Pegem Journal of Education and Instruction, Vol. 15, No. 3, 2025 (pp. 330-360)

RESEARCH ARTICLE WWW.PEGEGOG.NET

Design and Training a Routing Model for Predicting E-Commerce Logistics Distribution Systems in Transportation and Delivery of Educational Goods for Students Using Existing Algorithms, Big Data Analysis, and Machine Learning Methods

*

Shiva Khalili Gheidari

Master of Science in Information Technology (IT) Engineering, Mazandaran University of Science and Technology, Mazandaran, Iran

(Email: **Shivakhaliligheidari@gmail.com)**

Abstract

The problem of vehicle routing for delivering goods in e-commerce is one of the most important issues in the world's industries, which has received much attention today due to its real-world applications in industrial issues. Trying to increase the efficiency of the transportation system is a big step towards making the activities of this sector more economical. In this regard, reducing the distance traveled by vehicles in the transportation system, which has an impact on various economic areas (financial cost, energy consumption, time, etc.), including the delivery of goods in the educational system, is one of the issues raised in this area. In this regard, this study designed and trained a routing model to predict ecommerce logistics distribution systems in the transportation and delivery of educational goods for students using existing algorithms, big data analysis, and machine learning methods. The dataset used to build the model includes 12,399 observations from 12 variables including customer ID, warehouse block, mode of transportation, customer care calls, customer rating, product cost, previous purchases, product importance, male and female, discount offered on that specific product, weight in grams, and on-time delivery of the product to the customer, all related to Snap Company. In this study, 8 classification algorithms including rf or random forest algorithms, knn algorithm, ADB logistic distribution, and XGB algorithm were used to create machine models and predict on-time delivery of goods. Finally, the classification algorithms were compared and the best models were identified in order. This project showed that using machine learning techniques, it was predicted whether the company's cargoes reached their destinations on time or not, and to what extent this work will help optimize, improve, and develop the company.

Keywords: Logistics distribution, e-commerce, big data, training system, machine learning technique]

1-Introduction:

The new developments of the 21st century and the globalization of the market, followed by the emergence of new ways of business and marketing, which have been influenced by the rapid and comprehensive development of science and technology, and many economists, specialists and concerned futures have believed that In recent years, a revolution similar to the industrial revolution has taken

Corresponding Author e-mail: Shivakhaliligheidari@gmail.com

How to cite this article: Shiva Khalili Gheidari. Design and Training a Routing Model for Predicting E-Commerce Logistics Distribution Systems in Transportation and Delivery of Educational Goods for Students Using Existing Algorithms, Big Data Analysis, and Machine Learning Methods. Pegem Journal of Education and Instruction, Vol. 15, No. 5, 2025, 136-166

Source of support: Nil **Conflicts of Interest:** None. **DOI:** 10.48047/pegegog.15.05.10

Received: 12.04.2025

Accepted: 22.06.2025 **Published:** 01.07.2025

place. This revolution has brought the world into the "Information Age", has made many economic, social and cultural aspects of human life undergo a deep transformation and by using new technologies, it has reached a new and unimaginable level of dynamism and changeability. One of the dimensions of the information age is the profound changes that have occurred in the economic relations between individuals, companies and governments, so that a new phenomenon called business and electronic commerce has been introduced and many concepts, principles and It has transformed economic theories, to the extent that the word "Internet-based economy" is gradually replacing the word "traditional economy" in the Internet market, compared to the traditional market, there is this challenge that the business organization has with its competitors It is faced more and customer loyalty is less. Therefore, attracting, gaining trust, satisfaction and maintaining customers is an important and difficult issue for commercial organizations (2).

Due to the speed of efficiency, cost reduction and exploitation of fleeting opportunities, electronic commerce has opened a new arena in the competition, as far as it is said, lagging direct marketing and after-sales services (3). behind this evolution will have no result other than being isolated in the world economy arena. 1).

E-commerce includes various activities such as electronic exchange of goods and services, immediate delivery of digital claims, electronic transfer of funds, electronic exchange of shares, electronic bill of lading, business plans,

In the world of ecommerce, companies have the opportunity to communicate more with both suppliers and customers, which on the one hand leads to increased sales and on the other hand leads to access to more resources with reasonable prices and costs. becomes less (4).

Virtual stores have many advantages, including convenience, competitive prices, wide selection, and access to more information. On the other hand, in the case of shopping from home, the relationship between people in the society will be the lowest (5). In recent years, a large number of virtual stores have been created through moving pages and creating virtual realities and online communication (ON LINE CHATTING).) increase their relationships with people (6).

In the world of e-commerce, there are many factors that influence the decision and intention to buy through websites. The virtuality of the internet network, a little trust of the customer in online shopping, the quality of the site and its design model, especially in the way of presenting the product and relevant information, lead to the reluctance of customers to make electronic purchases. On the other hand, factors of people's characteristics also affect the decision and conduct of electronic shopping, such as people's loyalty to a brand name, attitudes, habits, people's

understanding of the environment and the opportunities that are presented in it, people's willingness to enter The world of new business and the use of technology and innovations. In the virtual store, you should have the answers to the following questions to be able to perform your activities well:

- What do we want to sell?

Who are our customers?

- Why does the customer want to buy from our online store? (7)

Online customers pay special attention to service performance indicators: time-related indicators, i.e. being on time and receiving products within a specified time frame - and delivery speed, i.e. minimizing the time interval between the customer's order and delivery (8-10).).

Distribution has always been one of the problems of organizations (11). Distribution operation occupied an important and meaningful position in product marketing when production increased and the distance between producer and consumer increased. In this way, the producer had to deliver his products to the consumer at long distances and through people called agents, wholesalers or retailers (12). Last-mile logistics is a phenomenon that is constantly growing, mainly due to ongoing urbanization and changes in consumer habits, with the strong growth of online retail and the resulting increase in online grocery shopping and ecommerce activities. The most recent statistics show that in the European Union the share of online shoppers was equal to 64% of all people aged 16 to 74 [1]. This growing pressure of freight traffic in urban areas brings with it a series of side effects that undermine the sustainability and liveability of our cities. This increases congestion and greenhouse gas emissions in urban areas due to the additional traffic generated by vehicles for deliveries that often have overlapping routes (25% CO2 and 30-50% PM and NOx). Furthermore, it reduces road safety due to the presence of heavy vehicles (13). The restrictive measures associated with the COVID-19 pandemic further accelerated the rise of online shopping, while being a “rare catalyst” for logistics innovation (14). However, despite recent efforts to improve logistics sustainability, issues related to the entire process are still debated. New technologies can play a key role in improving the efficiency of last-mile deliveries in urban areas. Beyond the wellknown smart technologies associated with online shopping, the entire logistics supply chain can also benefit from innovative processes and be a suitable field for automation. In particular, artificial intelligence (AI) and machine learning (ML) approaches have become a hot topic in the literature and practice, and are increasingly related to traditional vehicle routing optimization (VRO) models found in the literature. Their algorithms are actually able to provide accurate predictions – based on historical data – on common productivity problems in supply chains, such as demand forecasting, routing, and tracking, while being able to detect anomalies along the way (15). These algorithms are constantly evolving, and thanks to recent disruptions, the last few years have been very important for this area of research and development (16).

In this scenario, logistics management, and especially the management of the delivery process, acquires a fundamental role (13). In fact, the complexities of the physical distribution of products should not be underestimated, and among the various aspects that play a role in the success of an ecommerce initiative. have, logistics plays an essential role (17).

Logistics consists of planning, implementation and control in order to deliver raw materials to production centers until the product reaches the customer (18).

According to the definition provided by the American Logistics Association, logistics management is: the process of planning, implementing and controlling the flow and storage of raw materials, the inventory of semi-finished parts and final goods and their related information from the point of origin to the point of consumption for supply. Customer needs and satisfaction (19).

Success in the commercial environment requires the management of a complex communication system, which actually determines the distribution model between logistics agents, which aims to deliver raw materials to production centers until they reach the customer (20).

Today, coordinated and integrated decisions among each member of the supply chain are very vital. Two of these important decisions are the scheduling of production and distribution of customer orders, which takes place at the level of operational planning and leads to the reduction of operational costs and timely delivery of orders, and leads to customer satisfaction. In the conventional and previous point of view, supply chain management includes the guidance of all members of the supply chain in a unified and coordinated manner with the aim of improving performance, to improve productivity and profit, and supply chain managers seek to deliver goods and services faster, reduce costs and increase quality. were (21).

Electronic supply chain management is the use of information technology to facilitate communication between organizations and customers within a chain (22). Supply chain management is the process of integrating the activities of the supply chain and sharpening the information flows related to it, through the improvement and coordination of the activities in the supply chain of production and supply of the product. Supply chain management has three major processes, which are: 1- Information management

2-Logistics management

3- Relationship management

The scope of logistics is not only the flow of materials and goods, but also the focus of supply chain activities, where relationships and information are supporting tools for improving activities (19).

Business consumers in the digital age have increasingly increased (23). Technologies such as artificial intelligence (AI) and big data analytics provide unprecedented opportunities for companies to exploit data assets. have created their own for business-to-business (B2B) market initiatives (24).

Companies such as Google, Amazon, Facebook, and Apple have all made significant efforts in the field of industrial marketing through the collection and use of big data. All this emphasizes the importance of big data as an important factor in global marketing operations (26).

Data also plays a key role in various decisions regarding business supply chains and logistics operations, which are closely related to the field of industrial marketing. Supply chain management is concerned with creating and maintaining links between different entities with specific responsibilities, from purchasing raw materials to finished product interactions.

Logistics management ensures that related work support methods are in place, such as traffic management, warehouse management, inventory management, packaging and order tracking. By using a wide and diverse range of data in logistics and supply chain management, companies can understand the needs and preferences of their customers. Ecommerce

giants such as Amazon, Flipkart, and Snapdeal collect and explore customer data, orders, inventory, and other information (26). The success of e-commerce companies today largely depends on how efficiently they collect, store and use data. The emergence of the era of big data has further strengthened the relationship between logistics distribution and e-commerce, and this provides new opportunities such as the expansion of organizational information. Since the logistics ability of an e-commerce company has become an important indicator for its competitiveness. is, mode selection The integration of decision-making in the supply chain examines the dependence between the location of facilities, the allocation of suppliers and customers to facilities, the structure of the transportation system and their routing (27).

