each of the drivers has an identifier that allows the monitoring part to associate a trip with a given driver.
Trip ID	Similar to the Driver ID, each trip is identified and tracked by an identifier
Departure coordinates	Start coordinates are the GSP coordinates of the location through which a journey is initiated, expressed in decimal degrees. It includes 2 elements C1: Latitude i; C2: Longitude i
Arrival coordinates	Arrival coordinates are the GPS coordinates of the destination to which a given trip is destined, expressed also in decimal degrees. It includes 2 elements: C3: Latitude f; C4: Longitude f
Delta T - Min	This is the length of time an ID trip takes from departure to arrival. This duration is expressed in minutes as the distances and durations are short. Later will be expressed in Hour for the sake of the unification of units of measurement.
Distance - Km	Distance of a trip traveled between 2 points (Departure-Arrival). It is expressed in kilometers.
Tire Pressure Xi / I from 9 to 12	Pressure defines the amount of air inside the tire. Good pressure is essential to drive safely. Tires that are not standard cause overconsumption of energy and do not help the auto-driving. Tire Pressure 1: Drive wheels, located in front of the car. Tire Pressure 2: Drive wheels, located in front of the car. Tire Pressure 3: Steering wheels are the wheels That are in the rear part of the vehicle. Tire Pressure 4: Steering wheels are the wheels that are in the rear part of the vehicle.
Temperature	This is the temperature maintained in the refrigerated part of the vehicle. This temperature varies depending on the type of pharmaceuticals transported as well as what is required. Expressed in Celsius.
Weight freight	It is the weight carried on a specific trip. This point varies and decreases along the distribution to be fed again on the following working day. Expressed in Kilogram.
Delta T - H	Travel time is expressed this time in hours.
Speed - Km/H	The speed with which the goods are transported. The latter affects the quality and duration of the delivery.
Consumption per trip - L	It is the consumption in terms of energy (liters of diesel) consumed during a trip.
Emission CO_2 - kg	Ecology CO_2 emissions are all emissions of carbon dioxide into the atmosphere by urban vehicles in our case.
Number of stops per trip	By motion sensors, the number of times the engine of the vehicle has been declared stopped following the stopping of the vehicle is quantified.
Level of service	An estimate of the level of delivery service for the goods, which varies from one trip to another depending on the quality of the goods received, the deadlines and the conformity of the products delivered.
Unforeseen delays	Delay not taken into account during planning. Excluding loading, unloading transport and payment times. This includes contingencies and issues not taken into account. Expressed in Hour.
Number of vehicles	Representation of vehicles circulating in a location x at a time t (Veh)
Concentration	Concentration in terms of vehicles occupying the neighboring space per unit length expressed in Veh/M
Flow speed	The average space velocity of vehicles occupying space [x-Δx, x] expressed in M/S
Rate of flow	Number of vehicles flowing towards a point of abscissas during a time interval [t, t+Δt] expressed in (Veh/S)
Fossil energy consumption per ton of freight	Any energy consumed by the vehicle when transporting a Ton of weight, expressed in Kilogram.

3.3 Data Collection

Following algorithm creation and parameterization within the coding environment, model performance evaluation ensues. Commonly employed metrics guide selection of the most representative.

(1) Mean Absolute Error (MAE) measures the average absolute difference between predicted and actual values. It provides a direct, easy-to-interpret gauge of predictive accuracy. MAE assesses predictive errors' magnitude, crucial in our case where true output costs are compared to estimated values from the cost-predictive model.

$$MAE = \frac{\sum_{i=1}^{n} |e_i|}{n} \tag{1}$$

(2) Mean Squared Error (MSE) calculates the average of the squared differences between predicted and actual values. It amplifies larger errors due to squaring. MSE proves useful for penalizing substantial errors and commonly features in optimization algorithms. However, it lacks straightforward interpretability in the original data units.

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (Y_i - Y_i')^2 \tag{2}$$

(3) Root Mean Squared Error (RMSE) constitutes the square root of MSE, rendering a metric directly interpretable in the initial data units. RMSE serves extensively as a performance metric sensitive to sizeable errors, facilitating comprehension of a model's overall prowess.

