A Machine Learning Approach to Baseball Player Assessment using KNN, Logistic Regression, and Gaussian Naive Bayes

OSMAN AKANAY CANBULAT[1], SAFIYE TURGAY[2], ESMA SEDEF KARA[3]

[1, 2]Department of Industrial Engineering
Sakarya University
54187, Esentepe Campus Serdivan-Sakarya
TURKEY
[3] Rüstempaşa Mahallesi, İpekyolu Caddesi, No:120,
Sapanca (54600) Sakarya,
TURKEY

Abstract: - This research employs a machine learning approach to assess and predict baseball player performance, utilizing three distinct algorithms: K Nearest Neighbors (KNN), Logistic Regression, and Gaussian Naive Bayes. The purpose of the study is to discover trends and insights for an end-to-end comprehension of player skills, which assists coaches, scouts, and team management to make well-informed decisions. Player dataset involves batting averages, overall game statistics and defensive approaches. Along with the data set applied in the study, the sports analytics process is also developed and the assessment of the baseball players being done. Moreover, this study leads to talent identification, strategy development in the game and planning. KNN is used to get player clusters, logistic regression used to make binary predictions and Gaussian Naïve Bayes approach used to get probability of occurrence.

Key-Words: - Machine Learning; Player Assessment; Logistic Regression; Gaussian Naïve Bayes; KNN

Received: April 6, 2024. Revised: September 13, 2024. Accepted: October 15, 2024. Available online: November 25, 2024.

1 Introduction

Machine learning methods aim at helping make smarter and more effective choices as well as to analyze the success of each baseball player during the game. Player evaluation, stats on batting averages and results in the game also are taken into account. The spatial reasons are ok for KNN, Logistic Regression suitable for binary tasks, and the probabilistic insights for Gaussian Naive Bayes are considered as the unique strengths which are needed for every algorithms of a comprehensive player assessment framework. Generated after the figure 1 shows the following research objectives.

The data source containing many baseball game records, we use Python capabilities for collecting, processing and mining this huge data set to a complementary format previously prepared for analysis. With Python libraries such as Pandas and NumPy, we are able to demonstrate the art of feature engineering. Along this path, we inch toward features that bring forth the game prediction's essence. However, the very cause of concern is in the predictive output itself. With Scikit-learn, TensorFlow, and XGBoost libraries as our trusted companions we start our journey, the next step is to enter a regression analysis, time series prediction, Pecota prediction, and the ensemble method. Based on such tools as thorough research and comparison we reveal the most impactful approach in modelling so as to present predictive analysis for baseball games' outcomes.

Going down the levels we admit the secret technique of feature importance analysis as well as correlation studies. Python libraries, such as Scikit-learn and Seaborn, can be called the ones steering our way, while they help us find which trend or factors influence baseball tilt one way or another. Struggle between a player statistic (s), team result (s), weather condition (s) and the match condition (s) start to come out onto the field; this reveals the most dinstinguible features which are regularities of the game. It is significant, so we check the prediction models we have designed. The historical data match set, which for the purpose of

development was separate from the learning file, is used for training. In this work, machine learning algorithms were used with the aid of the python-Scikit- learn library.

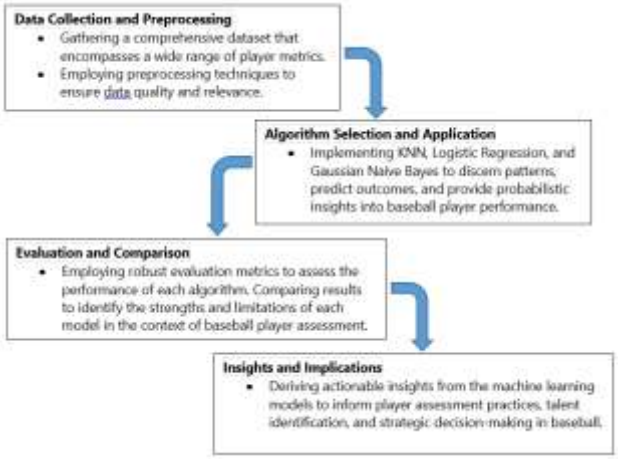

Figure 1. Main objectives

The remainder of the paper is structured as follows: section 2 presents a literature review of relevant milk collection issues. Section 3 gives detail about the machine learning methods. Section 4 shows the model and a sequential three-step solution approach. Finally, section 5 explains the conclusions.