In production environments, not paying attention to the delivery or due date of a job may cause the loss of a customer or the cost of compensating for a delay in the production and assembly line (28). In an integrated system, the ability to control the delivery date is an important factor in improving the performance of a system, and increasing the delivery date of an order leads to costs (29). The purpose of supply chain scheduling is to optimize short-term and medium-term decisions according to the balance of economic goals such as minimizing costs or maximizing profits and intangible goals such as customer satisfaction or customer service level (30). One of the common methods in the industry for sending goods is delivery in the form of batch shipment, with the aim of reducing transportation and distribution costs (31).

The most important drawback of the single shipping method is the delay in sending the goods, because in order for that batch to be sent, all the orders of a batch must be completed. In order to adopt a suitable strategy, a balance should be made between the shipping cost and the delay of orders (29).

The vehicle routing problem with a heterogeneous fleet is another form of the routing problem that Golden et al. raised and in which the vehicles have different fixed and variable costs (30).

2- The basics and background of the research

1-2-logistics management and distribution

This word has Greek roots and is used in military cases to move weapons, ammunition and food rations when moving from the main place to the front line. In the Greek, Roman and Eastern Roman Empire, there were soldiers named LOGISTIKAS, who were responsible for financial matters and distribution of necessities. In the Oxford dictionary, logistics is defined as: a part of military science that is responsible for the preparation and delivery of supplies and the movement of materials, people, and equipment. Logistics means the integration of two or more activities with the aim of planning, implementing, efficiently controlling the flow of materials and products from the main location to the point of consumption. Logistics includes information integration, transportation, inventory, warehouse, goods movement and packaging. Due to its nature, logistics is often synonymous with distribution activity, both for physical product distribution, supply chain management, pipeline management, and supply and transportation (31).

Logistics distribution directly affects the company's distribution quality and costs, as well as the coordination of its supply chain. There are three main modes of logistics distribution for an e-commerce company, which include self-made logistics, third-party logistics, and joint distribution mode (combined).) in order to meet the requirements of e-commerce development and improve customer satisfaction, it is necessary

for e-commerce companies to fully understand and examine the advantages and disadvantages of various logistics distribution methods and choose the most appropriate ones, for example, Meituan, China's largest online food ordering platform, developed a scientific distribution system with professional logistics and population distribution to avoid the cost pressure of self-made logistics. Meituan believes that the reasonable choice of the distribution mode is an important way to effectively save its costs (24).

Transferring, moving, processing and accessing logistics information to integrate shipping, ordering and manufacturing processes, order changes, production scheduling, logistics plans and warehousing operations; Data exchange and processing among partners, such as exchange and processing of technical information, orders, etc., collection and processing of information to analyze the process of sourcing and evaluation, selection and development of suppliers; Collecting and processing supply and demand information, etc. to predict market trends and future conditions of supply and demand; Creating and improving relationships between partners (31).

2-2- Electronic commerce

Studying the electronic business system is essential in today's turbulent and complex world, because the key trends in traditional businesses have changed a lot and are moving towards fast and global access, speed in the entire supply chain and conducting business electronically. the river Electronic business is a broad concept that includes all aspects of using information technology in business and does not only include buying and selling. Rather, it includes creating coherence and integration in all business processes and communications inside and outside the organization.

In other words, electronic business is a general term used to describe work processes used in virtual or electronic environments such as the World Wide Web.

According to Zvas, "e-business is the exchange of information through an electronic network, at any stage of the supply chain, whether within a company or between companies, between the company and the customer, or between the private and public sectors of the economy. And whether it is accompanied by payment of money or no payment is made. »

Turban (2002) also considers electronic business "something beyond electronic business in which, apart from the company's business activities, many other interorganizational relationships and between the customer and the organization are also electronic in nature. take »

E-business is the acquisition of customers and merchants for commercial exchanges through self-working of transactions, exchanges, communications and interactions through communication and computer technologies for economic goals that include interorganizational systems, Such as: telephone, internet, e-mail, or internal computer networks to support online business transactions. So it can be said that the goal of this type of new business is to automate business transactions and work flow.

The general definition provided by IBM company of the term e-commerce is: doing ecommerce using electronic media such as the Internet.

The intended e-commerce processes in IBM's definition include research and development, marketing, production and procurement (support) internally and externally, ecommerce buyer side exchanges with suppliers and seller side exchanges with customers. These processes are the key business processes.

» A safe, flexible and integrated approach to achieve value in different e-commerce by creating a combination of systems and processes that are based on the core activities of e-commerce while maintaining simplicity and using Internet technology. (33)

A step beyond e-businesses is a concept called dynamic e-businesses. Dynamic e-business is about how organizations can integrate and integrate systems in the three dimensions of intranet, extranet and internet in a dynamic state. Dynamic e-business allows them to modify existing systems more quickly and easily when business processes require changes, and focuses on the integration and substructural complexity of business-tobusiness relationships. By leveraging and influencing the benefits of Internet standards and public substructures to create optimal efficiencies for computer disciplines inside and outside the organization, electronic commerce means conducting business transactions through the network, especially the Internet (32). E-commerce has created important opportunities for companies through which companies can expand their market share and enter the international market. This can be implemented with the lowest cost and minimum investment (33).

E-commerce goes beyond having an internet site and includes placing orders, conducting job tenders, recruiting and exchanging information using information technologies. In e-commerce, money can be paid in both traditional and electronic ways. Many electronic merchants offer different payment systems to their customers (2).

Figure 1- Definition of e-commerce and e-business

Department of Trade and Industry of England (2003) considers electronic commerce as something more than a technology. This department believes that electronic commerce is related to the attitude of commercial companies to customers. Companies in ecommerce use the Internet and other communication and information tools to improve the performance of e-commerce and achieve success.

Turban (2002) also considers e-commerce as a more general meaning than e-commerce, which includes electronic buying and selling, customer services, business partnerships and supply chain management.

2-2- Integration of electronic commerce

In the last part, we explained the different models of the proposed framework based on the customer's wishes. In the following, first of all, the field of business and activity and

segmentation of customers should be determined, and then by combining the fields of activity and business models, an integrated framework of the company's business in all fields should be prepared.

1. Types of customers and fields of activity

The most famous division of the field of activities and types of customers is the comprehensive division according to the two factors of the company (B) and the customer (C) in the form of 4 B2B, B2C, C2B, C2C. Here, instead of using this famous framework, we use a more precise and detailed framework with the involvement of two other factors, employee and agent. It should be noted that "Agent" here means "intelligent agents that are used by the customer to automatically carry out the pricing and purchasing process."

	B	**C**	**A**	**E**
Business	B:B	B:C	B:A	B:E
Consumer	C:B	C:C	C:A	C:E
Agent	A:B	A:C	A:A	A:E
Employee	E:B	E:C	E:A	E:E

For example, B2E refers to the supply of services and products by the company to its employees, such as providing online insurance services. The reason for using this framework is to use the maximum capacity of organizations and individuals in providing service to specific customer segments, instead of the general segments specified in the famous company/customer framework. (39)

2. Integrated framework

Since each of the fields of activity and types of customers are supported by various types of values and business models, it becomes important to provide a multi-layered integrated model for the company's business method in different fields. The figure below presents an integrated framework of a company's business in different areas. In this figure, each square represents a business area and a section of customers. For each type of customer, there are different combinations of value and earning methods, and these different methods are marked on the sides of the square. It should be noted that the company may not have any activity in a number of business areas. The way of providing services in each field and the ability to carry out commercial processes for different customers are issues that companies should address and act intelligently about. (39)

2-1-factors affecting the ideal model of electronic commerce

- Delivery on time

The core of just-in-time production and delivery philosophy is to prevent wasted energy, waste of time, wasteful activity, excessive and useless production, prevent readjustments, defects, prevent product returns and finally adjust time.

Just-in-time production and delivery for lowcost production is the production and delivery of quality essential parts in the required quantity, at the right time and place, with the minimum use of facilities, equipment, materials and manpower.

Just-in-time production and delivery strengthen companies' competitive advantages, increase cash flow, and reduce inventory and working capital. In addition, just-in-time production and delivery can facilitate changes in designs to better respond to changing market conditions. The product must be produced in the fastest possible time and at the lowest cost, but in all these cases, the quality desired by the customer and fast delivery must be taken into consideration.

Just-in-time manufacturing and delivery essentially eliminates anything that isn't directly related to the value of the product. Apart from this, defective and incompatible products with the desired coordinates are also removed.

This waste includes everything that is used to fix the defect and even the wasted time associated with setting up the press. Just-intime production and delivery makes the printer complete his work in minimum time. (17)

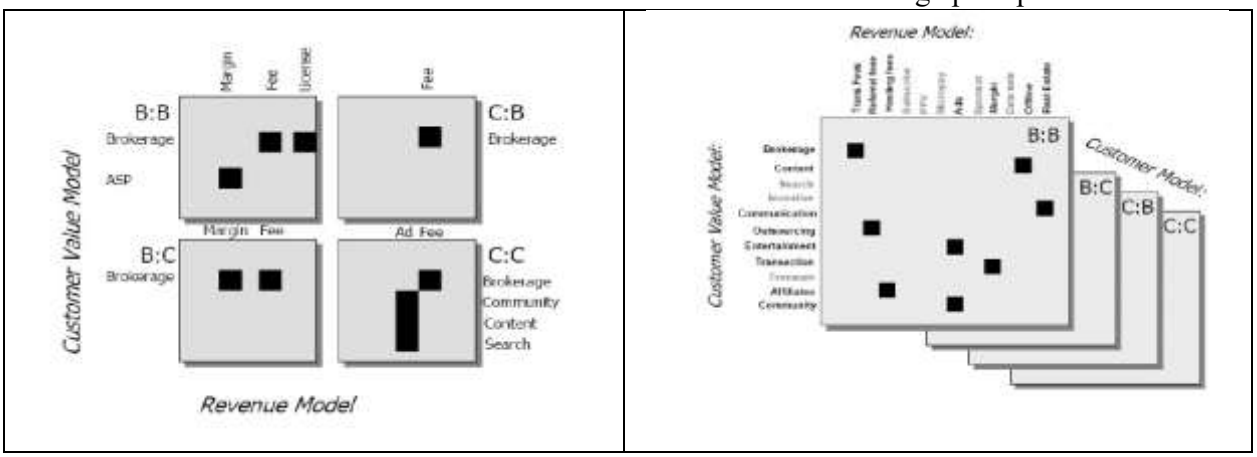


Figure 2: Integrated framework of business models for eBay company

- On-time production and delivery and comprehensive quality management

Just-in-time production and delivery can be applied both after the comprehensive quality management stage and in parallel with it.

