İSTANBUL UNIVERSITY
CERRAHPAŞA

ORIGINAL ARTICLE
Electrica 2024; 24(1): 238-246

Estimation of Future Number of Passes and Optimization of Number of Trips Based on Istanbul Hourly Public Transportation Data

Emel Geçici, **Zeynep Gürkaş-Aydin**

Department of Computer Engineering, İstanbul University-Cerrahpaşa Faculty of Engineering, Istanbul, Turkey

Cite this article as: E. Geçici and Z. Gürkaş-Aydin, Estimation of future number of passes and optimization of number of trips based on Istanbul hourly public transportation data, *Electrica*, 24(1), 238-246, 2024.

ABSTRACT

Considering the significance of transportation in everyday life, issues related to transportation are commonly encountered in developing and developed countries. The problems encountered in most countries are likewise evident in our country and manifest throughout several domains. Factors contributing to the issue include excessive passengers, inadequate trip availability to fulfill demand, ineffective line and route organization in addressing passenger requests, incomplete routes, and insufficient flight frequency. Insufficient service availability during peak hours and a preference for private automobiles are additional examples that can be mentioned. These negative factors reduce passenger happiness, lead to disturbances inside the company, and have a significant financial impact. Considering all these aspects, attempts were made to predict the future number of crossings using various forecasting algorithms based on the hourly public transportation dataset to achieve the lowest error rate. Based on the obtained results, an assessment was conducted on the areas with insufficient or missing flight frequencies, and recommendations were provided to enhance the number of flights in these regions, considering peak hours.

Index Terms— Artificial intelligence, frequency of trips, optimization, passenger demand in transportation, transportation intensity

I. INTRODUCTION

Istanbul boasts the most significant population in Turkey, primarily due to its abundant alternatives. Remarkably, around one-fifth of the entire country's population resides in this live metropolis. Due to the numerous opportunities and possibilities, the population steadily grows, resulting in a greater demand for particular necessities. Transportation is a fundamental necessity. Transportation refers to the movement of people and communities between different locations. It can be categorized into four main types: road, sea, airline, and train. Urban transportation predominantly favors highway transportation. Railway transportation options such as metro, tram, and sea transportation may also be favored based on the demand and physical location of the region. Urban transportation tends to avoid air transportation due to its high expense. Despite the abundance of transportation options, the overpopulation and inadequacy of vehicles compel many to rely on private vehicles. Individual car usage contributes to the occurrence of traffic congestion. Hence, despite the drawbacks and problems, passengers prefer public transit. The surge in demand for public transportation indicates that available resources will be insufficient. Therefore, it is imperative to take appropriate measures. This study aims to estimate the future number of passengers utilizing public transportation and the number of trips based on İstanbul hourly public transportation dataset. This estimation will be based on the hourly number of trips supplied by Istanbul Metropolitan Municipality [1]. Machine learning techniques such as regression, decision trees, and random forests, will be employed. Utilizing this dataset, various forecasting approaches were employed to estimate the future prediction of the number of passes and achieve the outcome with the lowest error rate. Based on the results, suggestions were proposed to augment the number of flights in these areas, considering the locations where flight frequency falls short of passenger demand and during peak periods. Subsequently, the clustering procedure for the districts and the identification of densely populated areas will be conducted based on the count of schools, hospitals, residences, public bread buffets, and vehicles in each district.

The remaining sections of the paper are organized as follows: The following section provides a comprehensive summary of relevant literature, along with a taxonomy table. Section III provides

Corresponding author:
Zeynep Gürkaş-Aydin

E-mail:
zeynepg@iuc.edu.tr

Received: December 20, 2023

Accepted: December 24, 2023

Publication Date: January 31, 2024

DOI: 10.5152/electrica.2024.23197

a concise overview of the approach and thoroughly describes the utilized dataset. The results are outlined in Section 4, while the study concludes in Section V.