2. Literature Survey

This paper will deal with the science behind predicting the outcomes of baseball games that will be aimed at offering an edge to the sports analysts, coaches and fans over the casinos. Python acts as a tool that can effectively carry out learned outputs like best team selection, strategic planning and even sports betting.

This study will be conducted by employing 3 machine learning algorithms from 889,234 records for 6737 different players (from 2002-2003 to 2021-2022) of the Major League Baseball (MLB). He describes using of data analytics as a device to dominate the competitors in the game. Though not specifically about machine learning, it, however, push forward statistics to be in area of player evaluation. While diverse reporting of the ML for player statistics can be given, such as is the case of machine learning on the player performance by researchers and the statistics comparison in the analysis of baseball games, the latter investigates the impact that machine learning has on the evaluation of players, rather than comparing the statistics. In addition, these studies were mostly focused on the modeling structure of the participant's performance by investigating the spatial distributions and patterns [7]. Some studies employed distinct machine learning techniques to include those factors. [8-14]. A case in point is our logistic regression that gives scores of the individual players themselves. Some of the studies introduce machine learning techniques along with their applications [15; 16]. One method that has been used to different studies in sports analytics approaches by Gaussian Naïve Bayes. It will provide you with a concise and introductory guide to using machine learning with Python if you are a beginner. It is best for people who are setting up a foundation and learning basics before moving on with practical applications in sports analytics.

An analysis of recent trends and progresses in baseball analytics science with a main emphasis on the growing usage of machine learning. This survey identifies unanswered questions from existing studies and provides a roadmap for future research studies on the intervention of player assessment in this area. However, data science in sports analytics is just one of the multi-faceted topics covered in the book and really shows its practical application. It provides for methodological assimilation from one sport to another which is baseball player assessment. The paper explores ethical concerns of KNN, Logistic Regression, and Gaussian Naive Bayes algorithms used for player rating and decision-making. Some of the researchers bring machine learning to practical applications of computer vision. While this technology is not baseball specific, it provides information on image identification applications, and this could be useful in players' motions and positions monitoring [21-23].

Comprehending the ethical aspects during the use of machine learning in the sports analytics. The literature review, in this case, discusses the ethical concerns associated with the use of algorithms such as KNN, Logistic Regression, and Gaussian Naive Bayes in player evaluation and decision-making. Some scientists concentrate on pratical applications of machine learning in computer vision is their objective. It may not be especial to baseball, however, it can teach about image recognition, which can be major in assessing athlete movements and positions [24,25].

The mentioned references are based on how sports analytics and machine learning provide a

relationship. The KNN method, logistic regression, and Gaussian Naïve Bayes methods were contemplated in the process of predicting the player's future performance in baseball. The next step is to introduce in methodology section of the study, the mathematical models of these three methods which are described in detail.

3. Methodology

Developing supervised machine learning solutions with Nearest Neighbors (KNN), Logistic Regression, and Gaussian Naive Bayes (GNB) approaches is a key aspect of the search for efficient processes and provides relevant solutions to different aspects of classification problems [26-27]. This exploration involves a foray into the techniques of KNN, Logistic Regression, and Gaussian Naive Bayes in which the idea behind their algorithmic variations to classification issues is further unfolded.

3.1. Logistic Regression

Logistic regression stands for the proper relationship between a collection of independent variables and a dependent variable that describes an underlying model. For logistic regression, there is the capability of dependent variable taking binary values like 0 and 1 in contrast with the limited nature of ordinary linear regression.