But just-in-time production and delivery cannot be implemented without a successful comprehensive quality management program, because just-in-time production and delivery forces the manufacturer to avoid producing substandard products.

Most of the printers, packaging and converting industries are in an extreme situation in terms of equipment and furniture, a high volume of paper is an added cause, while most of them have poor solutions for their problems.

The entire production time, that is, from the receipt of the prescription to the delivery of the final product, has many considerations in different stages. Pre-printing processes, receiving raw materials and goods, and machine preparation time are all part of these steps. In fact, the total production time includes the times related to the production of waste or the times related to the production of incomplete products.

If color separation and other pre-press services are to be provided from commercial workshops, a large amount of coordination and planning is needed, in this case, a situation like creating logical relationships between departments of a workshop is needed. (41) **- Reduction of warehouse inventory**

Inventory always goes up when we don't control or plan for orders on the assembly line or on our printing press. Adding inventory as a solution to prevent future problems is one of the most expensive solutions that increases working capital, and one of the losses always occurs with inventory that suddenly loses its market and must be scrapped. (40)

Not only the amount of stock in the warehouses should be reduced, but also the production processes should be simplified so that the time of adjusting and preparing the machines is reduced and the time of fixing defects and implementing immediate solutions is accelerated.

A visitor to the largest printing shop in Tokyo reported that the owner always allowed only 24 hours for paper delivery, and actually only stocked paper for 24 hours. And while this storage time is 60 to 90 times longer for a printer engaged in publishing in America.

There is no doubt that manufacturers must have a long-term commitment to suppliers and rely on them to achieve their goals. This type of relationship is unfamiliar to many business people in America and it takes a long time to establish it. (18)

- Resistance to production and delivery just in time

When management shies away from commitment, it is the skeptics of management who actively or passively begin to resist. For this reason, just as there are different experiences in America, in England and Europe, it can be seen from success to failure. In any case, if you are competing with a Japanese company, you are actually competing with the phenomenon of just-intime production and delivery, and although this phenomenon has probably failed in some cases in Japan, it has undoubtedly seen its greatest growth and success. has also experienced this in Japan.

However, just-in-time production and delivery, like total quality management, is an unfamiliar concept, and people, as well as managers, resist unfamiliar things and do not like to change the procedures they are familiar with. This is why, for example, in America, it is very difficult for some people to get rid of unnecessary things in the warehouse, such as empty boxes or paper rolls. (44)

- The effect of just-in-time production and delivery on packaging printing

The long time required for the preparation of letterpress machines and the high cost and time of production of gravure cylinders and the high waste of web offset machines cannot in themselves determine the processes that should be used, but it is absolutely necessary to take into account the considerations related to them.

The process of "coloring" the offset (that is, reaching the equilibrium point of ink and water) requires time and wastes paper. In fact, zinc scanning devices as well as densitometers were produced for this to minimize the time of color adjustment and actually reduce the time of composite adjustment, but flexo and gravure machines have an inherent advantage in this regard.

The long time of sampling and getting the customer satisfied with the sample is one of the places where you can reduce wasted time and unnecessary costs.

Application of total quality management principles can reduce all these unnecessary things. In addition, using the phenomenon of just-in-time production and delivery in the case of customers requires the fact that all problems are clearly identified and a plan to solve them.

In order to achieve complete satisfaction with products, cooperation between suppliers and customers is absolutely necessary, and although the application and use of comprehensive quality management and just-in-time production and delivery, most of the time leads to a reduction in the final price of the products, but what ultimately The determining factor of profitability and expansion of market share is nothing but the strict implementation of product and customer requirements and demands.

The concept of just-in-time production and delivery is that less time is spent, order sizes are smaller, and quality is better and more desirable. In fact, in order to achieve these goals, the triangle of quality, cost and service should always be considered: better technology, better management methods and better personnel training.

Printers, those involved in the converting and packaging printing industries, like other manufacturers, are reviewing the events they use, they should not be looking to buy cheaper ink or paper and zinc, but should be looking to buy products that allow them to meet the demands of customers. Gives at the lowest cost. The goal of the producer is not cheaper raw materials or lower wages, but the maximum profit. (35)

2-2-Machine learning techniques

In the science of machine learning (Machine Learning), the issue of designing machines that learn by using the examples given to them and their own experiences is addressed. In fact, in this science, an attempt is made to design a machine in such a way that it can learn and act without explicitly planning and dictating each and every action by using algorithms. In machine learning, instead of programming everything, data is fed to a generic algorithm, and it is this algorithm that builds its own logic based on the data it is fed. Machine learning has various methods, including supervised learning, unsupervised learning, and reinforcement learning. Algorithms used in machine learning are among these three categories. Machine learning is an application of artificial intelligence (AI) that allows systems to learn and improve automatically from experience without explicit programming. Machine learning is about developing computer programs that can access data and use it to learn. The learning process begins with observations or data such as examples, direct experience, or instructions to look for patterns in the data and make better future decisions based on the examples we provide (33).

The main goal is for computers to learn automatically without human intervention or help and adjust actions accordingly.

- **types of machine learning**

There are three types of machine learning algorithms:

• Supervised learning

• Unsupervised learning

• Reinforcement learning

- Supervised learning

Most machine learning methods use supervised learning. In supervised machine learning, the system tries to learn from the previous examples it is provided with. In other words, in this type of learning, the system tries to learn the patterns based on the examples given to it. (34)

Mathematically speaking, when the input variable (X) and the output variable (Y) are available and based on them, an algorithm can be used to obtain an input-output mapping function, in fact supervised learning is done. The mapping function is shown as (Y = f(X).

In machine learning, the algorithm is given data set(s) and the machine forms its own logic based on that data set(s). This data set has rows and columns. Rows (also referred to as records and data samples) represent data samples. For example, if the data set is about soccer games (weather conditions), one row contains information about a specific game. Columns (which are also referred to as features, attributes, characteristics) are actually features that describe each data sample (36).

Supervised machine learning problems can be divided into two categories: "classification" and "regression".

Categorization: A problem is considered to be categorized when the output variable is a category or group. For an example of this, we can refer to the belonging of a sample to the "black" or "white" categories and an email to the "spam" or "non-spam" categories.

Regression: A problem is regression when the output variable is a real value such as "height". In fact, it is used in classification with discrete variables and in regression with continuous variables.

- Unsupervised learning

In unsupervised learning, the algorithm should search for interesting structures in the data by itself. Mathematically speaking, unsupervised learning is when there are only input variables (X) in the data set and no output data variables. This type of learning is called unsupervised because, unlike supervised learning, there is no given correct answer and the machine itself must search for the answer (58), in other words, when the algorithm uses a data set to work with that lacks Labeled data (output variables) uses another mechanism for learning and decision making. This type of learning is called unsupervised. Unsupervised learning can be divided into clustering and association problems. (41)

Association rules: A learning problem is association rules when the goal is to discover rules that describe a large part of the data. For Example:

example, "a person who buys product A has a tendency to buy product B."

Clustering: A problem is considered clustering when there is an intention to discover inherent groups (data that are inherently included in a specific group) in the data. For example, grouping customers based on their buying behavior.

- Reinforcement learning

A computer program that interacts with a dynamic environment must achieve a specific goal (such as playing a game with an opponent or driving a car). This program provides feedback in the form of rewards and punishments and directs its problem space accordingly. Using reinforcement learning, the machine learns to make specific decisions in an environment that is constantly subject to trial and error.

Reinforcement learning is one of the trends in machine learning that is inspired by behavioral psychology. This method focuses on the behaviors that the machine should perform to maximize its reward. Due to its scope, this issue is investigated in various fields such as game theory, control theory, operations research, information theory, multi-agent system, crowd intelligence, statistics, genetic algorithm, optimization based on simulation. Reinforcement learning has two major differences with supervised learning, firstly, it does not use correct pairs of input and output, and inefficient behaviors are not corrected from the outside, and secondly, there is a strong focus on live performance, which requires finding a suitable balance between It has the discovery of new things and the exploitation of accumulated knowledge. Augmented learning in manufacturing, selfdriving cars, business management, computer systems, machine vision, education, energy, finance, gaming, healthcare, natural language processing (NLP), robotics, "science, engineering and art" and transportation is used. (17)

B: SVM

The original SVM algorithm was invented by Vladimir Vapnik in 1963 and extended to the nonlinear case by Vapnik and Corinna Cortes in 1995.

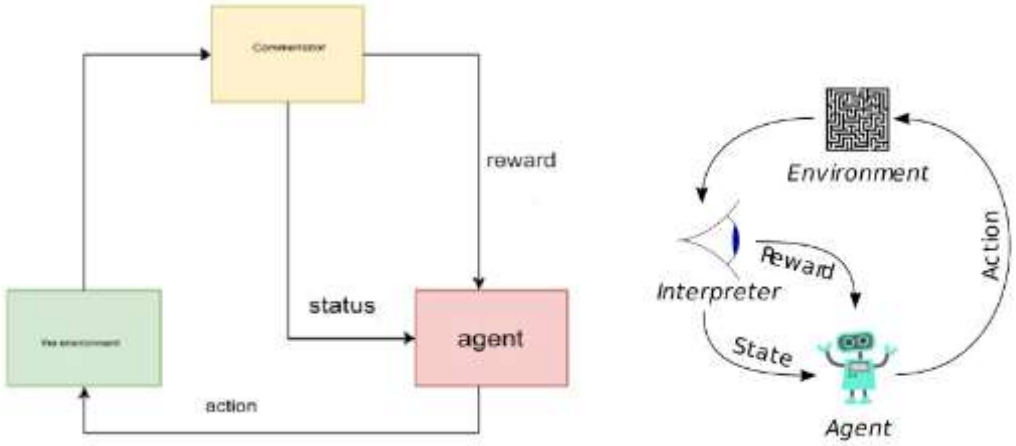


Figure 3- The functional system of reinforcement learning (23)

The SVM algorithm is classified as pattern recognition algorithms. The SVM algorithm can be used wherever there is a need to recognize patterns or classify objects into specific classes.

