MULTI-LAYER PERCEPTRON BASED TRANSFER PASSENGER FLOW PREDICTION IN ISTANBUL TRANSPORTATION SYSTEM

Anıl Utku[1] and Sema Kayapınar Kaya[2]*

[1] Munzur University, Department of Computer Engineering, Tunceli, Turkey
[2] Munzur University, Department of Industrial Engineering, Tunceli, Turkey

Original scientific paper

Abstract: *Predicting passenger movement in transportation networks is a critical aspect of public transportation systems. It allows for a greater understanding of traffic patterns, as well as efficient system evaluation and monitoring. It could also help with precise timing to emergencies or important events, as well as the improvement of urban transport system weaknesses and service quality. The number of transfer passengers demand in Istanbul, Turkey's biggest and most developed metropolis, was used to construct a real-world forecasting model in this study. The number of transfer passengers has been forecasted using popular machine learning methods such as kNN (k-Nearest Neighbours), LR (Linear Regression), RF (Random Forest), SVM (Support Vector Machine), XGBoost and MLP. The dataset utilized is made up of hourly passenger transfer counts gathered at two public transportation transfer stations in Istanbul in January 2020. Using MSE, RMSE, MAE and R^2 parameters, each model's experimental data have been thoroughly evaluated. MLP has more successfully other machine learning algorithms in the majority of transportation lines, according to the experimental results.*

Key words: *Machine learning, passenger flow management, transfer data.*

1. Introduction

In 2020, the city, which straddles the Bosporus and is located in both Europe and Asia, have a population of over 15 million people, contributing for 20 percent of Turkey's total population. (TUIK, 2021). In according to world demographics data, Istanbul is the most crowded city in Europe and the world's fifteenth most densely populated metropolis (Statista, 2020). The number of passengers utilizing public transportation is significantly higher as a result of the high density of the population (Pamucar et al., 2020). While nearly 11 million 500 thousand people use public

* Corresponding author.
        E-mail addresses: anilutku@munzur.edu.tr (A. Utku), semakayapinar@munzur.edu.tr (S.K. Kaya)

transportation in Istanbul every day, passengers who prefer highway transportation (metrobus, public urban transportation, private bus, etc.) account for nearly 84 %, followed by railway transportation (metro, light metro, tram, etc.) at around %14, and sealine transportation at just under 2% (IETT, 2021). Despite the constant increase in the number of people and vehicles, the fact that the proportion of cars per thousand people is also constantly rising is important in terms of showing the increased traffic density in Istanbul. According to the traffic index published by TomTom, Europe's largest navigation systems company, Istanbul is the fifth city with the highest traffic density of 51% in the world (TomTom Traffic Index, 2021). With the increase of urban transportation challenges, forecasting the number of people entering and departing Istanbul's transit terminals has become more challenging. Passenger flow forecasting provides a better understanding of travel patterns, efficient monitoring and evaluation of the system status of Istanbul transportation system. It may also help in the prompt response to crises or special events, as well as the correction of defects and enhancement of public transportation service quality.

Several predicting methodologies have been proposed to enhance the effectiveness of passenger forecasting models, encompassing mathematical modelling methods, statistical methods, and non-parametric methodology. The machine learning-based (ML) framework is one of the most well-known non-parametric approaches today (Boukerche and Wang, 2020). It's a subset of AI that integrates the problem of learning from data samples with the concept of reasoning in generally (Boukerche & Wang, 2020). It's a subfield of AI that relates the difficulty of learning from sample data with the concept of reasoning in overall (Tom Mitchell, 2006). There are two stages to any learning process: (i) Particular a given dataset, calculation of unknown relationships in a system Particular a given dataset, calculation of unknown relationships in a system (ii) predicted connections are used to forecast new platform outputs. Machine Learning has also been shown to be an interesting topic of study in passenger demand prediction, with several applications (Liu et al., 2020, Zheng et al., 2021, Wang et al., 2021, Hayadi et al., 2021; Gummadi and Edara (2018); Ye et al. (2019); Messinis and Vosniakos, (2020); Liu et al. (2020); Hayadi et al. (2021); Guo et al. (2021); Wang et al. (2021); Bozanic et al. (2021); Yang et al.(2021); Ge et al. (2021); Kamandanipour et al. (2022); Müller-Hannemann et al. (2022). The ability to predict passenger traffic in transportation networks is critical to public transportation management. It helps to improve transportation services, provide early warnings for unusual traffic situations, and make cities smarter and safer. Furthermore, transfer passenger flow prediction can improve the transfer operation efficiency reduce the transfer waiting time and enhancing passengers' satisfaction. To address this problem, the transfer passenger flow transferring a various modes transportation (metro & tram, bus & metrobus, rail, and ferries & sea-bus) in Istanbul has been developed for the first time in the literature.