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (Y_i - Y'_i)^2} \tag{3}$$

(4) R^2 (R-squared) denotes the coefficient of determination, conveying the proportion of variation in the dependent variable explained by the model. It gauges model fit to data, ranging from 0 to 1, where higher values signify superior data fitting.

$$R^2 = 1 - (\text{ SSres / SStot }) \tag{4}$$

(5) Root Mean Log Squared Error (RMLSE) calculates the average logged difference between predicted and actual values. It proves useful when the target variable exhibits skewed distribution. RMLSE penalizes both underestimation and overestimation, providing an accuracy measure on a logarithmic scale.

$$\text{RMLSE} = \sqrt{\frac{1}{n} \sum (\ln(\text{y\_pred} + 1) - \ln(\text{y\_actual} + 1))^2} \tag{5}$$

(6) Mean Absolute Percentage Error (MAPE) computes the average absolute percentage difference between predicted and actual values. MAPE furnishes a relative accuracy gauge, proving beneficial when comparing models across disparate datasets or scales. It elucidates predictive errors as a percentage.

$$MAPE = \frac{100\% \sum_{t=1}^{n} \left| \frac{A_t - F_t}{A_t} \right|}{n} \tag{6}$$

(7) Computation Time (CT) quantifies an algorithm's training or predictive run time. It assesses efficiency and scalability importance, especially for real-time or large-scale use, where briefer computation favors feasibility. Shorter time underscores desirability when weighed against predictive prowess gauged via additional performance metrics.

Each metric holds distinct significance and relevance contingent on contextual problem needs and requirements. They provide varied lenses into model performance and accuracy, empowering researchers and practitioners with effective evaluation and algorithm comparison. Selecting the most suitable metric(s) derives from specific aims, data traits, and how different error categories functionally impact goals within a given domain.

4 Methodology

This paper performs comparative algorithm evaluation, requiring holistic metric assessments for informed selection. Commonly analyzed metrics like Accuracy, Precision, Root Mean Log Squared Error, area under the curve (AUC), Mean Squared Error (MSE), and Computational Time (CT) provide unique insights into aspects like predictive precision, tradeoffs between false positives and negatives, overall performance, predictive deviations, and efficiency. Comprehensive examination ensures an algorithm aligning with specific task needs and aims.

Before comparative analysis, methodological process diagramming proves essential. Initial data collection derives from carrier histories and sensors. Next, value processing and cleaning. Then, evaluation method choices and performance calculation. Followed by tuning improvement. Lastly, chosen model definition. The diagram below delineates these steps.