Summary of practical use of SVM

We prepare the pattern matrix. We select the kernel function to use. We select the parameter of the kernel function and the value of C. To calculate iα values, we run the training algorithm using QP solvers. New data can be classified using iα values and support vectors (48).

Advantages and disadvantages of SVM

The training is relatively simple. Unlike neural networks, it does not get stuck in local maxima. It works pretty well for highdimensional data. The trade-off between classifier complexity and error rate is clearly controlled. It requires a good kernel function and C parameter selection.

C: Linear support vector machine

We have the test data set D including n members (points) which is defined as follows:

D= ⟦{(x_i,y_i)|x_i∈R^p,y_i∈{-1,1}}⟧ _(i=1)^n (1)

where the value of y is equal to 1 or -1 and each is a real p-dimensional vector. The goal is to find the separating hyperplane with the largest distance from the marginal points that separates the points with iy = 1 from the points with iy = -1. Every hyperplane can be written as a set of points x that satisfy the following condition:

w. x - b = 0 where . is the inner multiplication sign and w is the normal vector, which is perpendicular to the hyperplane. We want to choose w and b so that the maximum distance between the parallel superplanes separating the data is created. These hyperpages are described using the following relations.

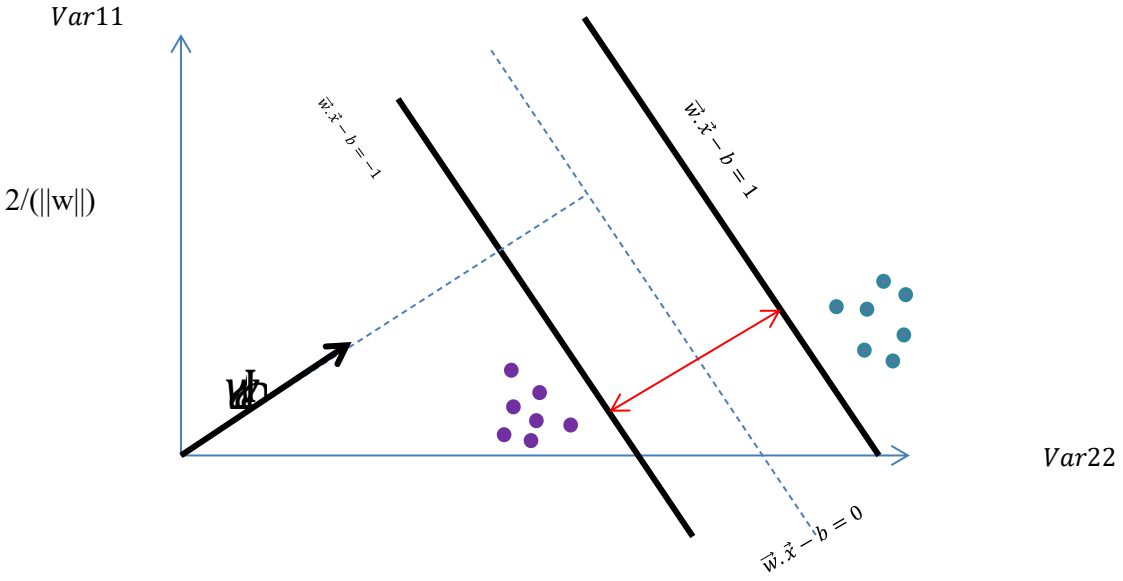
w. x - b = +1, w. x - b = -1 (2)

If the training data is linearly separable, we can consider two hyperplanes on the edge of the points, such that they have no points in common, and then try to maximize their distance. Using geometry, the distance between these two planes is 2/(||w||). Therefore, we must ||w|| minimize In order to prevent points from entering the margin, we add the following conditions: for each :iw. xi- b ≥ 1 ∀ xi with yi = 1 (13)

w. xi- b ≤ 1 ∀ xi with yi = -1 (3) These

conditions can be written as follows: yi(w.

xi + b) ≥ 1 ; i=1,...,N (4)



By putting these two together, an optimization problem is obtained: Minimize |(|w|)| (with w and b) provided yi(w. xi + b) ≥ 1 for all i=1,…,N.

2-4-Scientific record of research

• A study was conducted by Kangning Zhenga and colleagues in 2023 on the topic of ecommerce logistics distribution mode in the field of big data, case analysis of JD Com.

This article analyzes the existing distribution modes adopted by Chinese trading companies. Analytic Hierarchy Process (AHP) method and entropy value are used to investigate the distribution selection mode of e-commerce company and order priority method with similarity to ideal solution (TOPSIS) method is used to verify the model. The research results indicated that e-commerce logistics distribution specialists have strong managerial insights.

• A study was conducted by Govindan Kannan and colleagues in 2020 on the topic of big data analysis and its application in logistics and supply chain management. In this research, big data analysis and applications for logistics and supply chain management were discussed by examining new methods, practices and opportunities. Also, various opportunities to improve big data analysis and applications were discussed. For logistics and supply chain management, such as: exploring technologybased tracking strategies, financial performance relationships with data-driven supply chains, and supply chain capability implementation and maturity issues with big data were presented and analyzed.

• A study was conducted by Pei-Ju Wu et al. in 2018 on the topic of unstructured big data analysis to retrieve e-commerce logistics knowledge. In this research, e-commerce logistics business models were examined from unstructured big data. . In particular, this work develops a hybrid content analytic model to examine the essential knowledge of ecommerce logistics. The empirical results of the proposed model integrate the theories of Resource Dependence Theory (RDT) and Innovation Diffusion Theory (IDT) to generate logistics strategies.

A study was conducted by Taherdoost et al. in 2019 with the title of analysis of supplier selection criteria and methods. In this research, the main goal of the supplier selection process is to reduce the purchase risk, maximize the overall value for the buyer, and develop proximity. And long-term relationships between buyers and suppliers were stated. This article provides an overview of research on supply chain management, supplier selection criteria, and supplier selection evaluation methods (multi-criteria decision making). The results show that the use of structured decision-making technique is critical, especially in complex situations involving quantitative and qualitative criteria.

• A study was conducted by Hasanpour et al. in 2018 on the subject of multi-objective design of agile and valuable supply chain with the approach of multi-objective genetic algorithm. In this research, a study based on linear integer modeling has been conducted in the field of supply chain network design to address this research gap. The intended supply chain is presented at three levels of producers, distributors and customers in the form of multi-purpose, multi-product and multi-round. Objective functions include maximization of agility and pivot values. Games software is used to solve the mathematical model. Then the multi-objective genetic algorithm was presented with non-dominant sorting of the population members and its results were compared with the exact solution to validate the proposed algorithm. In the end, the results are analyzed.

• A study was conducted by Ghazanfari et al. in 2008 with the topic of developing the optimal model of the total cost of logistics distribution in the condition of one manufacturer and several distribution warehouses in supply chain management (SCM). In this article, the optimal model of logistics distribution in the condition of one manufacturer and several distribution warehouses have been developed. In this model, taking into account the route, the means of transport, the volume of the means of transport, as well as other transportation costs, the best value for transportation from the route and the desired means has been proposed, and limitations such as the capacity and number of distribution warehouses have been proposed. The volume of the vehicle is taken into account. An example has been solved in this field, by changing the costs, an attempt has been made to create the best solution and check the results.

data	No
ID	**1**
Block warehouse	**2**
how carry and quote	**3**
contact Hi care from customer	**4**
rank classification customer	**5**
cost product	**6**
Purchases Previous	**7**
Importance product	**8**
gender	**9**
discount presentation done	**10**
weight on according to hot	**11**
to time arrival of goods	**12**

Table 1- Data used for model input

3- Modeling tools and methods

1-3- Data used for model input

The data set used to build the model in this modeling included 10,999 observations of 12 variables. For Dij Kala company, it is product transportation tracking data, in order to answer questions including: What was the customer rating? And was the product delivered on time? Are customer questions answered? If the importance of the product is high. Top rated or on time delivery?

Therefore, for machine learning, the samples are entered in the order shown below with all the specifications.

The specifications of the samples include the customer's ID number, warehouse block, shipping method, customer care calls, customer rating, product cost, previous purchases, product importance, male and female, the discount offered on it. The specific product is divided according to the weight in grams and the timely delivery of the product to the customer.

2-3- Techniques and algorithms used in research 1-2-3-Machine learning technique

In this research, the learning process is used with observations or data such as examples, direct experience or instructions to look for patterns in the data and make better decisions in the future based on the examples we provide to model the on-time delivery of goods. In this research, the monitoring machine learning method is used. This method learns to map data (today's temperature, historical average, etc.) to outputs (tomorrow's maximum temperature) in the training phase or fitting the model.

Most machine learning methods use supervised learning. In supervised machine learning, the system tries to learn from the previous examples it has been given. In other words, in this type of learning, the system tries to learn the patterns based on the examples given to it.

Mathematically speaking, when the input variable (X) and the output variable (Y) are available and based on them, an algorithm can

be used to obtain an input-output mapping function, in fact supervised learning is done. The mapping function is shown as (Y = f(X). In machine learning, the algorithm is given a set of data (s) and the machine forms its own logic based on that set of data (s). This data set has rows and columns. Rows (also referred to as records and data samples) represent data samples. For example, if the dataset is about soccer games (weather conditions), one row contains information about a specific game. Columns (also referred to as attributes, attributes, characteristics) are actually attributes that describe each data instance.

Supervised machine learning problems can be divided into two categories: classification and regression.

Classification: A problem is considered to be classified when the output variable is a category or group. For an example of this, we can refer to the belonging of a sample to the "female" or "male" categories and an email to the "spam" or "non-spam" categories.

Regression: A problem is regression when the output variable is a real value such as "height". In fact, it is used in classification with discrete variables and in regression with continuous variables.

In more complex problems, the SVC method is used. The method used in this research. In the next chapter, we will see how to use this method to categorize the data of this research .

3-2-2-Methods for evaluating algorithms used in research

In supervised learning methods, there are two important data sets called training data and test data. Because the ultimate goal of data mining on these data sets is to find the system that governs them, so the efficiency of the classification model is very important. On the other hand, what part of the initial data set is used for training and what part is used for testing depends on the evaluation method used, which we will examine in the following types of popular methods.