The followings are some of the study's contributions:

I.   This paper offers a clear theoretical foundation and decision support for the practical work of using intelligent technologies to optimize the predictive performance of the number of passengers moving between different modes, including "metro and tram," "bus and metrobus," "rail," and "ferries and sea-bus."

II. As Istanbul has a very heavy traffic; the number of lines can be increased or decreased according to the number of passengers. Accurate transfer passenger volume is the fundamental of transportation scheduling in Istanbul.

III. This enhances the service standards of an urban public transportation system and exposes passengers with real-time transfer passenger demand information across several routes, allowing them to make greater decision to travel.

IV. Prediction transfer passenger flow assists Istanbul IETT authorities and management in increasing public transit reliability of the system, improving passenger experience, and maximizing routing plans.

The motivation of this paper is the prediction of the number passengers transferring in various lines in Istanbul recorded at 1-hour intervals between 1-31 January 2020. Istanbul is Turkey's biggest and most developed metropolis therefore, the dataset utilized comprises of passenger transfer numbers on several transportation lines in Istanbul. The goal of this research is to anticipate the number of transfer passengers in Istanbul. The number of transfer passengers was determined based on passenger data gathered during one-hour intervals. The Istanbul Public Urban Transport Company (IETT), Private Public Bus (OHO), motor/boat, and the IETT tunnel will have been subjected to empirical investigations. Time information such as hourly, daily, and weekly has been revealed in this fashion on certain lines. The goal of this research is to use machine learning techniques to predict the amount of transfer passengers on transportation lines using a different kNN, LR, RF, SVM, XGBoost and MLP methods.

2. Literature Review

With the development of big data technology, using machine learning algorithms to detect the principles of urban passenger movement has become one of the research hotspots in the field of public transportation. In recent decades, there has been a huge amount of work on passenger flow and forecasting using statistical methodological approaches notably Machine Learning. Xie et al. (2014) employed a combination of Seasonal Decomposition (SD) and Least Squares Support Vector Regression (LSSVR) methods to forecast air passenger volume for a short amount of time. Sun et al. (2015) proposed a hybrid Wavelet and Support Vector Machine (SVM) methods that consist of three significant levels to predict the number of people entering and leaving the subway in Beijing.

Roos et al. (2017) proposed a predicting technique, which is based on dynamic Bayesian Network (BN) built to function even passenger flow data is missing or uncertain. Ni et al. (2017) created a combination time series model based on seasonal ARIMA and Loss Function (LF), using data from the Twitter social media platform to monitor subway passengers. Toqué et al. (2017) addressed a passenger flow predicting in multimodal transport using ML methods such as Random Forest (RF) and Long-Short Term Memory (LSTM) neural networks. Zhang et al. (2017) predicted the short-term passenger data taken from GPS device and smart card system in favour of the two-Step Real Time Prediction (2RTP) approach based on the extended Kalman Filtering (KF) method. Milenković et al. (2018) estimated an ARIMA analysis to simulate the monthly number of train passengers while considering seasonal variations into consideration.

Gummadi and Edara (2018) employed the ARIMA and seasonal ARIMA to estimate bus passenger flow in India's transport industry over a short period of time. Ye et al. (2019) aimed to predict the daily bus passenger traffic amount using the ARIMA method and examined the outcomes of predictions in the case of complete weekday non-peak data collected from January to March 2018. Li et al. (2020) predicted shared passenger demand in various locations with a hybrid algorithm based on WT-FCBF-LSTM (Wavelet Transform, Fast Correlation-based Filter, and Long Short-term Memory). Liu et al. (2020) focused on a short-term estimation model for local bus passenger flow using SVM. Hayadi et al. (2021) proposed a Random Forest (RF) using

the location data from the GPS devices in the buses, the location of the bus stop used for operation management, and the volume of traffic estimated by an image processing method. Li et al. (2021) adopted the seasonal ARIMA and SVM to predict the periodic flow of railway passenger.

