ỨNG DỤNG THUẬT TOÁN K-MEANS CÂN BẰNG VÀ TABU SEARCH TRONG VRP

Application K-means balance and Tabu Search algorithm in VRP

line 1: Trần Hữu Ân   
line 2: HCM City *University of transport*   
line 3: *HCM City University of transport*line 4: HCM City, Viet Nam  
line 5: tranhuuan0342@gmail.com

*Abstract*— VRP (vehicle routing problem) là bài toán thách thức trong việc thiết kế các tuyến đường tối ưu từ kho đến một tập hợp điểm khách hàng với các ràng buộc cụ thể của doanh nghiệp, Chẳng hạn như hạn chế về phương tiện kiểm soát, hạn chế về chi phí, khoảng thời gian, nguồng lực liên quan đến quá trình xếp hàng tại kho,… . Trong bài báo này. Nhóm nghiên cứu muốn đề cập đến quá trình thiết kế đường đi cho tuyến đường vận tải dựa trên trên cơ sở của thuật toán Kmeans để phân loại các cụm khách hàng. Với mục tiêu là tối ưu hóa quá trình vận chuyển từ điểm kho đến tập khách hàng đã phân cụm sao cho điểm cuối cùng trở về là điểm kho. Bài báo đề xuất hướng giải quyết cho bài toán phân bổ đều ở các cụm khách hàng sử dụng thuật toán kmeans sau mỗi một cụm sẽ được một phương tiện di chuyển để giao hàng đến tất cả các điểm khách hàng nằm trong cụm đó bằng phương pháp Tabu search. Hướng giải quyết của nhóm nghiên cứu đã giảm thiểu được thời gian thực thi và cân bằng được khoảng cách di chuyển của các phương tiện với nhau. Mô hình bài báo cho thấy được sự cải thiện trong việc tối ưu hóa khoảng cách đường đi của mỗi phương tiện so với thuật toán phân cụm genetic. Giảm thiểu được sự chênh lệch tổng khoảng cách di chuyển của mỗi phương tiện.

Keywords— Vehicle routing problem, K-Means clustering, tabu search, supply chain management.

# Giới thiệu.

VRP là một bài toán được khái quát hóa của bài toán “Travel salesman problem”. Nó xuất hiện đầu tiên vào năm 1959 bài báo The Truck Dispatching Problem [1] của C. B Dantzig, J. H. Ramser, bài báo đề cập đến việc tối ưu hóa quá trình vận chuyển của một đội phân phối xăng từ một bãi kho đến các trạm dịch vụ do kho cung cấp.

VRP (vehicle routing problem) là một trong những vấn đề của tối ưu hóa trong quản lý và quá trình vận hành vận tải. VRP giúp các doanh nghiệp có thể giải quyết bài toán về tối ưu hóa được quá trình vận chuyển của một đội phương tiện của doanh nghiệp, giúp giảm thiểu được quá trình vận chuyển, chi phí mục tiêu tăng hiệu quả hoạt động của doanh nghiệp. Đây là một vấn đề quan trọng trong quản lý chuỗi cung ứng. Vấn đề này có thể mang lại nhiều tổn thất và hậu quả nặng nề cho doanh nghiệp nếu không biết cách tối ưu và quản trị hợp lý.

# Công trình liên quan.

“Cluster-based Hyper-Heuristic for Large-Scale Vehicle Routing Problem” [2] Bài báo nghiên cứu về một phương tiếp cận hai bước để giải quyết định tuyến xe tải các sức chứa (CVRP) sử dụng một đội xe tải không đồng nhất. Trong CVRP một độ xe tải có sức chứa và đặt ở một kho trung tâm sau đó sẽ đi giao sản phẩm đến một tập khách hàng với nhu cầu đã biết. Ở bước đầu tiên mô hình bài báo áp dụng gom cụm K-means để tổng hợp khách hàng thành cụm với nhau. Trong bước hai, các cụm đã tổng hợp được gán cho các xe tải và một mô hình tối ưu hóa cố định (MIP) để giải quyết. Trong bước này, các ràng buộc được thêm vào mô hình để giảm kích thước mô hình tối ưu hóa. Sau đó phân cấp để tìm ra các hanh trình cho mỗi xe.

“Solving the Heterogeneous Capacitated Vehicle Routing Problem using K-Means Clustering and Valid Inequalities” [3] tác giả đề xuất cách tiếp cận hai bước để tối ưu hóa quá trình giao hàng của một đội xe tải. Bước đầu tiên sử dụng thuật toán gom cụm K-Means để tổng hợp các cụm khách hàng với nhau. Sau đó, trong bước hai, các cụm này được gán cho các xe tải để tìm lời giải tối ưu cho hành trình đi của xe.