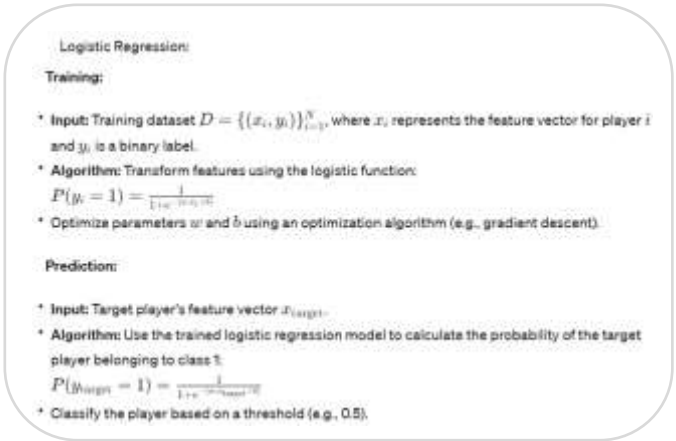

Figure 2. Logistic Regression Steps

This is perfect for helping with the understanding of event probabilities or for the difference between categories like success or survival. Unlike the linear regression, the logistics regression does not infringe the assumptions for the categorical variables as well as means with more than two categories. The functionality of logistic regression can be performed by effectively illustrating a graph as shown in Figure 2 [28].

$$a = \frac{\sum x_i \sum y_i - n \sum x_i y_i}{(\sum x_i)^2 - n \sum x_i{}^2} \quad (1)$$

3.2 K-Nearest Neighbor (KNN)

KNN stands for the K-Nearest Neighbor algorithm and it is among the most widely used machine learning algorithms in the problems related to classification and regression. The principle of operation for the k nearest neighbor algorithm is discovering the class for a new data point by considering the identities of majority class of the data points in the k closest neighboring of that point. By taking into account the same classes as the data points in the k neighborhoods, the new data point is assigned the most common category among those nearest neighbors. The new data point is usually considered in relation to how it differs from the other points and that can be done using distances like the Euclidean, Manhattan or Minkowski ones. Euclidean distance, in turn, reflects the straight-line distance between two objects, while the Manhattan distance or also called the "taxicab" distance is the sum of the horizontal and vertical distances on straight lines. While we consider Minkowski's distance as a generalization that takes into account the properties of both Euclidean and Manhattan distances. In the bottom: a mathematical model of K-nearest neighbors (KNN) that is demonstrated in Fig.3 [17].

$$\left(\sum_{i-1}^{n} |x_i - y_i|^p \right)^{\frac{1}{p}} \quad (2)$$

3.3 Gaussian Naive Bayes

Gaussion Naive Bayes is a kin variant of Naive Bayes and meets the features with the normal distribution assumption for classes. This may result in an inflation of the weights of the data points, situated near the center of masses of the classes. These allocations provide probabilities each feature's class label was estimated and further classifications for new instances are made. Bayes' theorem for dependent X and Y in Fig. 4 [16] is as shown:

$$P(y|X) = \frac{P(X|y).P(X)}{P(y)} \quad (3)$$

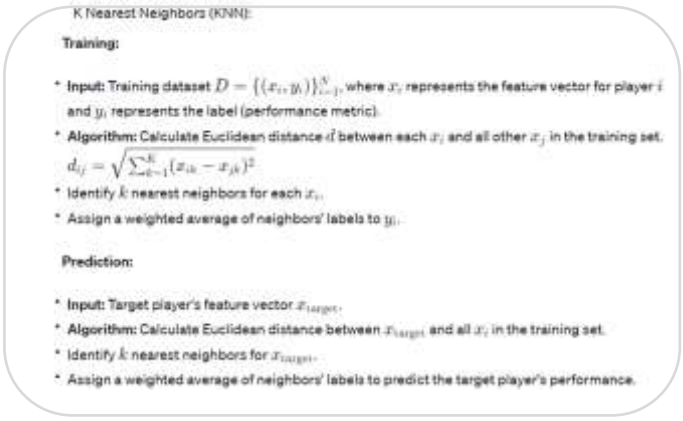

Figure 3. Mathematical Model of k-Nearest Neighbor (KNN)