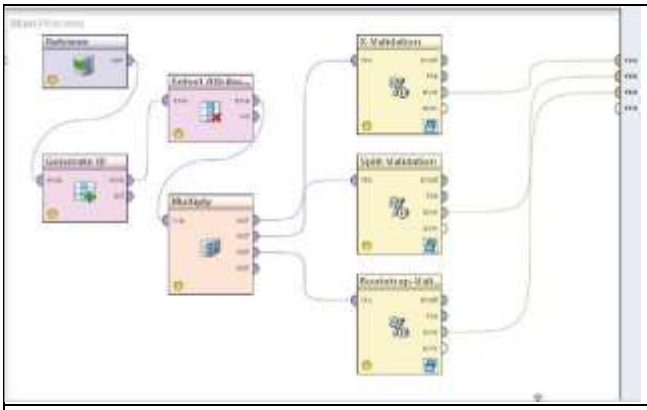


Figure 5- Overview of the use of evaluation methods

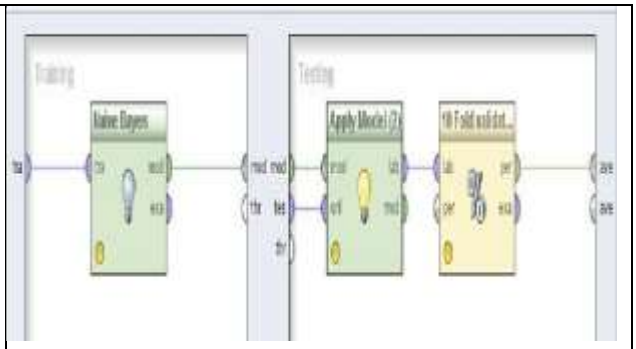


Figure 6 - Overview of the use of a model within an evaluation method

In this section, we will introduce and how to implement classification algorithms and see the results.

In this research, in the implemented algorithms, 5 algorithms were used to evaluate the model, which are: rf or random forest algorithm, knn algorithm, ADB logistic distribution and XGB algorithm, and the results were compared. Figures 5 and 6 show how to use an evaluation model in
Rapidminer.

4-Findings

This project is about e-commerce shipping data. In this chapter, with coding operations and using machine learning techniques, it is predicted that a commercial company with its characteristics or features will reach its destination on time or not, and this work will be optimized and improved. The development of this company will help.

The results of the classification algorithms are compared and the best models are determined in order. Therefore, in this project, using machine learning techniques, we predict whether this company will reach its destination on time with these features or not, and this work will help in the optimization and development of this company.

4-1-Library

The data set used to build the model includes 10,999 observations of 12 variables. In the first stage, 4 libraries were used, and in the first stage, these four libraries, which include numpy, pandas, matplotlip, and seaborn, are entered into the coding environment.

in such a way that the input data is cleaned and sorted using the pandas library. For example, deleting rows that have incomplete values, or completely deleting sections that have no value and are empty, or sorting or filtering specific columns based on specific conditions, and so on. Pandas has 1700 comments on GitHub and an active contributor community of 1200 people. which is heavily used for data analysis and cleaning. Pandas provides fast and flexible data structures, such as CDs dataframes, designed to be very easy and intuitive to work with structured data.

Using the numpy library to provide fast and precompiled functions for numerical routines. , array calculations for better performance, support for an object-oriented approach and compact and faster calculations with vectorization

From Matplotlib to analyze the correlation of variables, visualize the 95% confidence interval of the models. Discover outliers using a scatterplot, etc., and to gain quick insight into the distribution of data

The Seaborn library is used for statistical data visualization.

In Figure 7, the box related to the library is shown. In the next step and in the second box, the data entered into the model is shown in Figure 8.

```
In [1]: import numpy as np
        import pandas as pd
        import matplotlib.pyplot as plt
        import seaborn as sns


        import warnings
        warnings.simplefilter(action='ignore', category=FutureWarning)
```

Figure 7: The box for importing libraries

```
In [2]: df=pd.read_csv('G:/as.csv')
        df
```

Figure 8: The box for entering data

4-2- Importing data

The input data contains the information in Table 2. The specifications of the samples include the customer's ID number, warehouse block, shipping method, customer care calls, customer rating, product cost, previous purchases, product importance, male and female, the discount offered on it. The specific product is divided according to the weight in grams and the timely delivery of the product to the customer.

Table 2- Specifications and descriptions of the samples

no	Type	description
1	ID	ID Number of Customers.
2	Warehouse block	The Company have big Warehouse which is divided in to block such as A,B,C,D,E.
3	Mode of shipment	The Company Ships the products in multiple way such as Ship, Flight and Road.
4	Customer care calls:	The number of calls made from enquiry for enquiry of the shipment.
5	Customer rating	The company has rated from every customer. 1 is the lowest (Worst), 5 is the highest (Best).
6	Cost of the product:	Cost of the Product in US Dollars.
7	Prior purchases	The Number of Prior Purchase.
8	Product importance	The company has categorized the product in the various parameter such as low, medium, high.
9	Gender	Male and Female.
10	Discount offered	Discount offered on that specific product.
11	Weight in gms	It is the weight in grams.
12	Reached on time	It is the target variable, where 1 Indicates that the product has NOT reached on time and 0 indicates it has reached on time.

•ID: customer ID number
• Warehouse block: The company has a large warehouse that is divided into blocks A, B, C, D, E
• Shipping method: The company ships the products by various methods such as ship, air and road
• Customer Care Calls: Number of calls made from inquiry to shipment inquiry

- Customer Rating: The company has rated each customer. 1 is the lowest (worst), 5 is the highest (best).
- Product Cost: Product cost in US dollars
- Previous purchases: number of previous purchases
- Importance of the product: This company has categorized the product in different parameters, low, medium, high
- Gender: male and female
- Discount Offered: Discount offered on that particular product
- Weight in grams: Weight is in grams
- Arrived on time: is the target variable where 1 indicates that the product was not received on time and 0 indicates that it arrived on time.

3-4- Integration

In this section, in order to sort the data, the data indicating the non-delivery of the goods to the customer in one category, the results of which include 6563 numbers and the data indicating the timely delivery of the goods, which The obtained results are 4436, they are integrated separately according to Table 3 and Figure 9.

```
In [4]: df_id = df['ID']
        df = df.drop(columns = ['ID'])

In [5]: df['Reached.on.Time_Y.N'] = df['Reached.on.Time_Y.N'].astype('str')
        df['Reached.on.Time_Y.N'] = df['Reached.on.Time_Y.N'].replace({'1': 'No', '0': ''

In [6]: df['Reached.on.Time_Y.N'].value_counts()

Out[6]: No     6563
        Yes    4436
        Name: Reached.on.Time_Y.N, dtype: int64
```

Figure 9 – Data boxes 4 to 6 related to data integration

Table 3-Results related to data integration

Timely delivery of goods		Failure to deliver goods on time	
Abundance	number	% frequency	**number**
41	4436	59	**6563**

4-4-Noising and removing extraneous data

In this part of the research, it was done to obtain correct and real data. An outlier is a data that is observed at an unusual distance from the rest of the data values in a random sample of a population.

To some extent, the distance of this data from the rest of the data is unusual, which prompts the analyst to investigate what seems unusual in the data. In other words, outliers are values that are prominently and distinctly placed in the main pattern of the data set or graph of data. Figures 10 and 11 show the steps of noise and outlier data removal

```
In [11]: dfnum = df.drop(columns = ['Warehouse_block', 'Mode_of_Shipment',
                                   'Product_importance', 'Gender',
                                   'Reached.on.Time_Y.N'])
         fig,ax= plt.subplots(ncols=3,nrows=2,figsize=(16, 12))
         for variable, subplot in zip(dfnum.columns,ax.flatten()):
             boxdetails = sns.boxplot(x = dfnum[variable], ax=subplot)
             boxdetails.set_xlabel(variable, fontsize = 18)

         plt.show()
         #df_num=data.select_dtypes(include=[np.number])
```

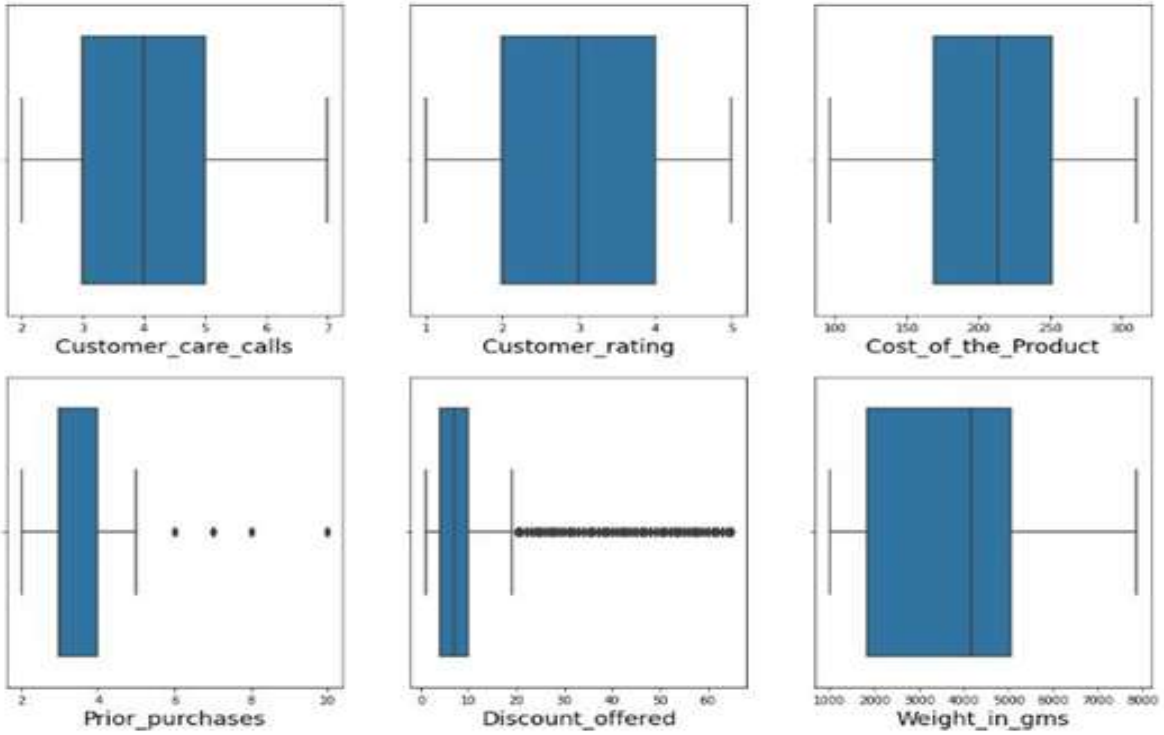


Figure 10-ure 12 The initial stages of noise

```
In [12]: q1 = df['Discount_offered'].quantile(0.25)
         q3 = df['Discount_offered'].quantile(0.75)
         IQR = q3-q1

In [13]: upper_limit = q3 + 1.5*IQR
         df.loc[df['Discount_offered'] > upper_limit, 'Discount_offered'] = upper_limit

In [14]: sns.boxplot(data = df['Discount_offered'])
         plt.show()
```