Guo et al. (2021) proposed a regression tree combined with copula-based simulations employing passenger level data to generate real-time distributional estimates of travels in an airport. Rajendran et al. (2021) developed a logistic regression (LR), artificial neural networks (ANN), RF, and gradient boosting (GB) for assessing air taxi demand considering various factors such as temperature, weather conditions and visibility.

Zheng et al. (2021) designed an integrated LR, a fully connected neural network (NN) and LSTM model for anticipating a metro station's abnormally substantial passenger movement. Rodríguez-Sanz et al. (2021) presented two RF algorithms that allow for the integration of flight data and passenger judgement for predicting the duration of queues at check-in counters and the security control area at Parma de Mallorca airport in Spain. Wang et al. (2021) established a LightGBM method to estimate railway high passenger parameters like railway specifications, past weather trends, and public transport time sequence. The LightGBM methodology outscored the XGBoost, RF, and ARIMA algorithms, as according their findings. Yang et al.(2021) proposed a prediction model based on transit passenger flow using the wavelet analysis (WA) and LSTM combination model for the short-term period. Abeyrathna et al. (2021) investigated the relationship between the Regression Tsetlin (RT) machine algorithm and pandemic events such as daily COVID-19 cases and deaths, pandemic control measures to estimate the number of transport passengers under different scenarios. Jackson et al. (2021) benefited from various Bayesian Network (BN) models for predicting bus schedule time. Ge et al. (2021) implemented a combination of differentially ARIMA and SVM to achieve a highly predictive model for passenger flow in Shanghai-Guangzhou railway station.

Kamandanipour et al. (2022) presented a multi-layer ANN system to forecast the strength of demand caused by seasonal conditions using train ticket service data. Müller-Hannemann et al. (2022) investigated a new technique of approximating scenario-based resilience employing XGBoost, Catboost, SVR and ANN models which are based on carefully selected important aspects of public transport systems. Wood et al. (2022) analysed its use of traditional LR analysis and a RF model to unveil future passenger occupancies on a bus when it reaches at next stops using real-time data from bus operating and meteorological data. Reitmann and Schultz (2022) developed the gradient boosting (XGBoost) algorithm and the point-of-interest (POI) model, helping in the reduction of the passenger flow forecast model's total training time, to forecast bus passenger flow in Beijing. Comparisons of these models are listed in Table 1 in detail.

Table 1. Literature review of passenger flow prediction

Author (year)	Models	Passenger type	Period
Xie et al. (2014)	SD, LSSVR	Air	short
Sun et al. (2015)	Wavelet, SVM	Subway	short
Roos et al. (2017)	BN	metro	short
Ni et al. (2017)	Seasonal ARIMA, LF	Subway	short
Toqué et al. (2017)	RF, LSTM	Multi model	long
Zhang et al. (2017)	2RTP, KF	Bus	short

Author (year)	Models	Passenger type	Period
Milenković et al. (2018)	ARIMA	railway	short
Li et al. (2020)	WT-FCBF-LSTM	railway	long
Liu et al. (2020b)	SVM	bus	short
Li et al. (2021)	Seasonal ARIMA, SVM	urban	short
Guo et al. (2021)	RT	urban	short
Rajendran et al. (2021)	LR, ANN, RF, GB	Taxi urban	short
Zheng et al. (2021)	NN, LSTM, LR	Metro	short
Rodríguez-Sanz et al. (2021)	RF	airport	long
Wang et al. (2021)	LightGBM, XGBoost, RF, ARIMA	railway	long
Yang et al. (2021)	WA, LSTM	Transit	short
Abeyrathna et al. (2021)	RT	Public transport	short
Jackson et al. (2021)	BN	bus	short
Ge et al. (2021)	ARIMA, SVM	railway	long
Kamandanipour et al. (2022)	ANN	Railway	short
Müller-Hannemann et al. (2022)	XGBoost, Catboost, SVR and ANN	Public transport	-
Wood et al. (2022)	LR, RF	Bus	short
Reitmann and Schultz (2022)	XGBoost, POI	Bus	short

3. Machine Learning-Based Passenger Flow Prediction

The amount of immediate data produced by urban transportation systems is also expanding, thanks to the growth of big data, internet of things, sensor networks, and cloud computing applications. In topics like safety management, emergency response efficiency, and urban traffic management, passenger flow forecast in urban transportation networks is critical. Passenger flow planning is important for concerns including scheduling, traffic planning, and passenger flow control. The goal of this research is to anticipate the number of transfer passengers in Istanbul, Turkey's largest and most developed metropolis, using passenger flow data. The dataset utilized comprises of passenger transfer numbers on various transportation lines in Istanbul, such as transfers and normal boarding, recorded for one month between January 1, 2020 and January 31, 2020.