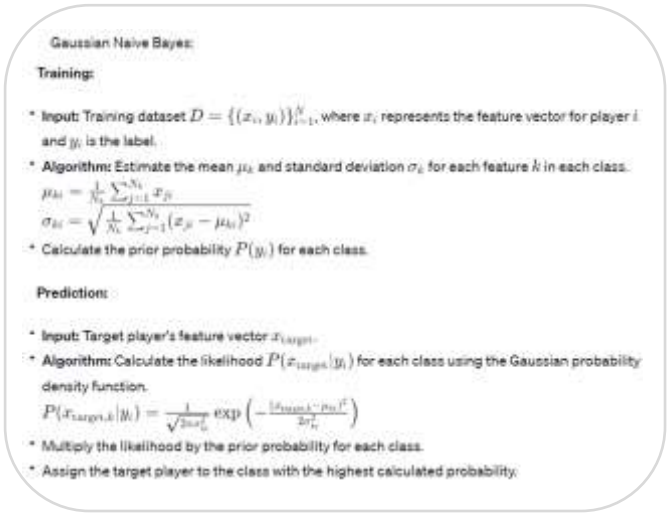

Figure 4: Gauss Naive Bayes Function

4. Case Study

Ever since the sports competitions have been introduced to the world, they have become one of the most to be watched programs around the planet. This is one of those many institutions that are responsible for gatherings ranging from few to thousands. At the moment, teams are established in the most part of the countries that are sometimes solely because of the interest. Leagues and the other contests are implemented because the fans are becoming more fanatic about the game Yet, baseball has transformed into more of a showbusiness than a sport. Likewise to the way run by business units, each baseball team become a separate entity functioning in the league. Not different from each other, every team independently manages its own finance, players turbulence task and marketing plan. Teams build revenue through the multi-tiered streams; namely, ticket sales, broadcast deals, merchandise business, sponsorships, and concessions. Just as commercial businesses hunt for good stuff, the ball games club like the baseball teams sign up star players in order to grow a competitive team. Player contracts are highly valued and act similar to employment contracts so that they often involve investment agreements, most of which are financially heavy in the anticipation of the performance returns. Contracts of Player are valued depending on several factors. In this report, we serves as the predictor of the players performance using data of 360,542 players of 6737 different players from 2002-2003 through to 2021-2022 in Major League Baseball (MLB) drawn from three different machine learning methods.

4.1 Obtaining the Data

In order to accomplish this goal, we utilize historical data about the MLB league, here www.fangraphs.com. The available data is herein provided; a number of matches played between the years 2002 and 2022 are featured. The data spreads several statistics that are of paramount importance to the teams' as well as the players' performance as a whole. Here are observed monitors like padawan walks, how many outs by a pitcher, stolen bases, batting average in game, homers, on-base percentage, on-base slugging percentage, unsuccessful at-bat, or how much a pinch-hitter have contributed.

4.2 Modelling the Data

The first step while analyzing the data is to decode the symbol "idfg" to get distinct ID for each player. The distinct identities are used to group players according to their performance. Hence, we separate players from the player's assigned group while dividing the data frame into groups based on player IDs. Having polished the data, sklearn.feature_selection is then employed to pick 20 key factors out of 132 features that are used to minimize inaccuracies the prediction processing step. Here the players identities are formed by choosing the predictors to set the attributes and this process helps to make the model better. Such a model assists to build a straight prediction approach.

4.3 Separation of the dataset into training and testing

It was attempted to see if the accuracy of the proposed approach is valid or not. Hence, both the datasets from Sklearn library module were classified into training and testing groups. With the exception of the match result variable, the rest of the dataset

are referred to as independent variables and for our dependent variable, we have the player's contribution to his team. During the process of training and testing DL models, the accurate models are selected from different algorithms based on the accuracy score. The train data set will be segmented into three separate groups of three-compound data sets. The system will use 296,414 data of the 6737 player data figured during the training process. This remaining data set is to the test data, a set consisting of the match data. The accuracy ratings for each algorithm were similarly achieved using a the testing set. This step is critically essential for judging the capacity of machine learning (ML) models to be able to generalize and adapt to new data.The training dataset is composed of 60% of random from the match and the testset of rests 40%. This technique randomly chooses data for training and testing which helps to represent the full range of the dataset and test the capability of the model when dealing with unseen data.

4.4 Training with Machine Algorithm and Finding Results

Figure 5a shows the logistic regression algorithm became successful 51% of the time by correctly predicting the results of the 10-player sample in different seasons. The logistic regression algorithm enabled 50% recall as depicted in Figure 5a. At 65% precision, the logistic regression algorithm was presented as shown in Fig.5. The K-Nearest Neighbor (KNN) regression model was successful with a rate of 56% on the correct identification of outcome.