# Xác định vấn đề.

Bài báo muốn đề cập đến vấn đề của VRP như sau. Một kho hàng sẽ có một đội phương tiện di chuyển đến các điểm khách hàng để phân phối hàng đến các đại lý. Bài toán hoạch định đường đi tối ưu cho từng phương tiện theo sự phân bổ ngẫu nhiên cho từng khách hàng. Bài toán được xây dựng vấn đề như sau.

Xét một mạng lưới tuyến đường trong vận chuyển được biểu diễn với đồ thị . Trong đó đại diện cho các điểm tập khách hàng và điểm 0 là điểm kho, , . Với K là tập phương tiện, mỗi . là chi phí di chuyển từ điểm đến . là tổng số lượng đường đi giữa hai điểm bao gồm cả điểm kho, là tuyến đường phương tiện k có đi qua hay không được gán là 1 nếu đi qua và 0 nếu không đi qua.

Dựa trên các biến xác định trên mô hình bài toán được mô tả như sau:

* Hàm mục tiêu:
* Các ràng buộc.

|  |  |
| --- | --- |
|  | * Phương tiện k phải xuất phát từ điểm kho. |
|  | * Phương tiện k phải kết thúc tại điểm kho. |
|  | * Đường đi của phương tiện k không được lặp lại. |

# phương pháp nghiên cứu.

## Phân loại trong chuỗi cung ứng.

Phân cụm trong tối ưu hóa tuyến đường đi là một khía cạnh trong lĩnh vực quản lý giao thông vận tải. Một số công trình nghiên cứu đã chứng minh được sử dụng các thuật toán phân cụm mang lại hiệu suất tốt trong tối ưu hóa tuyến đường đi chẳng hạn như trong bài báo của [L Rizkallah](https://scholar.google.com/citations?user=vQs6-VUAAAAJ&hl=en&oi=sra), [M Farouk](https://scholar.google.com/citations?user=fBZNfNcAAAAJ&hl=en&oi=sra), [N Darwish](https://scholar.google.com/citations?user=aEmk2o8AAAAJ&hl=en&oi=sra) [4]. Đã chứng minh được phương pháp tiếp cận này đã thành công trong việc giải quyết vấn đề về VRP vượt xa các phương pháp hiện đại khác . Phân cụm trong tối ưu hóa tuyến đường đi là quá trình gom nhóm các tuyến đường dựa trên các yếu tố như các điểm khách hàng, tải trọng giao thông, khoảng cách, tốc độ và các mục tiêu tối ưu hóa khác nhau. Mục tiêu chính của phân cụm là tạo ra các nhóm nhỏ để có thể dễ dàng quản lý và quá trình vận chuyển được tối ưu hơn.

## Phương pháp giải quyết.

Trong thế giới kinh doanh ngày nay. Việc quản lý và lưu trữ các địa điểm khách hàng trở nên đơn giản bởi các phần hệ thống phần mềm quản lý. Vì vậy trong bài báo này mô hình được áp dụng dựa trên đầu vào là các tọa độ điểm khách hàng.

Phương pháp tiếp cận ở bài toán đã xác định trên được nhóm nghiên cứu sử dụng thuật toán Kmeans để phân cụm điểm khách hàng và thuật toán Tabu search để tìm đường đi ngắn nhất của một phương tiện trong cụm đó. Cách tiếp cận này của nhóm nghiên cứu được chia thành hai bước.

### K-Means.

Thuật toán Kmeans phân loại các cụm điểm khách hàng dựa tọa độ điểm khách hàng. Ở bước này có thể được lặp lại thuật toán với số lần n để tìm ra lần phân cụm đều nhất. Giả sử ta có một tập tập hợp các điểm khách hàng, mỗi điểm khách hàng chứa tọa độ và với K là số cụm phân loại trong thuật toán kmeans. Chúng ta cần tìm tập điểm là điểm nằm chính giữa ở mỗi cụm phân loại. Với mỗi ta có là nhãn cho mỗi điễm dữ liệu đó nếu điểm được phân vào cụm k thì và .

Để tìm được cụm phân bố đều nhau sau N lần vòng lặp thuật toán kmeans. Nhóm nghiên cứu sử dụng bằng cách tìm độ lệch chuẩn nhỏ nhất sau N lần. Giả sử thuật toán kmeans phân cụm, trong đó là cụm trong lần thực hiện thứ với là trong đó là số lượng chứa điểm khách hàng trong một cụm với .