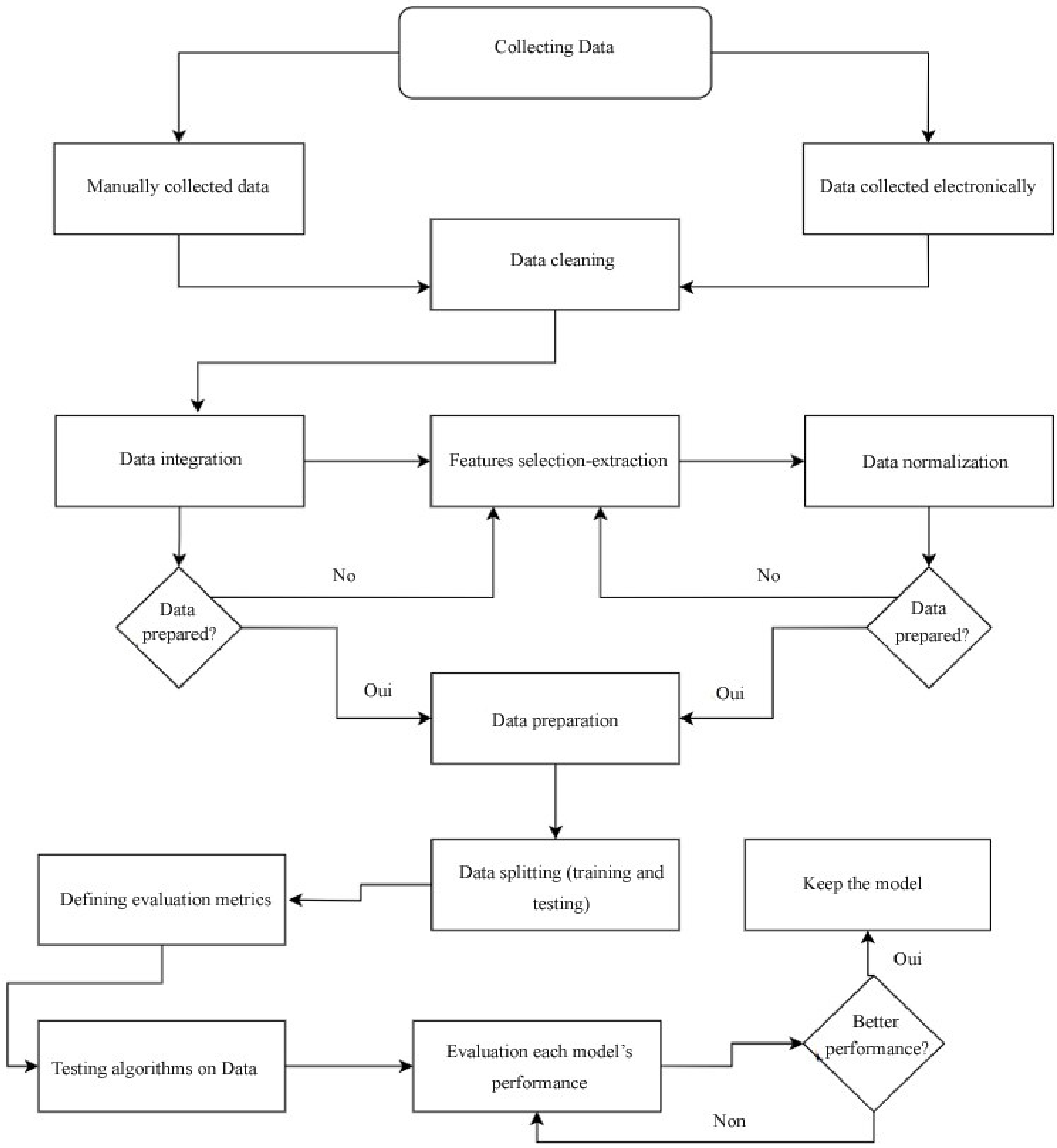


Figure 1. Process flowchart

4.1 Data Preparation

This part develops the preparation of the data, which we represent as a flowchart at the level of Figure 1.

Preparing and processing our pharmaceutical transportation data are crucial steps in building effective models. These steps involve transforming raw data (blank data from collection systems written as recorded) into a suitable format that can be used for training and testing phases. The process typically includes the following key stages.

Preparing and processing pharmaceutical transportation data comprise critical early steps for developing effective models. These activities transform raw collected information into a suitable format supporting training and testing. Typical processes include:

(1) Data collection: Gathering relevant variables from structured sources like databases and specially configured collection interfaces, or external sensor-derived datasets capturing transport characteristics like speed, temperature, or congestion estimates. Captured data should represent the problem domain adequately.

(2) Data cleansing: Identifying and managing missing, outlier, and inconsistent values through methods like imputation, detection, and validation to ensure quality and integrity.

(3) Data integration: Combining datasets from multiple origins (sensors, manual records, estimates) into a unified structure retaining all pertinent material consistently.

(4) Feature selection/extraction: Analyzing variable interrelationships and isolating the most informative characteristics contributing to the problem for model performance improvement and computational streamlining. Either deriving useful features from processed data or selecting a quality-focused subset.

(5) Data transformation/normalization: Standardizing data scales or distributions essential for algorithms sensitive to value ranges/distributions, applying techniques such as logarithmic transformation. Necessary here as data encompass diverse scales requiring conversion to compare and examine meaningfully as mentioned in Figure 2.

```
    return train,test,target,ignored_columns
```

```
In [73]: %%time
train,test,target,ignored_columns = load_data(ignored_columns=['Date', 'Vehicle per fleet', 'Fleet Number', 'Trip ID'],target="Tr

[INFO] We have detected test data
[INFO] Checking if the columns match ..
[INFO] No difference found..
CPU times: user 1.04 s, sys: 3.69 ms, total: 1.04 s
Wall time: 1.04 s
```

Model training

Getting data report

```
In [74]: reg = setup(data = train, target = target, normalize=True, silent=True,ignore_features=ignored_columns,fold_shuffle=True, sessior
```

Figure 2. Normalizing part in the code

(6) Data splitting: Partitioning the dataset into training, validation, and testing subsets, with the training set utilized for model training, validation set for hyperparameter optimization, and holdout test set reserved to appraise the final model performance independently. At this stage, specific split proportions have not been defined, remaining flexible to suit each applied algorithm.