Logistik Regresyon:

	IDfg	Season	Name	precision	recall	f1_score
0	2158	2008	Greg Dobbs	0.71	0.44	0.543304
1	1818	2007	Bobby Crosby	0.38	0.57	0.456000
2	13066	2021	Teoscar Hernandez	0.43	0.62	0.507810
3	1849	2014	Rickie Weeks Jr.	0.57	0.73	0.640154
4	14854	2021	Mike Yastrzemski	0.62	0.38	0.471200
5	4810	2007	Brian McCann	0.77	0.84	0.803478
6	5631	2010	Matt Kemp	0.82	0.28	0.417455
7	1875	2009	Josh Hamilton	0.39	0.72	0.505946
8	9166	2010	Buster Posey	0.52	0.19	0.278310
9	11579	2014	Bryce Harper	0.94	0.41	0.570963

Average Forecast Value (F1 Score): 0.536405
Average Recall Value 0.506
Average Precision Value 0.651

Figure 5: Logistic Regression Classification Report

As indicated in Fig. 5b, the K-Nearest Neighbor (KNN) algorithm achieved a recall rate of 54%. The KNN (K-Nearest Neighbor) algorithm secured 64% precision as indicated in Figure 6. The classifier Gaussian Naive Bayes algorithm exhibited 52% recall value as showed in Figure 7. The Gaussian Naive Bayes algorithm gave 45% precision as observed in the Fig 5c.

K-En Yakın Komşu (KNN):

	IDfg	Season	Name	precision	recall	f1_score
0	2158	2008	Greg Dobbs	0.57	0.52	0.543853
1	1818	2007	Bobby Crosby	0.40	0.61	0.483168
2	13066	2021	Teoscar Hernandez	0.59	0.58	0.584957
3	1849	2014	Rickie Weeks Jr.	0.56	0.64	0.597333
4	14854	2021	Mike Yastrzemski	0.61	0.77	0.680725
5	4810	2007	Brian McCann	0.73	0.62	0.670519
6	5631	2010	Matt Kemp	0.83	0.37	0.511833
7	1875	2009	Josh Hamilton	0.72	0.63	0.672000
8	9166	2010	Buster Posey	0.51	0.32	0.393253
9	11579	2014	Bryce Harper	0.90	0.37	0.524409

Average Forecast Value (F1 Score): 0.566236
Average Recall Value 0.5431
Average Precision Value 0.642

Figure 6: K-Nearest Neighbor (KNN) Classification Report

Gauss Naive Bayes:

	IDfg	Season	Name	precision	recall	f1_score
0	2158	2008	Greg Dobbs	0.38	0.41	0.394430
1	1818	2007	Bobby Crosby	0.21	0.60	0.311111
2	13066	2021	Teoscar Hernandez	0.39	0.62	0.478812
3	1849	2014	Rickie Weeks Jr.	0.37	0.60	0.457732
4	14854	2021	Mike Yastrzemski	0.42	0.54	0.472500
5	4810	2007	Brian McCann	0.54	0.78	0.638182
6	5631	2010	Matt Kemp	0.64	0.34	0.444082
7	1875	2009	Josh Hamilton	0.53	0.68	0.595702
8	9166	2010	Buster Posey	0.32	0.28	0.298667
9	11579	2014	Bryce Harper	0.74	0.39	0.510796

Average Forecast Value (F1 Score): 0.459467
Average Recall Value 0.524
Average Precision Value 0.451

Figure 7: Gaussian Naive Bayes Classification Report

4.5. Comparison of Results

In this section, the outcomes of each algorithm used for baseball player analysis and anticipation is compared with accuracy and verification scores (Table 1). These scores will be used to evaluate the performance of the algorithms and also see which algorithm is the best in terms of success.

Algorithms	Accuracy Scores
Logistic Regression	%51
K-Nearest Neighbour (KNN)	%56
Gaussian Naive Bayes	%46

Table 1: Algorithms and Accuracy Scores

After the algorithms (K Nearest Neighbor - KNN, Logistic Regression, and Gaussian Naive Bayes) implementation and evaluation, this figure (Figure 8-10) shows the results.