### Tabu search.

Ở bước hai mô hình áp dụng thuật toán Tabu search để tìm đường đi ngắn nhất từ điểm xuất phát (điểm kho) đi đến các điểm khác và kết thúc tại điểm xuất phát (điểm kho). Theo tác giả [Fred và Glover](https://pubsonline.informs.org/action/doSearch?text1=Glover%2C+Fred&field1=Contrib) trong “tabu search: A tutorial” cho rằng. Thuật toán Tabu Search là thuật toán tiếp cận theo phương pháp heuristic để giải quyết các vấn đề tối ưu hóa. Thuật toán Tabu Search đã thu được các giải pháp tối ưu và gần tối ưu cho nhiều vấn đề chẳng hạn như: Lập lịnh kế hoạch, …. Hàm mục tiêu và các điều kiện ràng buộc sẽ được dựa trên mô hình bài toán ở phần II. Thuật toán sẽ khởi tạo các lời giải ban đầu. Dựa vào hàm mục tiêu và các ràng buộc trên của mô hình bài toán, thuật toán sẽ chọn ra giải pháp tốt nhất.

### Lập trình bài toán.

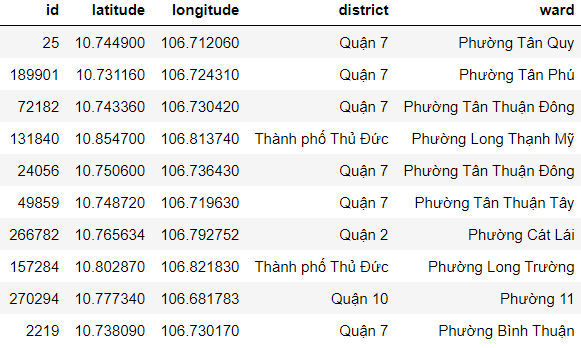
Mô hình sử dụng thuật toán kmean cân bằng sau n lần thực hiện và áp dụng thuật toán Tabu Search được thể hiện qua các bước dưới đây.

|  |
| --- |
| **Bước 1**: khởi tạo tập K điểm bất kỳ, mỗi điểm k chứa tọa độ .  **Bước 2**: Phân các điểm khách hàng vào cụm điểm khách hàng có tâm đó nhất gần nhất.  **Bước 3**: Nếu cụm điểm trung tâm của cụm không thay đổi so với lần tìm cụm trước thì dừng thuật toán.  **Bước 4**: cập nhật tập các vị trí M điểm trung tâm của các cụm bằng cách tính giá trị điểm trung bình của cụm khách hàng đó.  **Bước 5**: Quay lại bước 2.  **Bước 6**: Sau khi thuật toán lặp qua N lần 5 bước trên để thu được mỗi cụm. Thuật toán sẽ lấy cụm được phân ra đều nhất sau N lần lặp dựa trên độ lệch chuẩn của mỗi cụm.  **Bước 7**: Thuật Tabu search được áp dụng giải quyết tìm đường đi ngắn của mỗi phương tiện trong mỗi cụm đã được phân. |

# Thu thập dữ liệu.

Bộ dữ liệu được thu thập GPS là một tài nguyên quan trọng cho nghiên cứu khoa học trong lĩnh vực về kế hoạch đồ thị để thử nghiệm kết quả cho nghiên cứu. Bộ dữ liệu này bao gồm các thông tin về tọa độ địa lý điểm dữ liệu của khách hàng và Quận, Phường của điểm khách hàng đó.