After checking the data preparation phase, we move on to the configuration of the programming parameters. Part of the parameterization is presented in Figure 3.

	Description	Value
0	session_id	2
1	Target	Transportation Cost
2	Original Data	(307,28)
3	Missing Values	True
4	Numeric Features	16
5	Categorical Features	7
6	Ordinal Features	False
7	High Cardinality Features	False
8	High Cardinality Method	None
9	Transformed Train Set	(214,110)
10	Transformed Test Set	(93,110)
11	Shuffle Train-Test	True
12	Stratify Train-Test	False
13	Fold Generator	KFold
14	Fold Number	10
15	CPU Jobs	-1
16	Use GPU	False
17	Log Experiment	False
18	Experiment Name	reg-default-name
19	USI	3353
20	Imputation Type	iterative
21	Iterative Imputation Iteration	5
22	Numeric Imputer	mean
23	Iterative Imputation Numeric Model	Light Gradient Boosting Machine
24	Categorical Imputer	constant
25	Iterative Imputation Categorical Model	Light Gradient Boosting Machine
26	Unknown Categoricals Handling	least_frequent
27	Normalize	True
28	Normalize Method	zscore
29	Transformation	False
30	Transformation Method	None
31	PCA	False
32	PCA Method	None
33	PCA Components	None

Figure 3. Setting up descriptions

Prior to processing, requisite libraries are imported to enable full functionality of forthcoming Python-based functions as shown in Figure 4. For AI models to discern patterns and generate precise predictions/classifications, this preparation proves crucial. Libraries aid enhanced model performance, bolstered generalization, and reliable outcomes across AI applications by contributing to learning robustness.

```
In [66]: import numpy as np
         from pycaret.regression import *
         import pandas as pd
         from sklearn.model_selection import train_test_split
         import re,os
         from time import sleep
         from IPython.core.interactiveshell import InteractiveShell
         InteractiveShell.ast_node_interactivity = "all"
         import sys
         if "openpyxl" not in sys.modules:
             !pip install openpyxl > /dev/null 2>&1

In [67]: test_path = "./test_1.xlsx"
         train_path = "./train_1.xlsx"
         pred_path = "./pred.xlsx"
         model_name = "model_2"
```

Data Preprocessing

Figure 4. Importing libraries

4.2 Model's Selection

Table 2. Key performances of all tested models

Model	MAE	MSE	RMSE	R^2	RMILSE	MAPE	CT (sec)
Orthogonal Matching Pursuit	0.0126	0.0009	0.0278	1	0.0025	0.0011	0.039
Passive Aggressive Regressor	0.1906	0.202	0.3522	0.9984	0.0149	0.0098	0.065
Huber Regressor	0.1665	0.2442	0.3789	0.998	0.0134	0.0073	0.193
Ridge Regression	0.3242	0.6101	0.6136	0.995	0.0208	0.0146	0.043
Bayesian Ridge	0.1367	2.7444	0.5595	0.9781	0.0224	0.0045	0.168
Gradient Boosting Regressor	0.6259	3.7458	1.3794	0.9704	0.051	0.0279	0.301
Extra Trees Regressor	0.8068	4.7937	1.7443	0.9631	0.0609	0.0344	0.695
Random Forest Regressor	0.9368	5.3881	1.9475	0.9567	0.065	0.0389	0.83
Lasso Regressor	1.3358	5.8401	2.1777	0.9529	0.0838	0.0643	0.07
AdaBoost Regressor	1.4497	6.8977	2.3785	0.9438	0.0874	0.067	0.333
Decision Tree Regressor	1.198	8.1012	2.4166	0.9388	0.0766	0.0481	0.047
Light Gradient Boosting Machine	1.396	8.7488	2.6621	0.9318	0.0849	0.0568	0.146
Elastic Net	2.3108	12.1637	3.3469	0.9029	0.1307	0.1134	0.037
K Neighbors Regressor	2.556	17.9078	3.9477	0.858	0.1481	0.1167	0.059
Lasso Least Angle Regression	9.285	127.5931	11.2039	-0.0246	0.5264	0.5862	0.04
Dummy Regressor	9.285	127.5931	11.2039	-0.0246	0.5264	0.5862	0.028
Linear Regression	1.9324	218.3226	8.8808	-0.8446	0.1485	0.0575	1.067
Least Angle Regression	8060791 4551 37.2637	5981941576 8894858439 24779008	36019212 48272 6.5156	-84320506 92063587 77849088	18.4241	353170 84282 7.7862	0.1490