II. RELATED WORK

Transportation covers all transportation systems, which consist of all types of vehicles. Commonly, transportation-related issues lead passengers to opt for public transit. Inadequate or unusable existing resources necessitate the development of studies on the subject. The research mostly focuses on flight frequency, route optimization, estimating flight frequency, and identifying new routes. Yaylalı and Dilek analyzed the need for air travel in Erzurum. They suggested that there could be airline companies interested in establishing new routes to this region if there is sufficient demand [2]. Ceylan and Özcan studied the frequency of trips in urban bus transportation utilizing Harmony research optimization [3]. The Harmony research is a simulation model that operates on two levels and is designed specifically for the purpose of optimization. Özcan, however, did a study on optimizing the frequency of metropolitan public transit networks [4]. A sample study showcased the methodology's efficacy on the Çorlu bus network. Demirkollu researched to ascertain the number of trips and provided recommendations for the frequencies of these trips [5]. He analyzed the bus frequency in Denizli province and proposed that the bus schedules should be determined based on the passengers' demands during the peak hours of bus usage throughout the day. Deri utilized smart card data in his study [6]. Utilizing this dataset, he analyzed the frequency of public transportation trips in Izmir. Uludağ is the author of another study on transportation in Izmir [7]. Uludağ utilized fuzzy optimization techniques to develop a model for bus line modeling. As a result, Uludağ identified two particular stops in Izmir's Lausanne and Montreux regions. Gencer and his colleagues studied the Ankara metro line [8]. They conducted an analysis of the Kızılay-Batıkent line and developed projections based on an assessment of past passenger figures. In addition to the studies conducted on calculating aircraft frequencies, research has been undertaken to develop alternate passenger routes. Engin conducted a study where he implemented his selected methods on public transportation in Çanakkale, developed novel models, and evaluated their performance [9]. Sağbaş and Polat performed a study where they optimized bus scheduling for urban transportation by considering passenger demands. They employed the linear goal programming method [10] for this purpose. Sağbaş and Polat published a study where they developed a linear goal programming model specifically for the public transportation systems of Tekirdağ/Çorlu Municipality [11]. Palamutçuoğlu's main objective in his study was to calculate the landing points using the boarding points [12]. Nar studied passenger demand on the Yenikapı M1 - Kirazlı M1 route, considering the rapid population growth and increasing passenger demands [13]. Kızıltaş and Akkaya studied the correlation between meeting demand and the supply–demand relationship [14]. According to Erpik's analysis, urbanization is on the rise, necessitating improvements in public transportation methods to accommodate the growing population [15]. Korkmaz and his colleagues performed research on predicting passenger numbers within short time intervals for public transportation in the city [16]. While executing their tasks, they utilized the Python programming language and the Spyder interface. In his study, Şaşmaz identified transportation as a crucial issue in Istanbul, highlighting that the growing population and the rising number of vehicles in traffic had a detrimental impact on this problem [17]. In addition to road and railway transportation, air transportation has been the focus of significant research. Efendigil and Eminler did a study on predicting future demand in the field of air transportation [18]. The authors indicated that the artificial neural networks yielded the most reliable outcomes compared to Adaptive Neuro-Fuzzy Inference System (ANFIS) and other approaches. The railway is also the focus of comparable research. Çakır performed a study on railways, analyzing the demand for passenger transit. He utilized regression analysis and artificial neural networks to generate a model for this demand [19]. Yaşar examined the issue of rising demand in urban transportation in his work [20]. He performed demand forecasting in his study by utilizing demand analysis and employing novel mathematical techniques. Furthermore, transportation-related issues are the focus of research conducted abroad. Carmona-Benítez et al. studied passenger transportation and estimated the number of passengers at the airport [21]. In a separate study conducted by Carmona-Benítez et al., they put forth an Econometric Dynamic Model as a means to calculate the passenger count in airplane transportation [22]. Jin and his colleagues employed many techniques to accurately assess the intricate circumstances in passenger forecasting inside the transportation sector [23]. In his work, Kim emphasized the crucial role of passenger estimation when determining airport operations [24]. Gunter and his colleague carried out a study on the estimation of passenger numbers in airplane transportation [25]. According to Suryani et al. 's study [26], it is possible to estimate the number of passengers and demand through various studies, allowing for the appropriate arrangement of runway and terminal capacities. Estimating passenger numbers in airline planning has been the research focus in numerous countries. The estimation of domestic passenger demand in Nigeria was the subject of a study by Adeniran et al. [27]. Gelhausen and his colleagues developed and executed a four-stage approach [28]. This method was employed to evaluate the impact of Brexit on the volume of German airport traffic from 2016 to 2018. Srisaeng and his colleagues employed a genetic algorithm and completed experimental testing [29]. A total of 74 data sets were utilized to evaluate the effectiveness of the genetic algorithm. Scarpel conducted a study on the estimation of passenger numbers at Sao Paulo International Airport [30]. He employed a combination of models developed by local experts in his research and affirmed that the predictions generated from the employed methodology were considered satisfactory. In their investigations, Koç and Arslan discussed the exponential growth of airplane transportation and the highly competitive landscape [31]. The researchers employed an artificial neural network model to analyze domestic air transportation in Turkey. Hsiao and Hansen conducted a study on the regulation of travel prices [32]. Anvari and his colleagues studied transportation demand in rail systems. They utilized the Box-Jenkins technique [33] to make a forecast. Jafari, in his research, discussed the profound influence of coronavirus disease 2019 (COVID-19) and the pressing need it has created. He emphasized the need to accurately predict passenger demand [34]. Plakandaras and his colleagues studied the prediction of air, road, and train transportation in domestic routes inside the USA [35]. Marie-Sainte et al. addressed the matter of ascertaining passenger demand for planning purposes in their study [36]. Zhang and Wang developed a stochastic frontier analysis model that takes into account model uncertainty in their investigation. A wide range of alternatives for this model have been identified [37]. In their study [38], Zhao and Mi introduced a novel hybrid model for high-speed train transportation. Cyril and his colleagues discussed demand forecasting to facilitate planning [39]. Xiao et al. introduced a method that integrates Singular Spectrum Analysis, ANFIS, and enhanced particle swarm optimization [40]. Srisaeng and his colleagues utilized an artificial neural network in their study on Australia, marking the first instance of its application in this context, despite the scarcity of examples of artificial neural networks in their research [41]. In their study [42], Alekseev and his colleague discussed the issue of reliable forecasting techniques. Terekhov et al. determined this by conducting a comparative analysis of modeled and verified data from 3 distinct years [43]. Consequently, they stated the need to establish a source index.

Upon examining the existing literature, it was noted that previous research on public transportation in Istanbul did not utilize both optimization and forecasting approaches simultaneously, despite their usage in other countries. This represents the novel element of the research in the existing body of research. Table I provides a concise overview of the research conducted on the topic as documented in the literature on the subject.