The objective of this research is to use machine learning algorithms to forecast the amount of transfer passengers on transportation lines. In practice, kNN (k-Nearest Neighbors), LR (Linear Regression), RF (Random Forest), SVM (Support Vector Machine), XGBoost (eXtreme Gradient Boosting), and MLP (Multi-layer Perceptron) have been examined then, each model's experimental findings have been thoroughly examined using MSE, RMSE, MAE, and R^2 metrics.

3.1. Original Data Analysis

In this study, a dataset consisting of the transfer numbers of passengers such as transfer and normal boarding in different transportation lines in Istanbul recorded at 1-hour intervals between 1-31 January 2020 by Istanbul Metropolitan Municipality has been used. The dataset used consists of 23163 rows of transportation data. The

dataset contains id, date_time, transport_type_id, transport_type_desc, line, transfer_type_id, transfer_type, number_of_passenger parameters.

In this study, IETT, ÖHO, motor/boat and IETT tunnel transfer lines have been selected for prediction because they have the highest number of transfer passengers. IETT transfer line refers to all bus lines offered by the Istanbul Metropolitan Municipality. ÖHO transfer line refers to all bus lines offered by private public bus companies. Motor/boat, on the other hand, refers to all sea transportation made by marine vehicles. IETT tunnel refers to all transfers made using the underground metro. Table 2 shows the first 10 rows of the dataset used as an example.

Table 2. A sample from the dataset

Date_time	Line	Transfer_type	Number_of_passenger
1.01.2020 00:00	Motor_Tekne	Normal	1393
1.01.2020 00:00	Kabataş_Bağcılar	Normal	4310
1.01.2020 00:00	Aksaray_Airport	Normal	2936
1.01.2020 00:00	Kabataş_Bağcılar	Transfer	1586
1.01.2020 00:00	Kadıköy-Kartal Metro	Transfer	677
1.01.2020 00:00	Kirazlı-Olimpiyatköy	Transfer	10
1.01.2020 00:00	Edirnekapı-Sultançiftliği	Normal	793
1.01.2020 00:00	Şehir Hatları	Transfer	59
1.01.2020 00:00	Taksim-4.Levent	Normal	8119

3.2. Methodology

In this study, popular machine learning algorithms commonly used in the literature such as kNN, LR, RF, SVM, XGBoost and MLP have been applied. The dataset has been pre-processed before applying to the models. Possible blank or incorrect fields in the data have been checked. After the data pre-processing step, training, validation, and test datasets have been selected. 80% of the dataset is split into training and 20% testing. 10% of the training data have been split for validation. Validation data has been used for the optimization of model parameters.

Time series data refers to series of numbers ordered according to a time index. Time series data refers to series of numbers ordered according to a time index. In supervised learning problems, it is aimed to estimate the output from the inputs by using a function like $y=f(x)$. Time series data can be transformed into supervised learning problem for use in time series analysis. The time series data can be transformed into a supervised learning problem by using the values from the previous time step to predict the value in the next time step as seen in Figure 1.

213
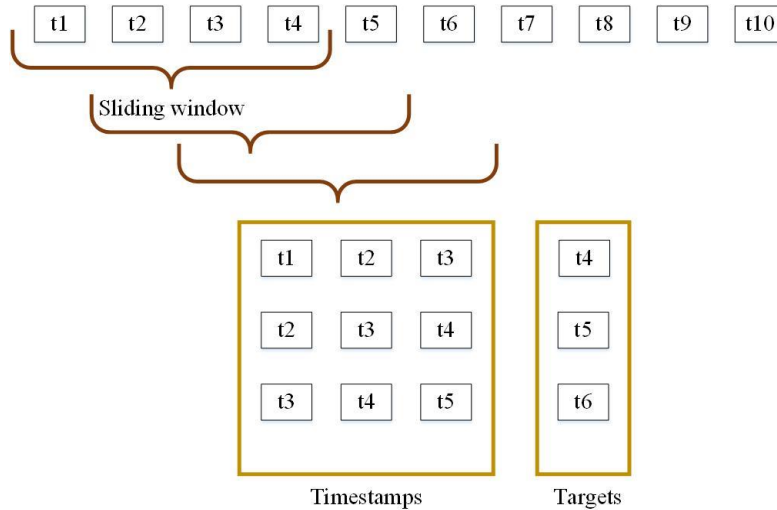
Figure 1. Converting time series data to supervised learning problem

In this study, time series data has been converted to supervised learning problem by using the sliding window method as seen in Figure 1. The number of previous timestamps determines the size of the sliding window. In this study, the size of the sliding window has been determined as 3 as a result of the experimental studies.