Upon comparative study completion, we select the top-performing model via highest metric scores. Specific values evidence Orthogonal Matching Pursuit (OMP) achieving best outcomes. For pharmaceutical transport forecasting contexts, OMP stands out suitably because of inherent multi-dimensional data handling and relevant feature identification abilities while promoting sparse model representation.

Pharmaceutical transport prediction often involves myriad variables like routes, temperatures, schedules and external factors. OMP adeptly captures complexity and interdependencies within this system by accommodating high-dimensionality. Furthermore, such data commonly exhibit sparse patterns where only select attributes meaningfully contribute to prediction. OMP excels at discerning and capitalizing on this sparsity through iterative selection of most impactful features, optimizing precision. This dimensionality reduction enhances computational efficiency and prevents overfitting. Table 2 summarizes performance calculations for each approach.

Apart from this choice, others also show a very significant performance such as: Passive Aggressive Regressor, Huber Regressor, Ridge Regression, Gradient Boosting Regressor, Extra Trees Regressor, Random Forest Regressor with accuracy values respectively equal to “0.9984; 0.9980; 0.9950; 0.9704; 0.9631; 0.9567”.

While others deviate outright and can neither be tuned or parameterized to give better results. The algorithm least suited to the context of pharmaceutical transport and the nature of our data is the “Least Angle Regression” with a MAE “8060791455137.2637” very large in relation to the nature of the values.

4.3 Model’s Description

Within our contextual variable set, OMP can establish relationships between characteristics and associated transport costs. The process involves several steps.

First, OMP initially selects a small number of highly impactful independent variables (characteristics/predictors) on cost. At each step, OMP chooses the variable most correlated with transportation cost, sequentially prioritizing consumption, distances, environmental factors, congestion, and lastly customer satisfaction - maximizing variable trend similarity to observed actual cost.

For example, among distances traveled, fuel consumption per trip, weight transported and other factors, OMP identifies the cost-closest variable.

Once the first variable is chosen, OMP attempts adding other variables providing significant predictive contribution while minimizing redundancy. This occurs by seeking other characteristics orthogonal (uncorrelated) to those selected.

For example, the algorithm can add fuel consumption per tonne transported as an additional, separately relevant feature informing total cost prediction.

OMP iterates until reaching a stopping criterion like a set number of selected variables or satisfactory prediction accuracy (in our case, precision reaching 1 in the Figure 5 below).

With characteristics selected, OMP then builds a cost prediction model applicable to new data, estimating future costs based on selected characteristic values.

The Orthogonal Matching Pursuit (OMP) algorithm is particularly well suited for predicting costs in a small data set due to its feature selection nature. It considers only the most influential variables to predict cost, which is crucial since data is limited and it is necessary to focus on the most significant factors. Then, as we are working on a small data set, dimension reduction is essential to avoid information overload. The OMP selects a small number of the most informative features, which reduces the complexity of the model while maintaining good prediction accuracy. It ensures maximizing the similarity between the selected features and the actual data, resulting in high prediction accuracy even with a small data set. It focuses on the variables that have the greatest impact on the target variable (in this case, cost).

Fold	MAE	MSE	RMSE	R2	RMSLE	MAPE
0	0.0090	0.0003	0.0173	1.0000	0.0018	0.0009
1	0.0097	0.0003	0.0166	1.0000	0.0007	0.0005
2	0.0066	0.0001	0.0099	1.0000	0.0009	0.0005
3	0.0098	0.0003	0.0178	1.0000	0.0019	0.0009
4	0.0152	0.0011	0.0335	1.0000	0.0034	0.0013
5	0.0168	0.0011	0.0325	1.0000	0.0024	0.0012
6	0.0130	0.0006	0.0240	1.0000	0.0023	0.0010
7	0.0201	0.0028	0.0529	1.0000	0.0058	0.0021
8	0.0122	0.0006	0.0250	1.0000	0.0016	0.0009
9	0.0178	0.0009	0.0297	1.0000	0.0029	0.0015
Mean	0.0130	0.0008	0.0259	1.0000	0.0024	0.0011
Std	0.0041	0.0007	0.0115	0.0000	0.0014	0.0005