Anova

Summary

Groups	Number	Sum	Average	Variance
Lojistik	10	5,19462	0,519462	0,019299
KNN	10	5,66205	0,566205	0,008724
Gauss	10	4,602014	0,460201	0,011779

ANOVA

Variance Resource	SS	df	MS	F	P-Value	F-Scale
Between Groups	0,05644	2	0,02822	2,12723	0,13871	3,35413
Inner Groups	0,35822	27	0,01327			
Total	0,41466	29				

Figure 8: Anova Test Report

	Lojistik	KNN
Average	0,519462	0,566205
Variance	0,019299	0,008724
Observation	10	10
df	9	9
F	2,212019	
P(F<=f) tek	0,126305	
F-Critic two sided	3,178893	

	KNN	Gauss
Average	0,566205	0,460201
Variance	0,008724	0,011779
Observation	10	10
df	9	9
F	0,740696	
P(F<=f) tek	0,331004	
F-Critic two sided	0.314575	

	Lojistik	Gaus
Average	0,519462	0,460
Variance	0,019299	0,011
Observation	10	
df	9	
F	1,638435	
P(F<=f) tek	0,236731	
F-Critic two sided	3,178893	

Figure 9: F Test Report

Figure 8: T Test Report

5. Conclusion

Although, player analysis is one of the important subject matter that is studied in machine learning. The comfort of grabbing historical data has encouraged the building up of player performance prediction studies to a considerable extent. In other words, baseball competitions are more likely to be the type of sports where such studies happen. Here, different machine-learning algorithms are compared, and the best model is chosen as well as best fitting algorithm to predict educational performance.

The primary aim of this research is to come up with a technique to forecast the level of baseball competition that is based on organized analysis of player data using machine learning methodologies. In this approach, a new venue is created to showcase a viewpoint not encountered in baseball business, which is done through evaluating player data by various algorithms. It is suspected that such assesments based on the results of sports games or players' performance are better to predict such results and creation of new possibilities in the betting industry take place. Concerning the future sustainability of this study, it is perceived that this modus of conduction of data collection and updation can be adopted as a sustainable method. Automated structure being in place will allow the system work perpetually across that specific time period. Indeed, the solution does have drawbacks and querying more data along with additional analysis and features would be very helpful. This research draws on data from all professional baseball athletes of the past. Nevertheless, it is speculated that the prediction success can be enhanced by passing over cross-references into the dataset. At the same time, it must be taken into account that factors like playOef the players, weather causes and fans' influence, team structure can also influence results and moreover these factors are relationships. The hypothesis is that including these exogenous elements in the dataset will be the cause of the increase in the correct rates of prediction. In addition to that is the present study design useful in any other sports, the set of data may be also assessed and recommended for the better quality.

To get higher accuracy while doing player analysis, real player data needs to be simplified and effect of unforeseen characteristics needs to be reduced. This framework is designed to make forecasts more reliable and precise. This feature, which depicts the freshness and relevancy of the database, is perceived as an improvement factor. In the forthcoming, these things will make emotional responses in the sector. In situations where the chances of a player-based prediction are high, people tend to lose interest in such prediction too, and consequently its enjoyment also reduces. Additionally, these adjustments will also affect the betting business with economic implications.

References:

[1] Grzegorz Bocewicz, G., Nielsen, P., Zbigniew, B., Milk-run routing and scheduling subject to different pick-up/delivery profiles and congestion-avoidance constraints, 10th IFAC Symposium on Intelligent Autonomous Vehicles, Gdansk, Poland, July 3-5, 2019.

[2] Hormes F, Siala A, Lieb C, Fottner J (2020). Fleet Sizing of Dynamically Routed In-plant Milk-run Vehicles Based on a Genetic Algorithm. Logistics Journal: Proceedings, Vol. 2020. (urn:nbn:de:0009-14-51420)

[3] Bocewicz G., Nielsen I., Zbigniew B., A decision support model for prototyping in-plant milk-run traffic systems, (2019) IFAC-PapersOnLine, 52 (13) , pp. 814-819.