In order to optimize the parameters of the machine learning algorithms used, 10% of the training data has been used for validation. By using the optimized parameters, algorithms have been applied and prediction values have been obtained. The pseudo code of the developed system is presented below:

Input: Passenger transfer data on IETT, OHO, motor/boat and IETT tunnel lines
Output: Predicted passenger numbers
1: Start.
2: Checking the missing and incorrect areas in the data (data pre-processing).
3: Splitting training, validation and test sets and normalizing the data.
4: Optimizing model parameters using validation data.
5: Walk forward validation.
6: Have the parameters with the lowest MSE value been selected?
If yes go to step 7, if no go to step 4.
7: Creation of the model.
8: Making predictions using the created model.
9: Calculation of MSE, RMSE, MAE and R^2 values according to the prediction results.
10: Finish.

3.3. Developed Model

In this study, a comparative analysis of the passenger number estimation problem of the MLP-based model developed with popular machine learning algorithms is presented. MLP is a neural network model inspired by the neuron structure in the brain. MLP is a combination of perceptron's that bind in different ways and operate in different activation functions. It consists of input nodes, hidden nodes and output

nodes. Input nodes provide input information to the network. No computation is performed on any of the input nodes. This only relay information to hidden nodes. Hidden nodes are structures that are not directly connected to the outside world, perform calculations and transmit information from input nodes to output nodes. A hidden layer is created with a collection of hidden nodes. While a network has only a single input layer and a single output layer, it can have zero or multiple hidden layers. MLP has one or more hidden layers. Output nodes, on the other hand, are responsible for information processing and transferring information from the network to the outside world.

The developed MLP model takes the passenger flow data in the training dataset as input and predicts the passenger numbers in the test dataset. According to the obtained result, the training process has been continued. The architecture of the developed model is shown in Figure 2.
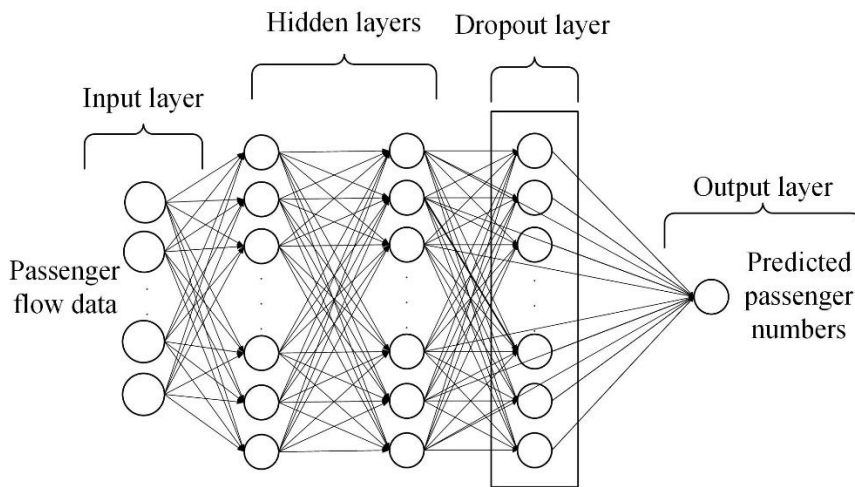
**Figure 2.** The architecture of the developed model

In the developed MLP-based model, there are an input layer, three hidden layers and an output layer as seen in Figure 6. Hidden layers represent an intermediate processing step that is combined using weighted sums to obtain the classification result. The developed model is a sequential model with linear layers. There is a dropout layer between the input layer and the hidden layer. In the output layer, there are two output units that return the prediction of the probability of customer loss. ReLU activation function is used in the input layer and hidden layers, and the sigmoid activation function is used in the output layer.