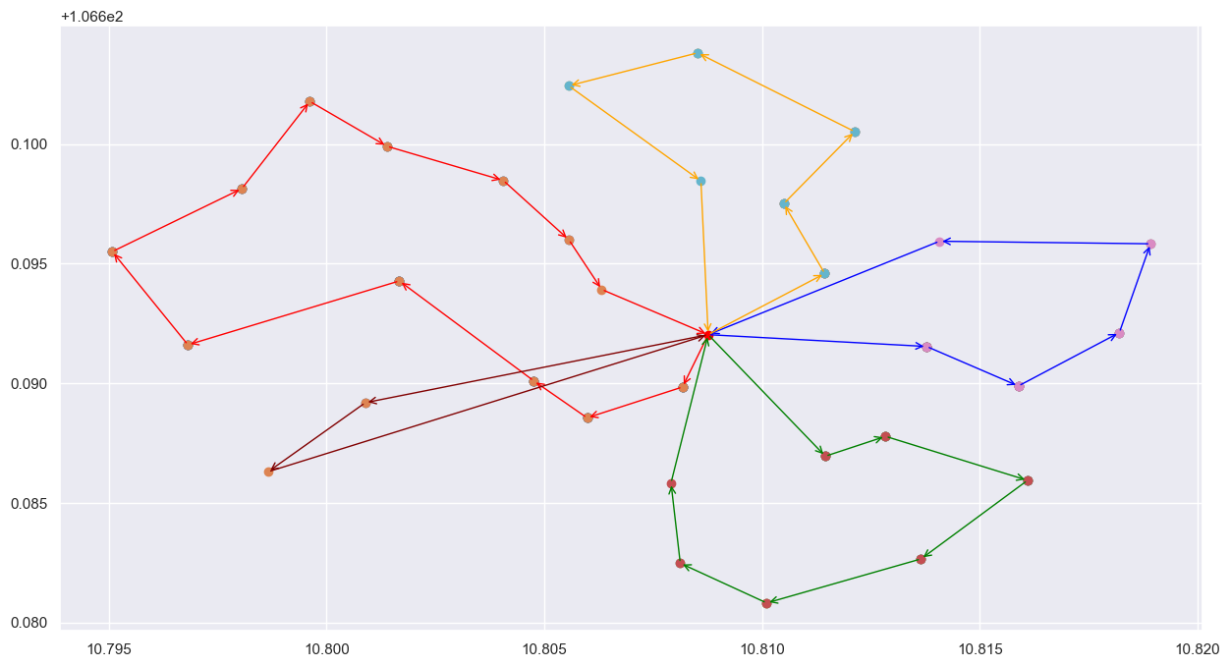
Dữ liệu GPS trên cho phép nhóm nghiên cứu có thể theo dõi quá trình tự phân cụm của thuật toán Kmeans và quá trình định tuyến tối ưu hóa của mô hình. Sử dụng bộ dữ liệu GPS giúp nhóm nghiên cứu có khả năng xây dựng mô hình dựa theo kết quả phân tích và đưa ra những cải thiện cho mô hình. Dưới đây là một vài dữ liệu mẫu trong bộ dữ liệu GPS.



Bảng : Dữ liệu mẫu của bộ dữ liệu GPS.

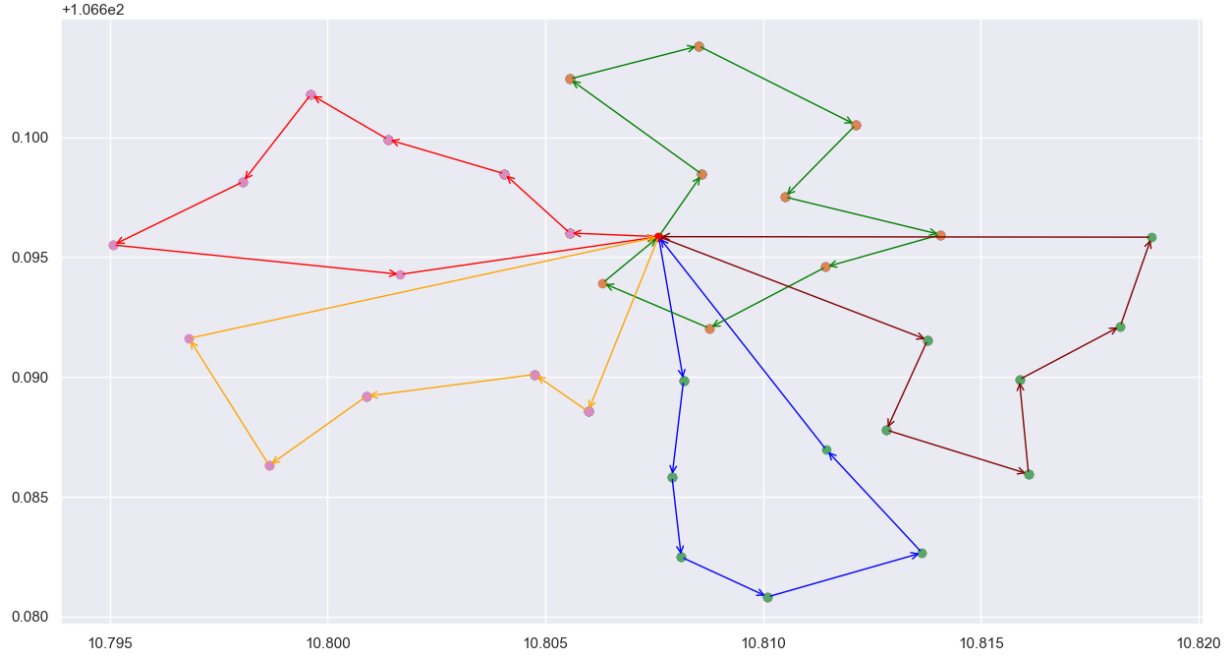
# Kết quả nghiên cứu.