III. METHODOLOGY

This study aims to predict the future number of passengers and trips for a public transportation system. The region of application for the research was selected as Istanbul. The dataset utilized was the "Hourly Public Transportation Data Set" [1] available on the "Open Data Portal" provided by Istanbul Metropolitan Municipality. A number of machine learning methodologies were implemented, comprising elastic-net regression, linear regression, ridge regression, lasso regression, decision trees, and artificial neural networks. Using this dataset, different forecasting methods were applied to anticipate the future number of passes, aiming to minimize the error rate. The clustering process for the districts and the determination of densely inhabited places will be carried out using the number of schools, hospitals, dwellings, public bread buffets, and cars in each district. In areas with heavy traffic, including that of public and private transportation, this study aims to forecast the pattern of demand and provide recommendations for its improvement.

Istanbul province comprises a total of 954 neighborhoods and 39 districts, with 14 located on the Anatolian side and 25 on the European side. For the purpose of this study, districts will serve as the designated demand points. First of all, the districts of Istanbul were clustered according to the determined criteria, and it was ensured that the districts with a similar structure or characteristics were brought together. The clustering technique in this study involves the utilization of relevant observation values, which specifically pertain to districts. The primary variables that can serve as indicators of Istanbul's character and be utilized for clustering are the number of hospitals, number of rail stations in the districts, housing sales rate in the districts, number of schools, number of literates by district, and number of cars. The initial function for k-means started from a random starting point. The number of clusters is expressed as [1, n) iterations (n, which is in the definition range and shown as the upper value, corresponds to the total number of districts in our data set). The parameter required to run the algorithm using different centers was 10. The maximum number of iterations performed in a single run was determined to be 300.

Considering these criteria, an elbow chart was generated to determine the number of clusters. The graphs were created using the Python programming language and the k-means technique from the Sklearn package. Subsequently, three distinct iterations were conducted using European side districts, Anatolian side districts, and all districts. The resulting elbow charts for each iteration are displayed in Fig. 1a–c, respectively. Upon analyzing the resulting graphs, it becomes evident that the most significant transformation occurred in forming the initial and subsequent clusters. The turning point in the graph occurs when five clusters are included in the elbow chart for the European side, three clusters for the Anatolian side, and nine clusters when considering all districts.

Once the number of clusters has been determined, the districts of Istanbul are divided into clusters, as seen in Table II, and the categorized districts are presented. It is essential to mention that the clusters are positioned randomly.

A. Dataset

Once the clustering procedure was finished, a forecasting process was conducted to determine the future demands of the demand points. The transportation data was obtained by utilizing the "Hourly Public Transportation Data Set" [1] provided on the "Open Data Portal" by Istanbul Metropolitan Municipality. This dataset offers hourly statistics on the number of passengers and trips in Istanbul. The data, including monthly data, has been consistently shared since January 2020. The dataset comprises characteristics such as date, time, mode of transportation, route, travel line, whether the passengers are transfer passengers, the number of passes for that line, and the count of passengers. Table III details the categories of variables.

To accurately predict the future, the analysis incorporated data from 2022, when all the necessary information was fully accessible. Upon examination of the data, it is obvious that there are a total of 7,866,727 rows of data. This pertains to all journeys undertaken in the year 2022. Fig. 2 displays a selection of observed values from this dataset.

Upon analyzing the data of the Transportation Line, it reveals that a grand total of 939 journeys are conducted throughout all lines. After examining the observed values in the dataset, it is clear that the lines "ALTUNIZADE-SULTANBEYLİ, USKUDAR-BEYKOZ, USKUDAR-CEKMEKOY" are the most frequently used when classifying the transportation line based on repetition, or in other words, when determining the mode of transportation. The Altunizade-Sultanbey li, Üsküdar-Beykoz, and Üsküdar-Çekmeköy lines were chosen as the areas for conducting forecasting of demand due to this justification. Initially, the dataset was filtered to generate sub-datasets that contained the necessary lines. Subsequently, appropriate modifications were made to the dataset to facilitate analysis. The data presented in text format, as shown in Fig. 1, was transformed into categorical data, as depicted in Fig. 3.

After this alteration, the designations of "Sea," "Rail," and "Highway" under the Route Name category were modified to 1, 2, and 3, respectively. The values for "Transfer" and "Normal" under the label "Transfer Type" have been modified to 1 and 2, respectively. Due to additional filtering, the variable "Transportation Line" was excluded from the dataset as it consistently represents the same line. Hence, the dataset, comprising 7 variables and a total of 16465 data lines, was prepared for analysis. Fig. 4 displays the descriptive statistics for this data set.

To make predictions using the dataset, it is necessary to categorize the variables into two groups: dependent and independent.