3.4. Experimental Results

In this study, a dataset consisting of the transfer numbers of passengers such as 1-month transfer and normal boarding in different transportation lines in Istanbul recorded at 1-hour intervals for 2020 has been used. IETT, ÖHO, motor/boat and IETT tunnel transfer lines with the highest transfer numbers have been selected for prediction. kNN, LR, RF, SVM, XGBoost and MLP algorithms, which are widely used in the literature, have been applied to the dataset. For each algorithm, the experimental results obtained using MSE, RMSE, MAE and R^2 metrics have been compared.

The IETT transfer line covers all bus lines offered by the Istanbul Metropolitan Municipality. IETT transfer line consists of passenger flow information transferring in 687 different time zones. 80% of this data is split for training and 20% for testing. After the train/test split, 6070 rows of data have been used in the training and 1518 rows of data have been used in the testing. Figure 3 shows the change in the number of transfer passengers on the IETT line over time. Table 3 show the average MSE, RMSE, MAE and R^2 results obtained for each algorithm for IETT line.
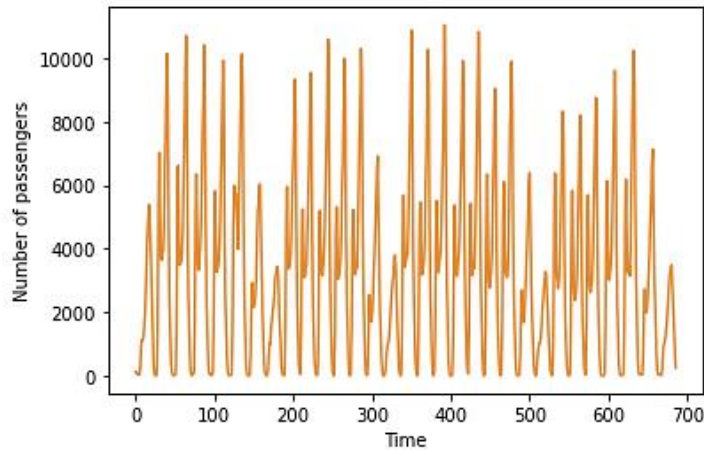
Figure 3. Change over time in the number of transfer passengers on the IETT line

Table 3. Experimental results for each model according to the MSE, RMSE, MAE and R^2 for IETT line

Model	MSE	RMSE	MAE	R^2
kNN	259682.920	509.590	323.066	0.958
LR	635367.832	797.091	653.525	0.906
RF	365886.694	604.883	354.582	0.944
SVM	237559.150	487.400	317.741	0.946
XGBoost	392332.530	626.364	411.415	0.942
MLP	227419.633	476.885	315.104	0.961

The experimental results show that the MSE values of kNN, LR, RF, SVM, XGBoost and MLP are 259682.920, 635367.832, 365886.694, 237559.150, 392332.530, 227419.633, respectively. The RMSE values of kNN, LR, RF and SVM are 509.590, 797.091, 604.883, 487.400, 626.364, 476.885, respectively. The MAE values of kNN, LR, RF and SVM are 323.066, 653.525, 354.582, 317.741, 411.415, 315.104, respectively. The R^2 values of kNN, LR, RF and SVM are 0.958, 0.906, 0.944, 0.946, 0.942, 0.961, respectively.

The ÖHO line covers all passenger transfers offered by private public bus companies. ÖHO transfer line consists of passenger flow information transferring in 716 different timestamps. 80% of this data is split for training and 20% for testing.