Đối với mô hình áp dụng thuật toán Kmeans và không áp dụng cân bằng các cụm. Giả sử các tuyến đường màu khác nhau là một cụm và ở mỗi cụm đó đều được do một phương tiện phân phối hàng theo hàm mục tiêu và các ràng buộc đã được phân tích ở phần trên. Nhìn vào biểu đồ có thể thấy rằng việc không cân bằng ở các cụm làm cho khoảng cách tuyến của mỗi phương tiện bị chênh lệch. Điều này dẫn đến thời gian thực thi của việc định tuyến cho phương tiện sẽ mất thời gian hơn. Chi tiết được thể hiện ở biểu đồ 1.



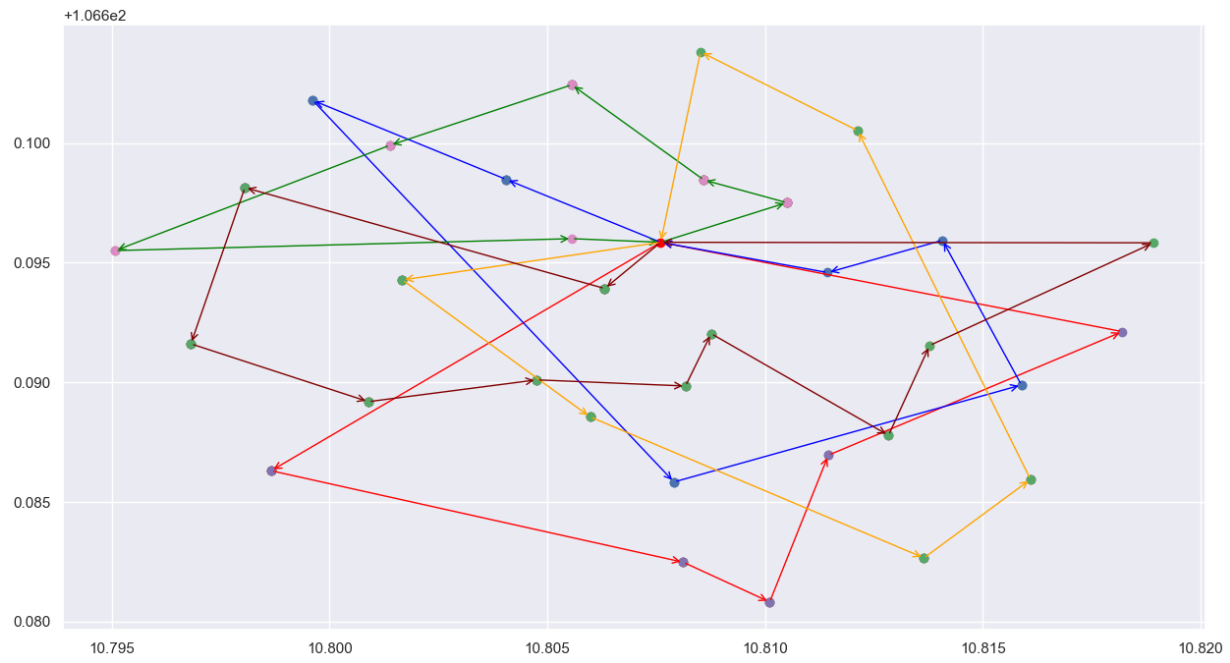
Biểu đồ 1: Biểu đồ áp dụng thuật toán kmeans.

Kết quả nghiên cứu thực nghiệm của mô hình đã nghiên cứu trên để giải quyết bài toán tối ưu hóa quá trình vận chuyển. Khi áp dụng mô hình này vào thực tế với bộ dữ liệu GPS trên đã cho thấy kết quả được cải thiện hơn so với việc sử thuật toán kmean phân loại ngẫu nhiên. Về thời gian thực thi, khoảng cách tối ưu. Dưới đây là hình bảng so sánh giữa cân bằng cụm và sử dụng genetic.