(a)

Fold	MAE	MSE	RMSE	R2	RMSLE	MAPE
0	0.0093	0.0003	0.0173	1.0000	0.0018	0.0009
1	0.0071	0.0002	0.0150	1.0000	0.0008	0.0004
2	0.0083	0.0002	0.0134	1.0000	0.0009	0.0006
3	0.0105	0.0005	0.0214	1.0000	0.0023	0.0010
4	0.0129	0.0017	0.0411	1.0000	0.0043	0.0013
5	0.0126	0.0008	0.0286	1.0000	0.0016	0.0008
6	0.0118	0.0006	0.0251	1.0000	0.0026	0.0010
7	0.0199	0.0027	0.0520	1.0000	0.0052	0.0019
8	0.0122	0.0006	0.0250	1.0000	0.0016	0.0009
9	0.0220	0.0015	0.0381	1.0000	0.0039	0.0020
Mean	0.0126	0.0009	0.0278	1.0000	0.0025	0.0011
Std	0.0046	0.0008	0.0119	0.0000	0.0014	0.0005

(b)

Figure 5. Performance results of the before (a) and after tuning (b) model performance

Saving the model

```
In [82]: save_model(tuned_model,model_name)

         Transformation Pipeline and Model Successfully Saved

Out[82]: (Pipeline(memory=None,
                   steps=[('dtypes',
                           DataTypes_Auto_infer(categorical_features=[],
                                                display_types=False,
                                                features_todrop=['Date',
                                                                 'Vehicle per fleet',
                                                                 'Fleet Number',
                                                                 'Trip ID'],
                                                id_columns=['Trip ID'],
                                                ml_usecase='regression',
                                                numerical_features=[],
                                                target='Transportation Cost',
                                                time_features=[])),
                          ('imputer',
                           Iterative_Imputer(add_indicator=False,
                                             classifier=LGBMCl...
                          ('dummy', Dummify(target='Transportation Cost')),
                          ('fix_perfect', Remove_100(target='Transportation Cost')),
                          ('clean_names', Clean_Colum_Names()),
                          ('feature_select', 'passthrough'), ('fix_multi', 'passthrough'),
                          ('dfs', 'passthrough'), ('pca', 'passthrough'),
                          ['trained_model',
                           OrthogonalMatchingPursuit(fit_intercept=True,
                                                     n_nonzero_coefs=8, normalize=False,
                                                     precompute='auto', tol=None)]],
                   verbose=False),
          'model_2.pkl')
```

(a)

Serving Model

```
In [83]: import re
         import pandas as pd
         import sys
         from pycaret.regression import *
         if "openpyxl" not in sys.modules:
             !pip install openpyxl > /dev/null 2>&1

         def predict():

             model = load_model(model_name)
             header_index = 0
             test_data = pd.read_excel(test_path,header=header_index)
             test_data.columns = [re.sub(r'[^\w\s]','',col).replace(u"\xa0","").strip().replace("Tire Pressure4","Tire Pressure 4") for cc
             while "Unnamed" in test_data.columns[0]:
                 header_index+=1
                 test_data = pd.read_excel(test_path,header=header_index)
                 test_data.columns = [re.sub(r'[^\w\s]','',col).replace(u"\xa0","").strip().replace("Tire Pressure4","Tire Pressure 4") fc

             predictions = predict_model(model, data = test)
             predictions["Predicted "+target] = predictions["Label"]
             predictions.drop(columns=["Label"],inplace=True)
             predictions.to_excel(pred_path,index=False)
             predictions.head()
             return predictions

         predict()

         Transformation Pipeline and Model Successfully Loaded
```

(b)

Figure 6. Coding of saving and serving the model

5 Results and Analysis

OMP's capacity for high-dimensionality, sparsity promotion, and domain knowledge incorporation render it suitable for pharmaceutical transportation prediction. Its attributes align well with domain complexity and needs, enabling effective modeling and optimization.

As specified in Sections 3, our dataset comprises a small observational sample from 30 calendar days of intra-Marrakeche transport, highlighting OMP's prevention of overfitting distortions. Table variables account for fuel/vehicle influences on costs. Their analysis elucidates factor-output relationships and consequent pricing through comparative metric formulations of model performance, errors and precision.