216

After the train/test split, 572 rows of data have been used in the training and 144 rows of data have been used in the testing. Figure 4 shows the change in the number of transfer passengers on the ÖHO line over time. Table 4 show the average MSE, RMSE, MAE and R^2 results obtained for each algorithm for ÖHO line.
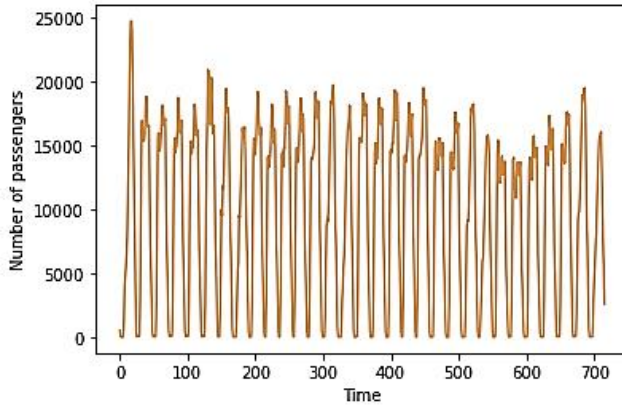
Figure 4. Change over time in the number of transfer passengers on the ÖHO line
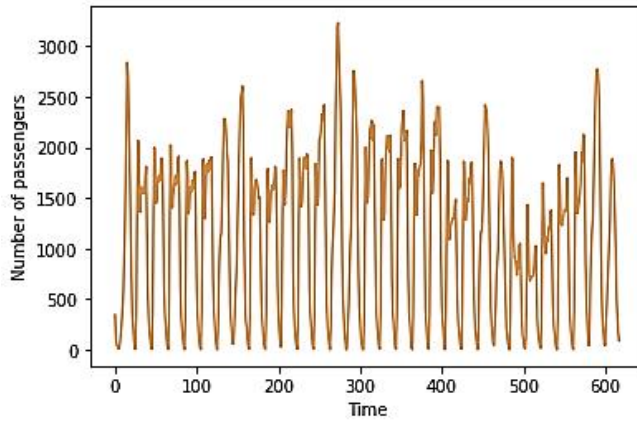
Table 4. Experimental results for each model according to the MSE, RMSE, MAE and R^2 for ÖHO line

Model	MSE	RMSE	MAE	R^2
kNN	1335463.741	1155.624	712.375	0.965
LR	2050366.100	1431.909	1117.355	0.949
RF	1332583.670	1154.375	710.408	0.967
SVM	1640037.728	1280.639	943.779	0.959
XGBoost	2236156.800	1495.378	931.525	0.944
MLP	1252185.815	1119.011	692.050	0.969

The experimental results show that the MSE values of kNN, LR, RF and SVM are 1335463.741, 2050366.100, 1332583.670, 1640037.728, 2236156.800, 1252185.815, respectively. The RMSE values of kNN, LR, RF and SVM are 1155.624, 1431.909, 1154.375, 1280.639, 1495.378, 1119.011, respectively. The MAE values of kNN, LR, RF and SVM are 712.375, 1117.355, 710.408, 943.779, 931.525, 692.050, respectively. The R^2 values of kNN, LR, RF and SVM are 0.965, 0.949, 0.967, 0.959, 0.944, 0.969, respectively.

Motor/boat transfer line refers to all transfers made by sea vehicles that provide sea transportation. Motor/boat transfer line consists of passenger flow information transferring in 618 different timestamps. 80% of this data is split for training and 20% for testing. After the train/test split, 494 rows of data have been used in the training and 124 rows of data have been used in the testing. Figure 5 shows the change in the number of transfer passengers on the motor/boat line over time. Table 5 show the

average MSE, RMSE, MAE and R^2 results obtained for each algorithm for motor/boat line.
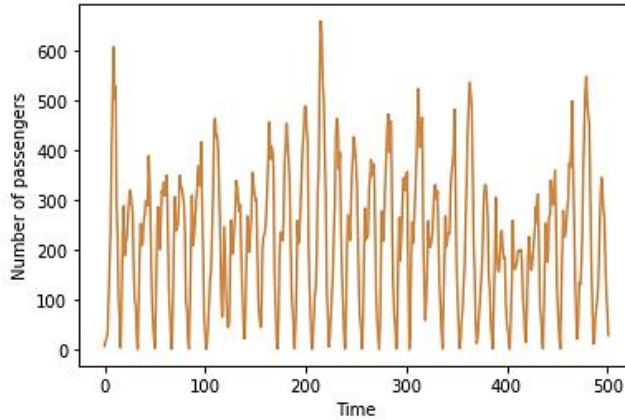

Figure 5. Change over time in the number of transfer passengers on the motor/boat line

Table 5. Experimental results for each model according to the MSE, RMSE, MAE and R^2 for motor/boat line

Model	MSE	RMSE	MAE	R^2
kNN	57453.940	239.695	160.711	0.884
LR	48366.810	219.924	168.390	0.903
RF	55962.547	236.565	159.844	0.887
SVM	34556.885	185.894	136.343	0.93
XGBoost	45494.010	213.293	144.571	0.907
MLP	30629.115	175.012	125.101	0.938

The experimental results show that the MSE values of kNN, LR, RF and SVM are 57453.940, 48366.810, 55962.547, 34556.885, 45494.010, 30629.115, respectively. The RMSE values of kNN, LR, RF and SVM are 239.695, 219.924, 236.565, 185.894, 213.293, 175.012, respectively. The MAE values of kNN, LR, RF and SVM are 160.711, 168.390, 159.844, 136.343, 144.571, 125.101, respectively. The R^2 values of kNN, LR, RF and SVM are 0.884, 0.903, 0.887, 0.930, 0.907, 0.938, respectively.