TABLE I. SUMMARY OF THE RELATED WORK

			Method		Transportation Type			
Author(s)	**Year**	**Country/City**	**Optimization**	**Prediction**	**Air**	**Highway**	**Seaway**	**Railway**
Alekseev and Seixas	2002	Brazil		X	X			
Yaşar	2009	Turkey		X		X		
Yaylalı and Dilek	2009	Turkey/Erzurum		X	X			
Ceylan and Özcan	2018	Turkey	X			X		
Demirkollu	2017	Turkey	X			X		
Sağbaş and Polat	2022	Turkey	X			X		
Efendigil and Eminler	2017	Turkey		X	X			
Adeniran, Kanyio and Owoeye	2018	Nigeria		X	X			
Marie-Sainte, Saba and Alotaibi	2019		X	X	X			
Cyril, Mulangi and George	2018			X		x		
Suryani, Chou and Chen	2010			X	X			
Uludağ	2010	Turkey/İzmir	X			X		
Hsiao and Hansen	2011			X	X			
Deri	2012	Turkey/İzmir	X	X		X		
Engin	2013	Turkey/Çanakkale	X			X		
Scarpel	2013	Brazil/Sao Paulo		X	X			
Xiao, Liu, Hu, Wang, Lai and Wang	2014	Hong Kong		X	X			
Anvari, Tuna, Canci, and Türkay	2016	Turkey/Istanbul		X				X
Carmona-Benítez and Nieto-Delfín	2015			X	X			
Srisaeng, Baxter, Richardson, and Wild	2015	Australia		X	X			
Erpik	2019	Turkey/Istanbul		X		X		
Plakandaras, Papadimitriou and Gogas	2019	USA		X		X		X
Şaşmaz	2019	Turkey/Istanbul	X			X		
Zhang, Zheng and Wang	2019			X	X			
Zhao and Mi	2019			X				X
Çakır	2020	Turkey		X				X
Jin, Li, Sun and Li	2020	China		X	X			
Palamutçuoğlu	2020	Turkey	X	X		X		
Gunter and Zekan	2021	Pacific Asia-America Caribbean		X	X			
Jafari	2022	USA		X	X			
Kızıltaş and Akkaya	2022	Turkey	X			X		
Korkmaz, İlyas and Efe	2022	Turkey/Bursa		X		X		

(*CONTINUED*)

TABLE I. SUMMARY OF THE RELATED WORK (*CONTINUED*)

			Method		Transportation Type			
Author(s)	**Year**	**Country/City**	**Optimization**	**Prediction**	**Air**	**Highway**	**Seaway**	**Railway**
Nar	2022	Turkey/Istanbul		X				X
Polat, Sağbaş and Dermenci	2022	Turkey/Tekirdağ	X			X		
Kim	2016			X	X			
Terekhov, Evans and Gollnick	2016	Global		X	X			
Carmona-Benítez, Nieto and Miranda	2017	Mexico		X	X			
Gelhausen, Berster and Wilken	2018	Germany		X	X			
Gencer, Alakaş, Eren, and Hamurcu	2018	Turkey/Ankara		X				X
Koç and Arslan	2018	Turkey		X	X			
Özcan	2018	Turkey	X			X		
Srisaeng, Baxter and Wild	2015	Australia		X	X			

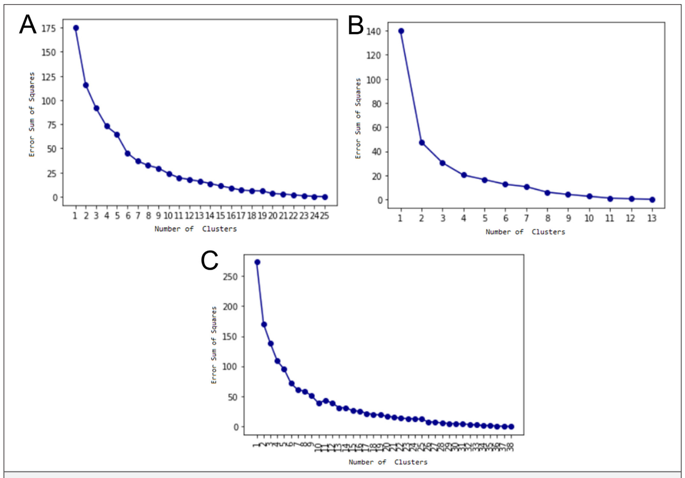

Fig. 1. Elbow charts for clusters: (a) European side. (b) Anatolian side. (c) All districts.

The dependent variable in this study is the "Number of Passengers," whereas the other variables are considered independent variables. Data scaling is required following the conclusion of this procedure.

TABLE III. VARIABLES IN ISTANBUL HOURLY PUBLIC TRANSPORTATION DATA

History	**Route Name**
Hour	Transportation Line
Transportation Type (1 = Highway, 2 = Rail, 3 = Sea)	Transfer Type (Normal, Transfer)
Number of Trips	Number of Passengers

This approach aims to mitigate the dominance of variables over one another. To accomplish this, the "Min–Max Scaler" was utilized. Applying this scaling operation to each column determines the maximum and minimum values within that column. The standard value is computed for each observation value using the maximum and minimum values in the corresponding columns, and a scaled value is obtained. Utilizing this procedure, values falling within the interval [0,1] are acquired. This procedure requires developing, training, and assessing models according to predetermined performance indicators to select the most suitable model. The scaled and standardized dataset is partitioned into training and test data for this objective. Following the existing and related work, the separation process