Overall, selected OMP demonstrates high potential for cost forecasting. It capably captures domain complexity and leverages key variables such as coordinates, congestion, durations, distances, pressures, temperatures, weights, speeds, flows, trips, and emissions. Incorporating these enables reliable predictions aiding optimization of transport operations and informed resource/budget decision-making.

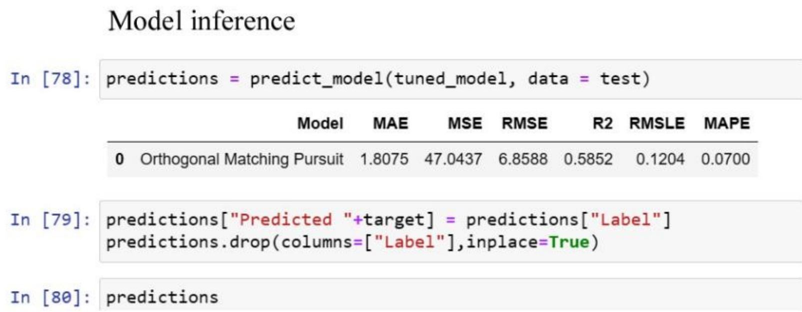
We next progress to prediction's "inference phase" - model deployment post-training/validation on learning and testing datasets respectively. This refers to OMP model execution and scoring application to real situations utilizing field data.

Following model creation, "Serialization" saves it - the process of object conversion allowing storage/transmission for reproduction if needed, utilizing inverse deserialization.

Once model saved, hosting enables cloud/premises functionality integration via API for application unification with AI.

After deployment and server placement, the model is made accessible from the server as presented in Figure 6.

For inference purposes, new data intentionally dissimilar to training datasets was loaded. Thoughtfully and justifiably considered according to inference aims and potential impacts. In real time, minor adjustments addressed errors and inconsistencies emerging within incoming information flows. This validates performance levels potentially reduced versus training and testing phases. Figure 7 below depicts the model's evaluated performance.



Model inference

```
In [78]: predictions = predict_model(tuned_model, data = test)
```

	Model	MAE	MSE	RMSE	R2	RMSLE	MAPE
0	Orthogonal Matching Pursuit	1.8075	47.0437	6.8588	0.5852	0.1204	0.0700

```
In [79]: predictions["Predicted "+target] = predictions["Label"]
         predictions.drop(columns=["Label"],inplace=True)

In [80]: predictions
```

Figure 7. Inference performance of the selected model

6 Discussion

OMP excels through judicious selection of most predictive characteristics, especially within limited datasets. Capably identifying preeminent cost-impacting variables enhances accuracy. Additionally, OMP preserves interpretability, uncovering causal links between constituents and projections, enriching comprehension. Testing involved over 300 observations, with over two-thirds comprising training and 30% retention for post-tuning testing via 10-fold cross-validation of best models.

During inference, new intentionally distinctive data was introduced thoughtfully in view of aims and ramifications. Ongoing minor improvements precisely addressed probable errors and irregularities. 58% performance confirms model resilience across contingencies, encouraging prospective applicability. Validation of real-time tuning underscores intelligent volatility-management, demonstrating theoretical knowledge's practical translation potential for tangible benefits within reality. Adherence to Section 3.3 criteria renders real-time implementation conceivably feasible.

This milestone represents analytical efforts' fulfillment, enabling theoretical understanding's manifestation as meaningful utility aiding real-world scenarios. Despite advantages, OMP possesses limitations. Susceptibility to outliers potentially impacts result quality. Additionally, expertise prerequisites on ideal constituent dimensionality constrain accommodation of intricacy. Identification of non-linear or interdependent associations remains difficult. Handling lacking information likewise warrants consideration.

7 Conclusions

This discussion explored algorithm selection for pharmaceutical transport cost prediction via a 300-observation small dataset. The importance of diverse algorithm comparisons for suitability determination was underscored. Evaluation metrics including MAE, MSE, RMSE, R2, RMLSE, MAPE, and CT were pivotal performance indicators. Metrics assessed various algorithms, with Orthogonal Matching Pursuit emerging most suitable due to high-dimensionality handling, sparsity promotion, and feature relevance permitting accuracy.