[4] Bocewicz, G., Banaszak, Z.A., Rudnik, K., Witczak, M., Smutnicki, C., Wikarek, J.,Milk-run Routing and Scheduling Subject to Fuzzy Pickup and Delivery Time Constraints: An Ordered Fuzzy Numbers Approach,2020 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)}, 2020,pp.1-10

[5] Kilic HS, Durmusoglu MB. "A mathematical model and a heuristic approach for periodic material delivery in lean production environment". The International Journal of Advanced Manufacturing Technology, 69(5-8), 977-92, 2013.

[6] Kilic HS, Durmusoglu MB, Baskak M. "Classification and modeling for in-plant milk-run distribution systems". The International Journal of Advanced Manufacturing Technology, 62(9-12), 1135-1146, 2012.

[7] Gotthardt, S., Hulla, M., Eder, M., Karre, H., Ramsauer, C., Digitalized milk-run system for a learning factory assembly line, Procedia Manufacturing, Volume 31, 2019, Pages 175-179, ISSN 2351-9789, https://doi.org/10.1016/j.promfg.2019.03.028.

[8] De Moura D.A., Botter R.C. (2016) Delivery and pick-up problem transportation – milk run or conventional systems. Independent Journal of Management & Production (IJM&P), vol. 7 (3), 746-770.

[9] Meyer A. (2015), Milk Run Design (Definitions, Concepts and Solution Approaches), PhD thesis, Institute of Technology. Fakultät für Maschinenbau, KIT Scientific Publishing, Karlsruhe.

[10] Mei, H., Jingshuai, Y., Teng, M., Xiuli, L., Ting, W., The Modeling of Milk-run Vehicle Routing Problem Based on Improved C-W Algorithm that Joined Time Window, Transportation Research Procedia, Volume 25, 2017, Pages 716-728, ISSN 2352-1465, https://doi.org/10.1016/j.trpro.2017.05.453.

[11] Lou, Z., Li, Z., Luo, L., Dai, X., Study on Multi-Depot Collaborative Transportation Problem of Milk-Run Pattern, MATEC Web Conf., 81 (2016) 01004

[12] Sipahioğlu, A., Altın, İ., A mathematical model for in-plant Milk-Run routing , Journal of Pamukkale Univ Engineering Science, 25(9), 1050-1055, 2019.

[13] Adriano DD, Montez C, Novaes AGN, Wangham M. DMRVR: Dynamic Milk-Run Vehicle Routing Solution Using Fog-Based Vehicular Ad Hoc Networks. Electronics. 2020; 9(12):2010. https://doi.org/10.3390/electronics9122010

[14] Vilda, F.G., Yague-Fabra,J.A., Torrents, A.S., An in-plant milk-run design method for improving surface occupation and optimizing mizusumashi work time, CIRP Annals - Manufacturing Technology 69 (2020) 405_408.

[15] Aragao, D.P., Antonio Galvao, A., Novaes, N., Lunac, M.M.M, An agent-based approach to evaluate collaborative strategies in milk-run OEM operations, Computers & Industrial Engineering, Computers & Industrial Engineering 129 (2019) 545–55.

[16] Klenk, E., Galka,S., Günthner, W.A., Operating Strategies for In-Plant Milk-Run Systems, IFAC-PapersOnLine 48-3 (2015) 1882–1887.

[17] Sahar, G.S., Okba, K., Abdelkader, L., Amine, Y.M., Euler, R., Bounceur, A., Hammoudeh, M., An Optimized Scalable Multi-ant Colony System for Multi-depot Vehicle Routing Problems Using a Reactive Multi-agent System, WSEAS Transactions on Systems, vol. 20, pp. 249-259, 2021.

[18] Turgay, S., Streamlined Supply Chain Operations: Leveraging Permutation-Based Genetic Algorithms for Production and Distribution, WSEAS Transactions on Information Science and Applications, Vol. 21, 2024, pp. 23-32, ISSN: 1790-0832.

[19] Turgay, S., Gujrati, R., Analyzing of B2B and B2C in Buying Centers using Rough Sets and MCDM Respect of Sustainable Supply Chain Management, Available at SSRN: https://ssrn.com/abstract=4295742 or http://dx.doi.org/10.2139/ssrn.4295742.

[20]ParedesBelmar, G., Marianov, V.,Bronfman, A., Obreque, C. and Lüer-villagra, A. (2016), "A milk collection problem with blending", Computers and Electronics in

Agriculture, Vol. 94, pp. 26-43, doi: 10.1016/j.tre.2016.07.006.