TABLE II. CLUSTERING OF ISTANBUL'S DISTRICTS

Cluster	**Districts in the Cluster**
Cluster 1	Bakırköy, Bayrampaşa, Beşiktaş, Beyoğlu, Esenler, Fatih, Güngören, Şişli, Zeytinburnu
Cluster 2	Esenyurt, Küçükçekmece, Pendik, Ümraniye
Cluster 3	Bağcılar, Bahçelievler, Kadıköy, Maltepe, Üsküdar
Cluster 4	Islands, Arnavutköy, Beykoz, Çatalca, Silivri, Sultanbeyli, Şile, Tuzla
Cluster 5	Ataşehir, Avcılar, Başakşehir, Beylikdüzü, Büyükçekmece, Çekmeköy, Eyüpsultan, Gaziosmanpaşa, Kağıthane, Kartal, Sancaktepe, Sarıyer, Sultangazi

	Date	Hour	TransportationType	Road	Line	TransferType	NumberofPasage	NumberofPassenger
6247610	2022-10-19	19	1	OTOYOL	ALTUNIZADE-SULTANBEYLI	Aktarma	235	235
5569277	2022-09-19	5	1	OTOYOL	ALTUNIZADE-SULTANBEYLI	Normal	461	452
438501	2022-01-21	18	1	OTOYOL	ALTUNIZADE-SULTANBEYLI	Normal	1426	1419
7837173	2022-12-30	16	1	OTOYOL	ALTUNIZADE-SULTANBEYLI	Normal	2077	2036
6319010	2022-10-23	4	1	OTOYOL	ALTUNIZADE-SULTANBEYLI	Normal	26	26
...	...	...	...	...	...	...	...	...
771727	2022-02-07	9	1	OTOYOL	ALTUNIZADE-SULTANBEYLI	Normal	1349	1337
894539	2022-02-12	19	1	OTOYOL	ALTUNIZADE-SULTANBEYLI	Normal	1578	1575
832404	2022-02-09	22	1	OTOYOL	ALTUNIZADE-SULTANBEYLI	Normal	741	739
4447554	2022-07-28	21	1	OTOYOL	ALTUNIZADE-SULTANBEYLI	Normal	977	948
6199310	2022-10-17	18	1	OTOYOL	ALTUNIZADE-SULTANBEYLI	Aktarma	228	227

Fig. 2. Structure of the Raw Data Set.

is executed to ensure that 20% of the dataset comprises test data (3293 observation values) and 80% comprises training data (13 172 observation values).

IV. RESULTS

The data used to determine the demand for public transportation in Istanbul, including all modes of transportation, consists of hourly public transportation statistics for 2022. The model was created using various regression approaches, including linear regression, ridge regression, lasso, polynomial regression, elastic net, k-nearest neighbors, decision tree regression, and random forest regression, employing the training data. The study's results revealed that the Lasso regression yielded the most efficient model. The linear regression method was employed to evaluate the model and meet the 5-year demand. The Altunizade-Sultanbeyli, Üsküdar-Beykoz, and Üsküdar-Çekmeköy lines that were part of the training data were utilized during the execution of this procedure. Consequently, all three lines were derived from a collection of 954 distinct lines and 7 866 727 intersection values of clustering and trip numbers. The proportion of the area covered by the three specified lines is 0.5852. Regions of demand were established for the points located on these lines, and the maximum cluster cover model was executed based on the outcomes of the linear regression model. An Üsküdar district line expansion was accomplished as a consequence of the optimal outcomes that were achieved.

A. Estimation of Future Number of Trips

After the dataset is completed, the optimal model for demand forecasting is identified. The available training data was used to perform several regression techniques, including linear regression, ridge regression, lasso, elastic net, k-nearest neighbor, decision tree, and random forest. As a result, Table IV was generated to summarize the results obtained. The optimal performance indicator values for ridge and lasso regression were achieved with alpha values of 0.5 and 10, respectively. The elastic net is a hybrid model that combines the features of lasso and ridge regression. Optimal performance is achieved when the function is used with specific parameter values. Polynomial regression sets the degree increment by the number of variables. The optimal values were achieved at degrees different from six degrees. Polynomial regression with the lowest degree was selected to avoid the complex structure.

Upon examination of the values of the training data in Table IV, it is evident that the decision tree model yields the most favorable performance metrics. However, upon examining the results, the decision tree was not selected as the prediction approach due to its excessive learning from the data, which was considered undesirable. Subsequently, the optimal linear regression technique was chosen. Upon running the test data on the model, it is observed that the resulting performance indicator is 0.871. This demonstrates that the resulting model can be utilized for predicting demand. It is a well-established fact that for the resulting model to be considered usable, the R2 value of the performance indicator must be no less than 0.7. The mean squared error represents the average of the squared difference between the original and predicted values in the data set. In contrast, the mean absolute error shows the average of the absolute difference between the real and predicted values in the dataset. The square root of the mean squared error is the root mean squared error. The mean

	Date	Hour	TransportationType	Road	TransferType	NumberofPasage	NumberofPassenger
215	2022-01-01	0	1	3	1	60	60
254	2022-01-01	0	1	3	2	134	133
293	2022-01-01	1	1	3	1	74	74
338	2022-01-01	1	1	3	2	254	253
390	2022-01-01	2	1	3	2	101	101
...	...	...	...	...	...	...	...
7865336	2022-12-31	21	1	3	2	745	706
7865587	2022-12-31	22	1	3	1	111	111
7865620	2022-12-31	22	1	3	2	584	557
7866372	2022-12-31	23	1	3	1	61	61
7866610	2022-12-31	23	1	3	2	358	331