Through rigorous evaluation and contrast, OMP surpassed others in accuracy, interpretability and efficiency. Comprehensive assessment of algorithmic functioning considered prediction exactness, error examination and computation time. OMP accurately predicts despite limited data by selecting few, information-rich characteristics clarifying cost-variable relationships. Complexation capturing may falter by selecting highly correlated features. This model demonstrates high-level cost forecasting.

Results underscore AI value for planning/resource optimization and performance management strengthening. Precision and transparency empower strategic, informed decision-making. Successes transcend pharmaceuticals, revolutionizing supply chains through more exact cost predictions and resource management via agribusiness expansion. Similarly, dangerous goods transport application complexity handling promises. Larger, diverse datasets could strengthen generalization/reliability through finer, complex pattern capture.

Data Availability

The data used to support the research findings are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflict of interest.

References

[1] M. Rasol, J. C. Pais, V. Pérez-Gracia, M. Solla, F. M. Fernandes, S. Fontul, D. Ayala-Cabrera, F. Schmidt, and H. Assadollahi, "GPR monitoring for road transport infrastructure: A systematic review and machine learning insights," *Constr. Build. Mater.*, vol. 324, p. 126686, 2022. https://doi.org/10.1016/j.conbuildmat.2022.126686

[2] Ü. Ağbulut, "Forecasting of transportation-related energy demand and CO_2 emissions in Turkey with different machine learning algorithms," *Sustain. Prod. Consum.*, vol. 29, pp. 141–157, 2022. https://doi.org/10.1016/j.spc.2021.10.001

[3] K. L. M. Ang, J. K. P. Seng, E. Ngharamike, and G. K. Ijemaru, "Emerging technologies for smart cities'transportation: Geo-information, data analytics and machine learning approaches," *ISPRS Int. J. Geo-Inf.*, vol. 11, no. 2, p. 85, 2022. https://doi.org/10.3390/ijgi11020085

[4] M. Woschank, E. Rauch, and H. Zsifkovits, "A review of further directions for artificial intelligence, machine learning, and deep learning in smart logistics," *Sustainability*, vol. 12, no. 9, p. 3760, 2010.

[5] M. Huff-Rousselle and F. Burnett, "Cost containment through pharmaceutical procurement: A Caribbean case study," *Int. J. Health. Plann. Manag.*, vol. 11, no. 2, pp. 135–157, 1996.

[6] K. Tadlaou, A. Chafi, and A. Ennadi, "Système d'approvisionnement pharmaceutique au Maroc: Opportunités et défaillances," in *Congrès International de Génie Industriel et Management des Systèmes (CIGIMS), Fès, Maroc*, 2015.

[7] H. Chrifi, A. Echchatbi, and A. Cherkaoui, "Modélisation de la chaîne logistique pharmaceutique marocaine Vers l'intégration du facteur risqué," in *Xème Conférence Internationale: Conception et Production Intégrées, Tanger, Maroc*, 2016.

[8] M. B. El Azizi, "Modélisation multi-agents de la coopération au sein des chaînes logistiques à deux échelons: Application à la distribution de produits pharmaceutiques au Maroc," 2014.

[9] P. P. S. De Carvalho, R. D. A. Kalid, and J. L. Moya Rodriguez, "Evaluation of the city logistics performance through structural equations model," *IEEE Access*, vol. 7, pp. 121 081–121 094, 2019. https://doi.org/10.1109/access.2019.2934647

[10] S. Forkert and C. Eichhorn, "Innovative approaches in city logistics: Inner-city night delivery," *Niches*, 2008.

[11] C. Liu and L. Ke, "Cloud assisted Internet of Things intelligent transportation system and the traffic control system in the smart city," *J. Control Decis.*, vol. 10, no. 2, pp. 174–187, 2022. https://doi.org/10.1080/23307706.2021.2024460

[12] A. Premkumar and C. Srimathi, "Application of blockchain and IoT towards pharmaceutical industry," in *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India*, 2020. https://doi.org/10.1109/icaccs48705.2020.9074264

[13] F. Elbichri, "Supply chain finance issues: An exploratory survey on Moroccan companies," *Int. J. Account. Fin. Audit. Manage. Econ.*, vol. 4, no. 2, pp. 518–532, 2023. https://doi.org/10.5281/zenodo.7829481