[21] Paredes-Belmar, G. Montero, E., Leonardini, O., A milk transportation problem with milk collection centers and vehicle routing, ISA Transactions, Volume 122, 2022, Pages 294-311, ISSN 0019-0578, https://doi.org/10.1016/j.isatra.2021.04.020

[22] Hosseini, S.D., Shirazi, M.A:, Karimi, B., Cross-docking and milk run logistics in a consolidation network: A hybrid of harmony search and simulated annealing approach, Journal of Manufacturing Systems,Volume 33, Issue 4, 2014, Pages 567-577, ISSN 0278-6125, https://doi.org/10.1016/j.jmsy.2014.05.004.

[23] Vahdani, B., Tavakkoli-Moghaddam, R., Zandieh, M., Razmi, J., Vehicle routing scheduling using an enhanced hybrid optimization approach, J Intell Manuf. 2012) 2 759–774, doi. 10.1007/s10845-010-0427.

[24] Musa, R., Arnaout, J.P., Jung, H., Ant colony optimiza tion algorithm to solve for the transportation problem of cross-docking network, Computers & Industrial Engineering, Volume 59, Issue 1, 2010, Pages 85-92.

[25] Goodarzi A, H., Zegordi S.H. (2016) A location-routing problem for cross-docking networks: a biogeography-based optimization algorithm. Comput Ind Eng 102:132–146.

[26] Urru, A., Bonini, M., & Echelmeyer, W. (2018). Planning and dimensioning of a milk-run transportation system considering the actual line consumption. IFAC-PapersOnLine, 51, 404-409.

[27] Klenk, E., Galka,S., Analysis of real-time tour building and scheduling strategies for in-plant milk-run systems with volatile transportation demand,IFAC-PapersOnLine, Volume 52, Issue 13, 2019, Pages 2110-2115.

[28] You, Z., Jiao, Y., Development and Application of Milk-Run Distribution Systems in the Express Industry Based on Saving Algorithm", Mathematical Problems in Engineering, vol. 2014, Article ID 536459, 6 pages, 2014

[29] Mao, Z., Huang, D., Fang, K., Wang, C., & Lu, D. (2019), Milk-run routing problem with progress-lane in the collection of automobile parts. Annals of Operations Research 26, 1-28.

[30] Turgay, S., Yaşar, Ö., Aydın, A., A Multi-objective Framework for Dairy Products Supply Chain Network with Benders Decomposition, Industrial Engineering and Innovation Management (2023), Vol. 6 Num. 5, DOI: 10.23977/ieim.2023.060509 ISSN 2522-6924.

[31] Meyer, A. (2017). Milk run design: Definitions, concepts and solution approaches (Vol. 88). KIT Scientific Publishing.

[32] Hižak, J.; Logožar, R., An Overview Of The Genetic Algorithm And Its Use For Finding Extrema ─ With Implementations In Matlab, Tehnički glasnik 10, 3-4(2016), 55-70.

[33] Turgay, S., Aydın, A., An Effective Heuristic Algorithm for Flexible Flow Shop Scheduling Problems with Parallel Batch Processing, Manufacturing and Service Operations Management (2023), ISSN 2616-3349 Vol. 4 Num. 1.,), DOI: 10.23977/msom.2023.040109.

[34] Meilinda F. Maghfiroh, N. , Darmawan, A., Yu, V.F., Genetic Algorithm for Job Shop Scheduling Problem: A Case Study, International Journal of Innovation, Management and Technology, Vol. 4, No. 1, February 2013.

[35] Süer, G.A., Yang, X., Alhawari, O.I., Santos, J., Vazquez, R., A Genetic Algorithm Approach for Minimizing Total Tardiness in Single Machine Scheduling, International Journal of Industrial Engineering and Management (IJIEM), Vol. 3 No 3, 2012, pp. 163-171.

[36] Rardin, RL (2000) Optimization in Operations Research, Prentive Hall, April

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

The authors equally contributed in the present research, at all stages from the formulation of the problem to the final findings and solution.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

No funding was received for conducting this study.

Conflict of Interest

The authors have no conflicts of interest to declare that are relevant to the content of this article.