Fig. 3. Latest Version of the Data Prepared for Analysis (Hourly Transportation Data of Altunizade-Sultanbeyli Line).

```
               Hour  TransportationType     Road  TransferType  \
count  16465.000000             16465.0  16465.0  16465.000000
mean      12.048709                 1.0      3.0      1.532038
std        6.778097                 0.0      0.0      0.498988
min        0.000000                 1.0      3.0      1.000000
25%        7.000000                 1.0      3.0      1.000000
50%       12.000000                 1.0      3.0      2.000000
75%       18.000000                 1.0      3.0      2.000000
max       23.000000                 1.0      3.0      2.000000

       NumberofPasage  NumberofPassenger
count    16465.000000       16465.000000
mean       648.927786         635.243243
std        688.409986         673.303385
min          1.000000           1.000000
25%        132.000000         131.000000
50%        242.000000         241.000000
75%       1208.000000        1172.000000
max       3312.000000        3274.000000
```

Fig. 4. Descriptive Statistics of the Data Used in Analysis.

absolute error (MAE) and its percentage variant (MAPE) or the mean square error (MSE) and its rooted variant (RMSE) are utilized in the majority of regression studies. For this reason, in this study, these performance indicators employed are evaluated based on their "MSE," "RMSE," "MAE," and "MAPE" values, which indicate the level of error. On the other hand, the "R2" value is used to evaluate the model's predictive capacity. Hence, it is crucial to minimize the error values and maximize the R2 value to get optimal performance for the model. Consequently, the estimated number of passengers for the following 5 years was approximated based on the obtained model.

B. Optimization of the Lines

The maximum cluster cover approach was employed to identify the areas necessitating an increase in the number of trips. The model determines if parts of the neighborhoods adjacent to the Altunizade-Sultanbeyli line should be added to the route based on the demand values. The neighborhoods located between the Altunizade-Sultanbeyli line were allocated proportionally based on variables such as population, accessibility to transit, and the need for services. The specific locations where these characteristics intersected were identified. Based on the acquired results, it was considered that the Altunizade line in the Üsküdar district should be expanded due to the high population density along that route.

TABLE IV. ERROR VALUES FOR GENERATED MODELS (TRAINING DATA)

Model	R2	MSE	RMSE	MAE	MAPE
Linear regression	0.960	0.0298	0.1726	0.0396	0.1297
Ridge (a = 0.5)	0.997	0.0298	0.1726	0.0396	3391
Lasso (a = 10)	0.930	0.0298	0.1726	0.0396	67 099
Elastic Net	0.434	257 602	507 545	419 157	686 529
Polynomial regression (degree = 2)	0.939	631.2	25 123	122 978	72 856
K-Nearest neighbors	0.975	11 188	3344	1349	0.239
Decision tree	0.999	0.001	0.001	0.001	0.001
Random forest	0.942	26 317	162 227	120 773	0.001
Artificial neural networks	0.872	12 712	3565	390 417	589 167

MAE, mean absolute error; MAPE, mean absolute error percentage; MSE, mean square error; RMSE, rooted mean square error.

V. CONCLUSION

Based on the various scenarios outlined in this study, an estimation was established for the number of transit passes that will be issued in the next 5 years. The estimating process utilized the "Hourly Public Transportation Data Set" [1], which the Istanbul Metropolitan Municipality provided on the open data portal. To achieve the highest level of accuracy, various machine learning techniques, including linear regression, ridge regression, lasso regression, elastic-net regression, decision trees, and artificial neural networks, were employed. Linear regression yielded the most favorable outcome among the employed methodologies. The obtained demand information was utilized in the maximum cluster coverage model. The demand information was provided as demand values, and it was determined which points among the demand points between Altun izade-Sultanbeyli, Üsküdar-Beykoz, and Üsküdar-Çekmeköy should have more trips added. Upon evaluating the acquired results, it was determined that there is a need to manage transportation at the Üsküdar location and increase the number of trips. Anticipated outcomes include improving future passenger satisfaction and reducing potential chaos, provided that restrictions are implemented and trip frequencies are increased. With these steps and the currently available data set in mind, future research may employ time series methods in addition to these algorithms to develop alternatives to the methods used in demand forecasting and to determine the optimal method among multiple models. Consequently, choosing models from a more extensive pool can determine the optimal approach. Furthermore, seasonal effect-considering factors are incorporated along with time series. Furthermore, the study's scope extends 5 years into the future, allowing for adjustments to be made to reflect a longer planning period. Furthermore, implementing a multi-period mathematical model enables the gradual processing of lines that require expansion instead of their simultaneous arrangement. In the end, it may be preferable to use a mathematical model that accommodates all passengers rather than one that attempts to accommodate the greatest number of individuals through an alternative approach.

Peer-review: Externally peer-reviewed.

Author Contributions: Concept – E.G., Z.G.; Design – E.G., Z.G.; Supervision – Z.G.; Materials – E.G.; Data Collection and/or Processing – E.G.; Analysis and/or Interpretation – E.G., Z.G.; Literature Search – E.G.; Writing – E.G., Z.G.; Critical Review – Z.G.

Acknowledgment: This study is a part of the MSc thesis titled "Forecasting the Number of Passes and Optimization of The Number of Transportation According to Istanbul Hourly Public Transportation Data" at the Institute of Graduate Studies, Istanbul University-Cerrahpasa, Istanbul, Turkey.