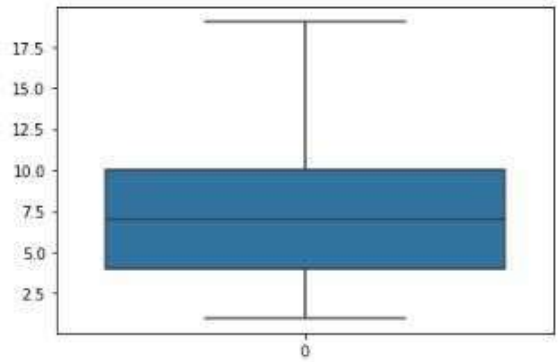
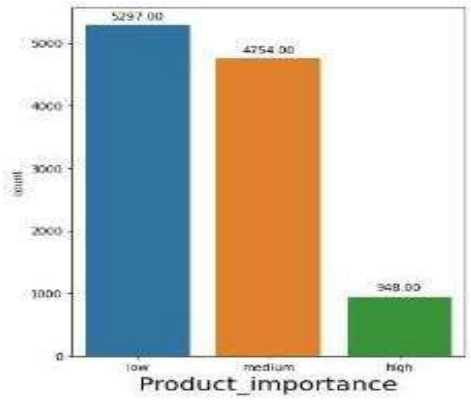


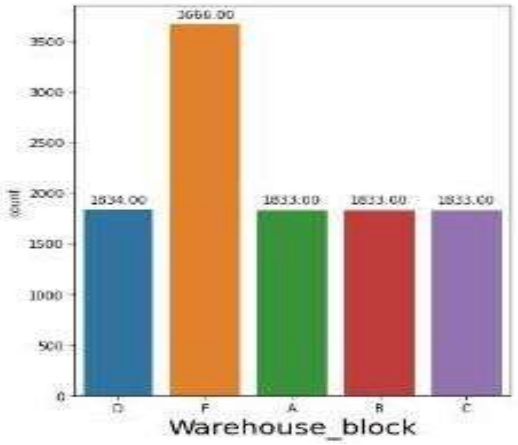
Figure11-The second stage of noise generation

4-5- Data optimization

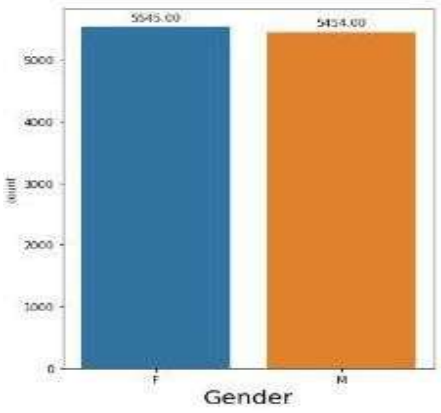
This part is related to the count chart. In this part, the optimal and final results of the data related to the warehouse-block, the importance of the product, the mode of transportation and the gender, which have been noised and come in an acceptable form, their results are shown in the diagram and graphs.



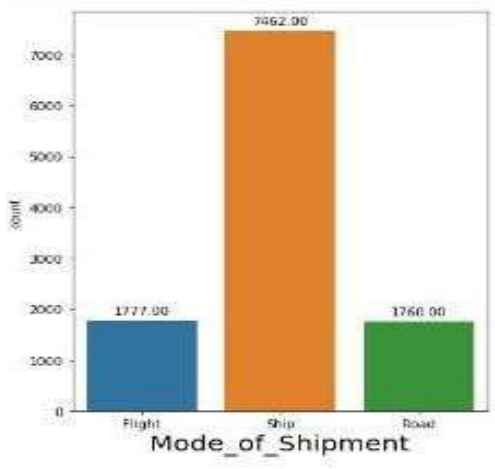
B: The importance of the product



A: Warehouse - block



D: Gender



A : mode of shipment

Figure 12 - count chart and final optimization of related data F: warehouse - block B: product importance C: transportation method D: gender

4-6-Correlation between features

In this part, the correlation between the desired indicators has been investigated according to the optimized data explained in the previous part. Table 4 shows the correlation between these indicators.

Table 4- The degree of correlation between indicators

	Customer care calls	**Customer rating**	**Product cost**	**Previous purchases**	**Discount provided**	**Weight in grams**
Customer care calls	1.000000	0.012209	0.323182	0.180771	0.133149-	0.276615-
Customer rating	0.012209	1.000000	1.000000	0.013179	0.001346-	0.001897-
Product cost	0.323182	0.009720	0.123676	0.123676	0.143676-	0.132604
Previous purchases	0.180771	0.013179	0.143876	1.000000	0.085184-	0.168213-
Discount provided	0.133149-	0.133149-	0.143876-	0.085184-	1.000000	0.389933-
Weight in grams	0.276615-	0.001897-	0.132604-	0.168213	0.389933-	1.000000

4-7- Data scaling

In general, the scale of data is of great importance in any type of machine learning problem or any illustration. The reason for this importance is that almost all algorithms in this field are based on mathematical calculations. Therefore, the lack of scale of the data causes errors in the visualization of the data and also causes mistakes in the calculations and performance of the algorithms and predictions. Therefore, in this part of the research, data scaling is discussed. In Figure 13, a sample of the scaled and standardized data is shown, and in Figure 13 to Figure 14, the transformation of the obtained and standardized numbers into mathematical language is shown.

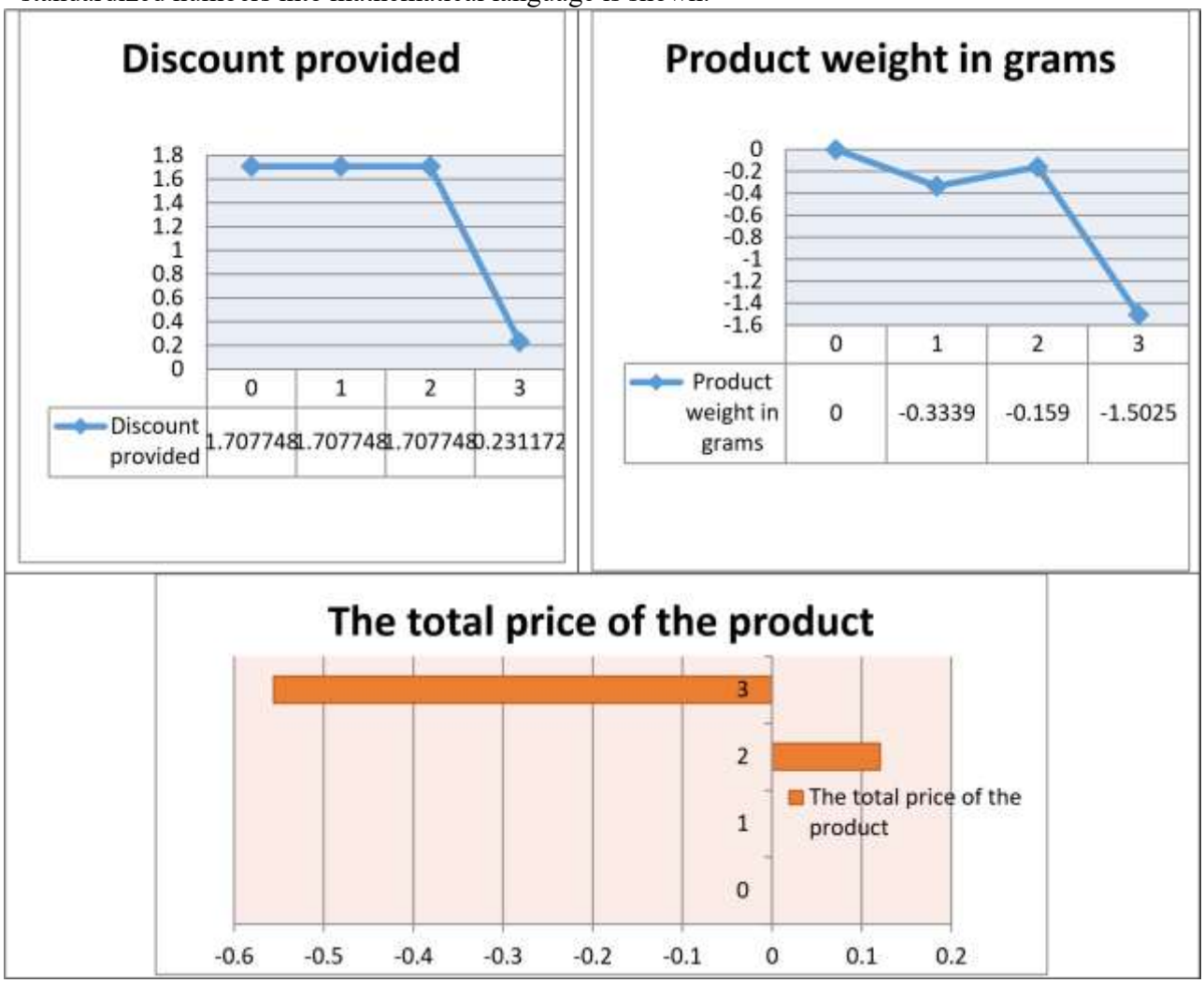


Figure 13- Sample of scaled and standardized data

	Cost_of_the_Product	Discount_offered	Weight_in_gms	Warehouse_block_B	Warehouse_block_C	Warehouse_block_D	Warehouse_block_F	Mode_of_Shipmer
0	-0.690722	1.707748	-1.468240	0	0	1	0	
1	0.120746	1.707748	-0.333893	0	0	0	1	
2	-0.985881	1.707748	-0.159002	0	0	0	0	
3	-0.711529	0.231172	-1.502484	1	0	0	0	
4	-0.548074	1.707748	-0.700244	0	1	0	0	

Figure 14- Converting the obtained and standardized numbers into mathematical language **4-8- Algorithms of car models and predicting the timely arrival of goods**

In this section, the desired algorithms for making car models and predicting the timely arrival of the goods to the destination and to the hands of the customer have been discussed. earned Then, in order to obtain a valid model, training data is given to the model, which is about 42%, then 0.2% of that data is selected for data testing. Finally, as shown in Figure 15, 42% of the data are used to create the random forest algorithm, and these data are fit.

```
In [39]: x1 = df_model.drop(['Reached.on.Time_Y.N'] ,axis = 1)
         y1 = df_model['Reached.on.Time_Y.N']

In [40]: X_train1, X_test1, y_train1, y_test1 = train_test_split(x1, y1, test_size=0.2, random_state=42)

         model =  RandomForestClassifier(random_state = 42)
         model.fit(X_train1, y_train1)

Out[40]: RandomForestClassifier(random_state=42)
```

Figure 15- Determine x-y to make the model and make the model fit the random forest algorithm

4-10- Determination of the best parameters for the model

Considering that each of the models has a type of settings, which should be determined using the trial and error method, which models are better than other models according to the important and main parameters that have been determined in the previous sections. Finally, rf algorithm or random forest, knn algorithm, logistic distribution, ADB, XGB were selected for final evaluation.