Biểu đồ 2: Biểu đồ tìm đường đi ngắn nhất sử dụng mô hình kmeans cân bằng.

Với mô hình sử dụng kmeans cân bằng để tìm ra đường đi ngắn nhất cho mỗi cụm trên là 151km cho tổng quảng đường của các cụm. Thời gian thực thi để tìm ra quảng đường cho mỗi cụm là 7.51s. Nhìn vào biểu đồ 1 có thể thấy rằng việc phân loại các cụm rõ ràng. Trung bình mỗi phương tiện di chuyển trong một cụm rơi vào khoảng 27 km đến 33 km.



Biểu đồ 3: Biểu đồ sử dụng thuật toán genetic để phân loại.

Đối với biểu đồ 2 sử dụng thuật toán genetic để phân cụm để tìm ra đường đi ngắn nhất cho mỗi cụm trên là 248km cho tổng quản đường đi của các cụm. Thời gian thực thi để tìm ra quảng đường ngắn nhất cho mỗi cụm là 7.11s. Nhìn vào biểu đồ trên có thể thấy rằng việc phân cụm của thuật toán này chưa rõ ràng và quảng đường di chuyển dài hơn so với thuật toán kmeans cân bằng. Các phương tiện di chuyển trong mỗi cụm rơi vào khoảng từ 32km đến 50km.

# kết luận.

Tối ưu hóa đường đi trong chuỗi cung ứng đang trở thành một lĩnh vực quan trọng trong quản lý chuỗi cung ứng. Tối ưu hóa định tuyến đường đi đóng một vai trò quan trọng trọng việc cải thiện hiểu quả vận chuyển và giảm thiểu chi phí trong quá trình di chuyển hàng hóa từ nguồn cung cấp đến điểm cuối cùng.

Tối ưu hóa tuyến đường đi trong chuỗi cung ứng nhằm tìm các tuyến đường và lịch trình vận chuyển tối ưu nhất để đáp ứng nhu cầu của khách hàng với chi phí tối thiểu. Điều này quan trọng đến việc xác định tối ưu hóa thời gian và chi phí trong quá trình quản lý.

Mô hình nhóm nghiên cứu đã phát triển và xây dựng lên với mục tiêu là dựa trên thuật toán kmeans để phân cụm và cân bằng các cụm với nhau để tổng khoảng cách di chuyển của các phương tiện trong doanh nghiệp không chênh lệch nhau quá nhiều. Ở mỗi cụm do một phương tiện phân phối mô hình áp dụng thuật toán Tabu search để giải quyết tìm một chu trình ngắn nhất sao cho phương tiện đó xuất tại điểm kho và điểm cuối cùng phương tiện đến là điểm xuất phát.

# Tài liệu tham khảo.

|  |  |
| --- | --- |
| [1] | J. H. R. G. B. Dantzig, "The Truck Dispatching Problem," *Management Science,* October 01, 1959. |
| [2] | J. G. C. a. M. Y. a. Z. M. Costa, "Cluster-based Hyper-Heuristic for Large-Scale Vehicle Routing Problem," *IEEE,* 03 September 2020. |
| [3] | A. E. Noha A. Mostafa, "Solving the Heterogeneous Capacitated Vehicle Routing Problem using K-Means Clustering and Valid Inequalities," *ResearchGate,* April 2017. |
| [4] | M. F. N. D. L Rizkallah, "A Clustering Algorithm for Solving the Vehicle Routing Assignment Problem in Polynomial Time," *International Journal of Engineering and Technology,* 2019. |
| [5] | H. G. H. Y. J. e. a. Zhang, "Review of Vehicle Routing Problems: Models, Classification and Solving Algorithms," *Arch Computat Methods Eng,* 19 April 2021. |
| [6] | J. L. Guoming Li, "An Improved Tabu Search Algorithm for the Stochastic Vehicle Routing Problem With Soft Time Windows," *IEEE,* vol. 8, pp. 158115 - 158124, 28 August 2020 . |
| [7] | A. Villa, "Introducing some Supply Chain Management problems," *International Journal of Production Economics,* vol. 73, pp. 1-4, 31 August 2001. |
| [8] | Y. a. M. W. a. X. R. a. S. Y. He, "A tabu search algorithm with variable cluster grouping for multi-depot vehicle routing problem," *IEEE,* 08 July 2014. |
| [9] | F. Glover, "Tabu Search: A Tutorial," *Interfaces,* vol. 20, pp. 74--94, 1990. |