Declaration of Interests: The authors have no conflicts of interest to declare.

Funding: The authors declared that this study has no financial support.

REFERENCES

1. I. B. B. Open Data Portal, "Hourly public transport data set," Available: https://data.ibb.gov.tr/dataset/hourly-public-transport-data-set.
2. M. Yaylali, and Ö. Dilek, "Determination of factors affecting passengers' airline transportation preferences in Erzurum," *Marmara Univ. J. Econ. Admin. Sci.*, vol. 26, no. 1, pp. 1–21, 2009.
3. H. Ceylan, and T. Özcan, "Optimization of service frequencies in bus networks with the Harmony Search Algorithm: An application on the Mandl test network," *Pamukkale Univ. J. Eng. Sci.*, vol. 24, no. 6, pp. 1107–1116, 2018. [CrossRef]
4. T. Özcan, *Trip Frequency Optimization in Urban Public Transportation Systems*. Pamukkale University Institute of Science and Technology, 2018.
5. M. Demirkollu, *Determining the Number of Bus Trips with Target Programming Management*. Pamukkale University Institute of Science and Technology, 2017.
6. A. Deri, *Determination of Public Transportation Travel Demand and Trip Scheduling Optimization Using Smart Card Data*. Dong-Eui University Institute of Science and Technology, 2012.
7. N. Uludağ, *Modeling of Bus Lines with Fuzzy Optimization and Linear Goal Programming Approaches*, 2010.
8. M. A. Gencer, H. M. Alakaş, E. Tamer, and M. Hamurcu, "Scheduling of Ankara Metro M1 line departure times: A decision support system application," *J. Intell. Transp. Syst. Appl.*, vol. 1, no. 2, pp. 108–128, 2018.
9. T. Engin, *Public Transportation System Movement Schedule Optimization with Genetic Algorithm: Çanakkale Example*. Turkey: Marmara University, 2013.
10. A. Sağbaş, and U. Polat, "Modeling of urban public transportation systems with linear goal programming approach," *Eur. J. Eng. Appl. Sci.*, vol. 5, no. 2, pp. 112–121. [CrossRef]
11. U. Polat, A. Sağbaş, and M. S. Dermenci, *Schedule Optimization on Urban Bus Lines for Sustainable Transportation Planning*, 2022.
12. B. T. Palamutçuoğlu, *Trip Frequency Optimization with Smart Card Data Mining in Urban Public Transportation Planning*, 2020.
13. M. Nar, *Station and Line-Based Passenger Demand Forecasting for Yenikapi M1-Kirazli M1 Line and Prioritization of Critical Success Factors in Determining Trip Frequency in Rail Transportation Systems with Multi-criteria Decision-Making Methods. Critical Success Factors in Headway Determination in Railway Systems with Multi-criteria Decision-Making Methods*, 2022.
14. M. Ç. Kızıltaş, and D. Akkaya, "Innovative public transportation and examination of frequency service parameter in this context," *Eur. J. Sci. Technol.*, vol. 41, pp. 315–323, 2022.
15. Z. Erpik, *Genetic Algorithm Approach to Line Scheduling Problem: İETT Example*. Social Sciences Institute, 2019.
16. E. S. Korkmaz, Ö. İlyas, and S. B. Efe, "Short-term passenger number estimation in urban bus transportation," *J. Intell. Transp. Syst. Appl.*, vol. 5, no. 2, pp. 133–145.
17. H. Şaşmaz, *Main Feeder Line Modeling in Public Transportation: The Example of Istanbul Mahmutbey Transfer Center*. Institute of Science and Technology.
18. T. Efendigil, and Ö. E. Eminler, "The importance of demand forecasting in the aviation industry: A forecasting model on passenger demand," *Yaşar Univ. J.*, vol. 12, pp. 14–30, 2017.
19. F. Çakır, *Prediction of Railway Passenger Transportation Demand with Artificial Neural Networks*. Aksaray University Institute of Science and Technology, 2020.
20. A. B. Yaşar, *Demand Forecast in Urban Bus Transportation*. Pamukkale University Institute of Science and Technology, 2009.
21. R. B. Carmona-Benítez, and M. R. Nieto-Delfín, "*Bootstrap estimation intervals using bias corrected accelerated method to forecast air passenger demand*". Paper presented at the Computational Logistics: 6th International Conference, Delft, The Netherlands: ICCL, p. 6, 2015.
22. R. B. Carmona-Benítez, M. R. Nieto, and D. Miranda, "An Econometric Dynamic Model to estimate passenger demand for air transport industry," *Transp. Res. Procedia*, vol. 25, pp. 17–29, 2017. [CrossRef]
23. F. Jin, Y. Li, S. Sun, and H. Li, "Forecasting air passenger demand with a new hybrid ensemble approach," *J. Air Transp. Manag.*, vol. 83, p. 101744, 2020. [CrossRef]
24. S. Kim, and D. H. Shin, "Forecasting short-term air passenger demand using big data from search engine queries," *Autom. Constr.*, vol. 70, pp. 98–108, 2016. [CrossRef]
25. U. Gunter, and B. Zekan, "Forecasting air passenger numbers with a GVAR model," *Ann. Tourism Res.*, vol. 89, p. 103252, 2021. [CrossRef]