4-11- Model evaluation

In this part, model evaluation is discussed in terms of the accuracy of training and testing and the final score of each model. The steps are shown in Figure 16 and the general results related to training and testing and the final score of the models are shown in Figures 16 to 19.

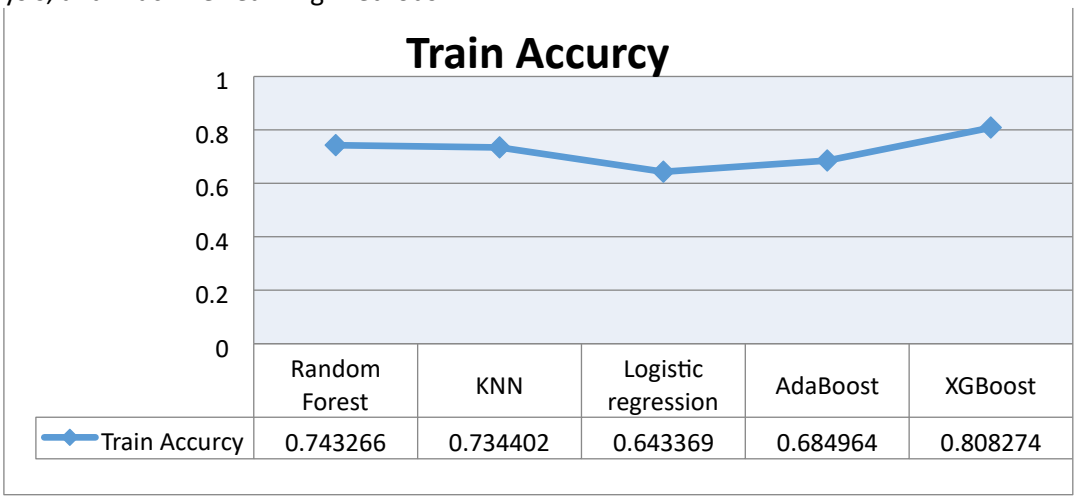


Figure 16- Data related to training accuracy of models (Train Accuracy)

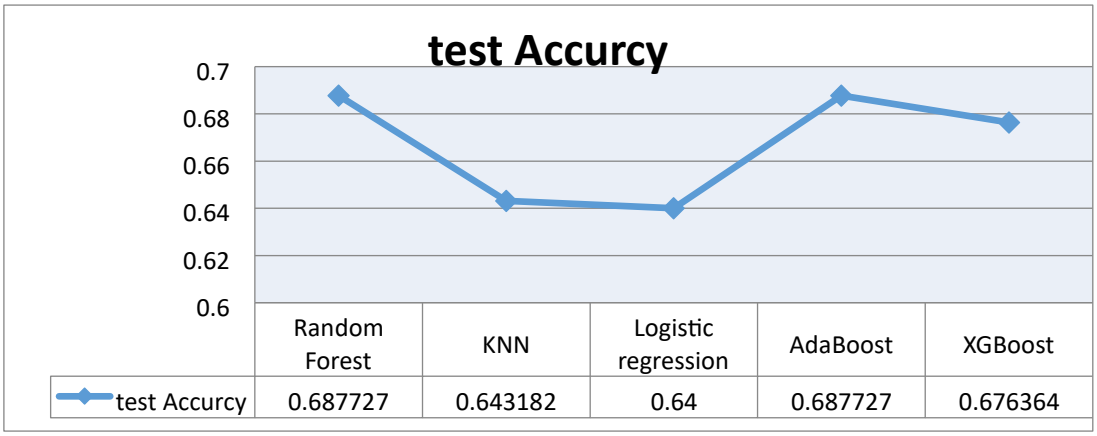


Figure 19: Data graph related to the accuracy of the models test (Test Accuracy)

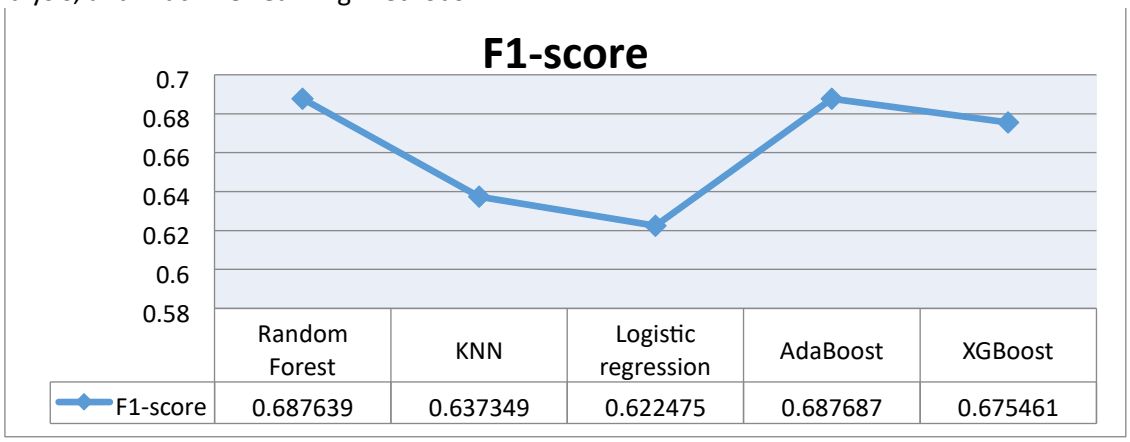


Figure 19: Data graph related to the accuracy of the models test (F1-score)

6-Final results

By comparing the researched models whose results are shown in the graph, it can be seen that among the 6 investigated models, the Stacking and XGBoost models got the highest score compared to other models.

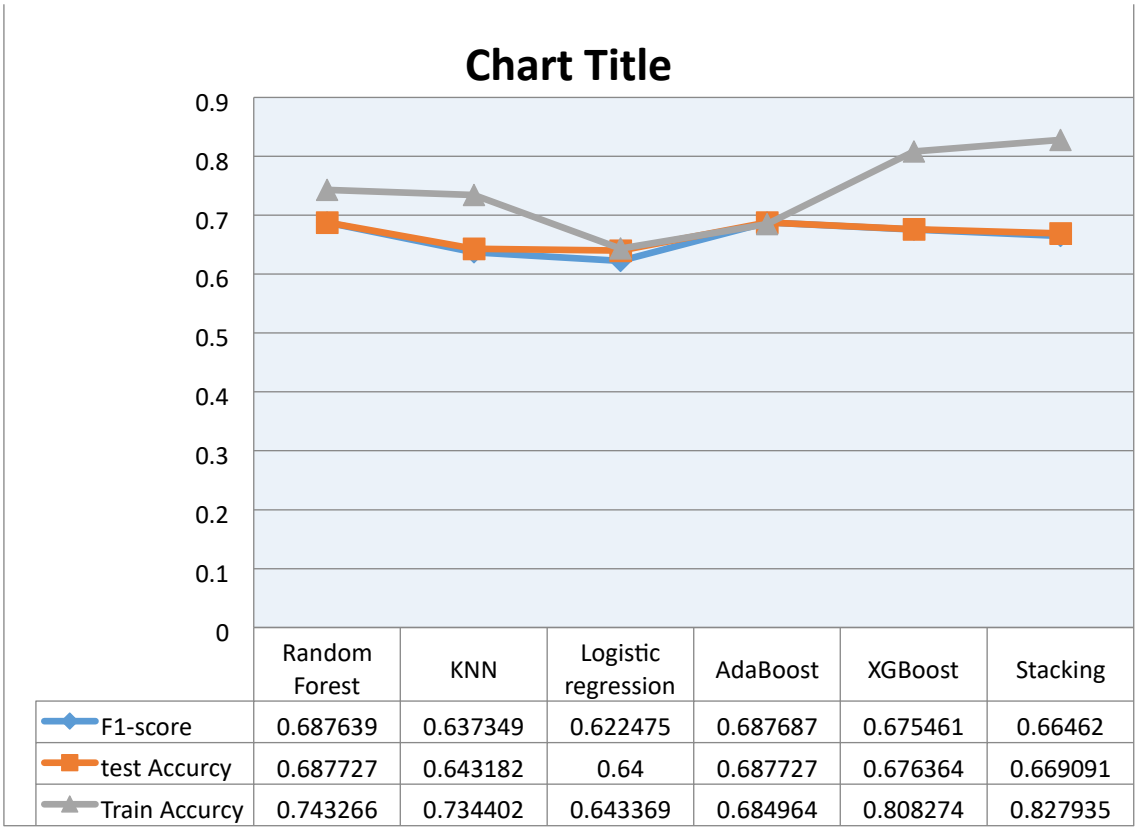


Figure 20- The comparison chart of the researched models

According to this Figure, we can say that the best model should be Random Forest or AdaBoost. The KNN VE logistic regression has a very low test score, but the lack of fit problem in the XGBoost and Stacking models is quite evident despite the high train score. If you have to choose a model it should be AdaBoost because the test scores and AUC are close to each other, but the random forest model seems inadequate.

5-Conclusion

The purpose of this research is to analyze the logistics distribution model in the context of ecommerce with an emphasis on big data analysis: designing a routing model by machine learning methods, the results of which are analyzed in the following discussion. Therefore, this research investigates whether the product was delivered on time in Snap company or not. In this research to meet e-commerce customer demand, an established company wants to discover key insights from its customer database. They want to use some of the most advanced machine learning techniques to study their customers and sell electronic products. The data set used to build the model includes 10,999 observations of 12 variables, including customer identification numbers, warehouse blocks, shipping methods, customer care calls, customer ratings, product cost, previous purchases, product importance, male And woman: The discount offered on that particular product is weight in grams and the product reaches the customer on time.

In this research, we were able to prove that with machine learning, models can be provided that determine the effective factors in the design of a logistics transport routing model in Snap using existing algorithms. In this research, 8 sorting algorithms including rf or random forest algorithms, KNN algorithm, ADB logistic distribution and XGB algorithm have been used to make machine models and predict the timely arrival of goods. Finally, the sorting algorithms were compared and respectively, the best Random Forest or AdaBoost model should be determined, so that in this project, using machine learning techniques, it was predicted that with these characteristics or features, the company reached its destination on time or not, and to what extent Optimizing and improving and developing this company will help.

Refrences

1. Badri M, Kazemi Seyyed Mohammad, Rahimi H, (2025), Presenting the behavioral model of citizens in selection of trip vehicle with emphasis on how to go to work, Case Studies on Transport Policy Volume 19 , 101304. https://doi.org/10.1016/j.cstp.2024.101304

2. Shokrollahi M, Eftekhari E, Mostahfezian M, Zahedi H, (2024), Adoption of Strategies of Electronic Digital Innovation and Transformation and New Information Technologies Via Application of AI based on Organizational Culture Within Commercial Companies , power system technology journal, Volume 48 Issue 2, 418.

3. Shokrollahi M, Eftekhari E, Mostahfezian M, Zahedi H , (2024). The Complex Relationship, Between The Culture Of An Organization And Its Significant Impact On Promoting Novel Principeles Entrepreneurship Within Knowledge Based Companies, In Iran. Educational Administration: Theory and Practice, 30(7), 903–910. https://doi.org/10.53555/kuey.v30i7.6886

4. Simsci-Levi D. Designing and managing the supply chain. McGraw-Hill; 2000.

5. Robertson C, Parrish, M., Casey, S., McBain, J., Ramos, L., Wizdo, L., ... Hartig, K. Predictions 2019: B2B marketing and sales. 2018.

6. Anderson B. Forrester B2B Marketing & Sales 2018: Meeting The Business Consumer On Their Terms Requires Dynamic, Customer-First Experiences. 2018.

7. Miguel JC, Casado MÁ. GAFAnomy (Google, Amazon, Facebook and Apple): The big four and the becosystem. Dynamics of big internet industry groups and future trends: Springer; 2016. p. 127-48.

8. Satish Meena SFA. Amazon And Flipkart Are Expected to Record $1.2 Billion To $1.5 Billion In Sales In India's Coming Festive Season. SEP 19 2017.

9. Borak M. Meituan dianping launches autonomous food delivery system. Aug 29, 2018(Retrieved from https://technode.com/2018/07/26/meitua n-dianping-autonomousdelivery/.

10. Alexander Osterwalder & Yves Pigneur. (2002). An e-Business Model Ontology for Modeling e-Business. Ecole des HEC, Université de Lausanne, 1015 Lausanne. Bled, Slovenia, June 17 – 19 .

11. Andrew N. K, et al. , (2006), "Strategies for effective Web services Adoption for dynamic e-businesses", Decision Support Systems, 42, pp: 789- 809

12. Interreg Europe. Sustainable Urban Logistics. A Policy Brief from the Policy Learning Platform on LowCarbon Economy. 2020. Available online: http://www.interregeurope.eu/fileadmin/user_upload/plp_uploads/policy_briefs/Sustainable_urban_logistics.pdf (accessed on 15 March 2022).

13. DHL. Logistics Trend Radar 5th Edition. 2021. Available online: https://www.dhl.com/globalen/home/insights-andinnovation/insights/logistics-trend-radar.html (accessed on 15 March 2022).

14. Zhao, M.; Ji, S.; Wei, Z. Risk prediction and risk factor analysis of urban logistics to public security based on PSO-GRNN algorithm. PLoS ONE 2020, 15, e0238443.

15. SENATOR, 2021. [D2.1] State of the Art in Optimization and Machine Leaning Algorithms Applied to Last Mile Logistics. Available online: https://www.senatorproject.eu/wpcontent/uploads/2021/10/D2.1-SotA-inoptimization-and-machine-leaningalgorithms-applied-to-last-milelogistics-Public.pdf (accessed on 19 February 2022).

16. Finnegan, P. , & Hayes, J. (2005). Assessing potential of e0business models: Toward a framework for decision-maker. *European Journal of Operation Research*, 160, 365-379.

17. Hanafizadeh, P. (2006). *E-commerce, definitions and barriers*. Tehran: Sharif university, (in Persian) .

18. Ambrazio, J. and cekotya, P. , (2005), "Assessing the of potential of e-business models: towards a framework for assisting decision-makers", European Journal of Operational Research 160, pp: 365-379 .

19. Khatami Firouzabadi, A. and Shafiei Nikabadi, M. (2008), "A Framework for Evaluation of E-Business Efficiency (a DEA Approach)" , Commercial Surveys Journal, Vol. 4, No. 30,pp: 19-29 (in persian)

20. Alcaraz, J.J.; Caballero-Arnaldos, L.; Vales-Alonso, J. Rich vehicle routing problem with last-mile outsourcing decisions. Transp. Res. Part E: Logist.

Transp. Rev. 2019, 129, 263–286. [Google Scholar] [CrossRef]

21. Baller, A.C.; Dabia, S.; Dullaert, W.E.H.; Vigo, D. The vehicle routing problem with partial outsourcing. Transp. Sci. 2020, 54, 1034–1052. [Google Scholar] [CrossRef]

22. Zhang, H.; Zhang, Q.; Ma, L.; Zhang, Z.; Liu, Y. A hybrid ant colony optimization algorithm for a multiobjective vehicle routing problem with flexible time windows. Inf. Sci. 2019, 490, 166–190.

23. Qin, G.; Tao, F.; Li, L. A vehicle routing optimization problem for cold chain logistics considering customer satisfaction and carbon emissions. Int. J. Environ. Res. Public Health 2019, 16, 576.

24. Elshaer, R.; Awad, H. A taxonomic review of metaheuristic algorithms for solving the vehicle routing problem and its variants. Comput. Ind. Eng. 2020, 140, 106242. [Google Scholar]
[CrossRef]

25. Costa, L.; Contardo, C.; Desaulniers, G. Exact branch-price-and-cut algorithms for vehicle routing. Transp. Sci. 2019, 53, 946–985.

26. LAL. K,. (2002), E-BUSINESS AND EXPORT BEHAVIOR, EVIDENCE FROM INDIA FIRM WORLD INSTITUTED FOR DEVELOPMENT ECONOMICS RESEARCH.

27. Linder, J. C. , & Cantrell, S. (2000). *Changing business models*. Chicago: Institute for Strategic Change .

28. Magertta, J. (2002). Why business models matter. *Havard Business Review*, 80, 86-92 .

29. Ng, E. (2005). An empirical framework developed for selecting B2B e-business models. *Journal of Business & Indusrrial Marketing*, 218-225 .

30. Pavic, S. K. , S. C. L. Simpson, M. Padmore, J. (2007). "Could e-business create a competitive advantage in UK SMEs?. " Benchmarking: An International Journal, pp 58-63.

31. Petrovic, O. , C. Kittl, et al. (2001). Developing Business Models for eBusiness .

32. Rappa, M. (2003). Managing the digital enterprise: *Business models on the web*. North Carolina State University .

33. Rayport, Jeffrey F. , Bernard J. Jaworski, (2001); Introduction to Ecommerce, Mc Graw-Hill .

34. Lickert, H.; Wewer, A.; Dittmann, S.; Bilge, P.; Dietrich, F. Selection of Suitable Machine Learning Algorithms for Classification Tasks in Reverse Logistics. Procedia CIRP 2021, 96, 272–277. [

35. Tamayo, S.; Combes, F.; Gaudron, A. Unsupervised machine learning to analyze City Logistics through Twitter. Transp. Res. Procedia 2020, 46, 220–228. Tian, Z.; Zhong, R.Y.; Vatankhah Barenji, A.; Wang, Y.T.; Li, Z.; Rong, Y. A blockchain-based evaluation approach for customer delivery satisfaction in sustainable urban logistics. Int. J. Prod. Res. 2021, 59, 2229–2249.

36. ang, S.; Zhang, Y.; Liu, R.; Jafari, M.; Kharbeche, M. Data-driven optimization for dynamic shortest path problem considering traffic safety. IEEE Trans. Intell. Transp. Syst. 2022, 23, 18237–18252.

37. Jain, N.; Husain, S.O.; Goyal, S.; Hariharasudhan, S.; Victor, M.; Manjula. Predictive Analytics for Network Traffic Management. In Proceedings of the 2024 IEEE International Conference on Communication, Computing and Signal Processing (IICCCS), Asansol, India, 19–20 September 2024; pp. 1–6.

38. Metrix, S. Optimisation des Transports: Technologies émergentes. 2023. Available online: https://solutionsmetrix.com/fr/article/optimisation-des-transports-grace-auxnouvelles-technologies/ (accessed on 2 April 2025).

39. Bharadiya, J.P. Artificial intelligence in transportation systems a critical review. Am. J. Comput. Eng. 2023, 6, 35–45.

40. Akhtar, M.; Moridpour, S. A review of traffic congestion prediction using artificial intelligence. J. Adv. Transp. 2021, 2021, 8878011.

41. Seyedhosseini SM, Ghazanfari H. Total Logistic Distribution Cost Optimization Modeling Approach, When there is one Producer and Several Storage Distribution Centers In a Supply Chain Management. International Journal of Industiral Engineering & Producion Management. 2008;19(1):117-22.