26. E. Suryani, S.-Y. Chou, and C.-H. Chen, "Air passenger demand forecasting and passenger terminal capacity expansion: A system dynamics framework," *Expert Syst. Appl.*, vol. 37, no. 3, pp. 2324–2339, 2010. [CrossRef]
27. A. Adeniran, O. Kanyio, and A. S. Owoeye, *Forecasting Methods for Domestic Air Passenger Demand in Nigeria*, 2018.
28. M. C. Gelhausen, P. Berster, and D. Wilken, "A new direct demand model of long-term forecasting air passengers and air transport movements at German airports," *J. Air Transp. Manag.*, vol. 71, pp. 140–152, 2018. [CrossRef]
29. P. Srisaeng, G. Baxter, S. Richardson, and G. Wild, "A forecasting tool for predicting Australia's domestic airline passenger demand using a genetic algorithm," *J. Aerosp. Technol. Manag.*, vol. 7, no. 4, pp. 476–489, 2015. [CrossRef]
30. R. A. Scarpel, "Forecasting air passengers at São Paulo International Airport using a mixture of local experts model," *J. Air Transp. Manag.*, vol. 26, pp. 35–39, 2013. [CrossRef]
31. İ. Koç, and E. Arslan, "*Demand forecasting for domestic air transportation in Turkey using artificial neural networks*", Paper presented at the 2018 6th International Conference on Control. Engineering & Information Technology (CEIT), 2018.
32. C.-Y. Hsiao, and M. Hansen, "A passenger demand model for air transportation in a hub-and-spoke network," *Transp. Res. E*, vol. 47, no. 6, pp. 1112–1125, 2011. [CrossRef]
33. S. Anvari, S. Tuna, M. Canci, and M. Turkay, "Automated Box–Jenkins forecasting tool with an application for passenger demand in urban rail systems," *J. Adv. Transp.*, vol. 50, no. 1, pp. 25–49, 2016. [CrossRef]
34. N. Jafari, "The chaos on US domestic airline passenger demand forecasting caused by COVID-19," *IJBFMI*, vol. 7, no. 3, pp. 241–258, 2022. [CrossRef]
35. V. Plakandaras, T. Papadimitriou, and P. Gogas, "Forecasting transportation demand for the US market," *Transp. Res. A*, vol. 126, pp. 195–214, 2019.
36. S. L. Marie-Sainte, T. Saba, and S. Alotaibi, "Air passenger demand forecasting using particle swarm optimization and firefly algorithm," *J. Comp. Theor. Nanosci.*, vol. 16, no. 9, pp. 3735–3743, 2019. [CrossRef]
37. X. Zhang, Y. Zheng, and S. Wang, "A demand forecasting method based on stochastic frontier analysis and model average: An application in air travel demand forecasting," *J. Syst. Sci. Complexity*, vol. 32, no. 2, pp. 615–633, 2019. [CrossRef]
38. S. Zhao, and X. Mi, "A novel hybrid model for short-term high-speed railway passenger demand forecasting," *IEEE Access*, vol. 7, pp. 175681–175692, 2019. [CrossRef]
39. A. Cyril, R. H. Mulangi, and V. George, "Modeling and forecasting bus passenger demand using time series method" 7th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), Vol. 2018, 2018, pp. 460–466.
40. Y. Xiao, J. J. Liu, Y. Hu, Y. Wang, K. K. Lai, and S. Wang, "A neuro-fuzzy combination model based on singular spectrum analysis for air transport demand forecasting," *J. Air Transp. Manag.*, vol. 39, pp. 1–11, 2014. [CrossRef]
41. P. Srisaeng, G. Baxter, and G. Wild, "Using an artificial neural network approach to forecast Australia's domestic passenger air travel demand," *World Rev. Intermodal Transp. Res.*, vol. 5, no. 3, pp. 281–313, 2015. [CrossRef]
42. K. Alekseev, and J. Seixas, "*Forecasting the air transport demand for passengers with neural modeling*," Proceedings of Paper presented at the VII Brazilian Symposium on Neural Networks, Vol. 2002, 2002.
43. I. Terekhov, A. Evans, and V. Gollnick, "*Forecasting a global air passenger demand network using weighted similarity-based algorithms*". Paper presented at the Complex Networks VII: Proceedings of the 7th Workshop on Complex Networks CompleNet, vol. 2016, 2016.

Emel GEÇİCİ received her bachelor's degree from Nişantaşı University Computer Engineering in 2019. She has completed her MSC degree in İstanbul University-Cerrahpaşa in 2023. She is currently working a backend developer in BELBİM in İstanbul.

Zeynep GÜRKAŞ-AYDIN is an Assistant Professor in the Department of Computer Engineering Cyber Security Department at Istanbul University in Istanbul, Turkey. She received her BSc and MSc in Computer Engineering from Istanbul University. In 2011 and 2014, she received PhD degrees in computer engineering from Istanbul University and computer science from Université Pierre-Et-Marie-Curie: Paris VI / Telecom SudParis. Her current areas of expertise include data and computer communications, wireless and mobile networks, the internet of things, and cyber security. She is also a member of Istanbul University-Cerrahpaşa's Internet of Things Security Test and Evaluation Center (ISTEC). Additionally, she has authored publications on mobility management, indoor localization, content delivery networks, and wireless communications.