IETT tunnel transfer line refers to all transfers made using the underground metro. IETT tunnel transfer line consists of passenger flow information transferring in 502 different timestamps. 80% of this data is split for training and 20% for testing. After the train/test split, 401 rows of data have been used in the training and 101 rows of data have been used in the testing. Figure 6 shows the change in the number of transfer passengers on the IETT tunnel line over time.

Figure 6. Change over time in the number of transfer passengers on the
IETT tunnel line

Table 6 show the average MSE, RMSE, MAE and R^2 results obtained for each algorithm for IETT tunnel line.

Table 6. Experimental results for each model for IETT tunnel line

Model	MSE	RMSE	MAE	R^2
kNN	1909.902	43.702	34.096	0.879
LR	2619.355	51.179	40.426	0.835
RF	2070.691	45.504	34.879	0.869
SVM	2336.181	48.334	36.738	0.852
XGBoost	2832.245	53.218	40.686	0.836
MLP	1904.229	43.637	32.560	0.88

The experimental results show that the MSE values of kNN, LR, RF and SVM are 1909.902, 2619.355, 2070.691, 2336.181, 2832.245, 1904.229, respectively. The RMSE values of kNN, LR, RF and SVM are 43.702, 51.179, 45.504, 48.334, 53.218, 43.637, respectively. The MAE values of kNN, LR, RF and SVM are 34.096, 40.426, 34.879, 36.738, 40.686, 32.560, respectively. The R^2 values of kNN, LR, RF and SVM are 0.879, 0.835, 0.869, 0.852, 0.836, 0.880, respectively. The prediction results of the developed MLP-based model are shown in Figure 7.
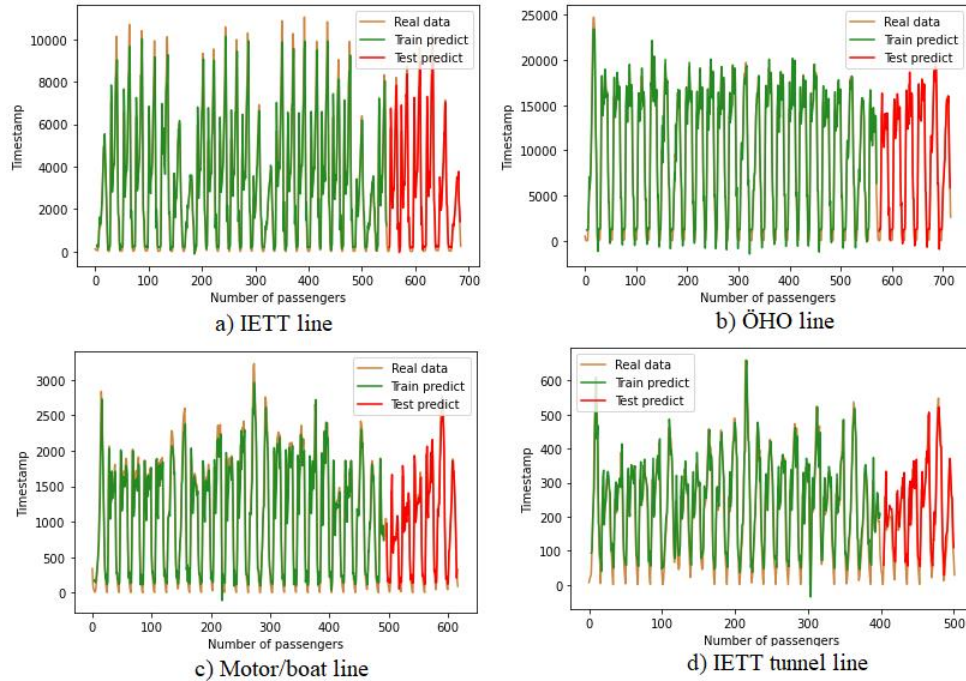
Figure 7. Prediction results of developed MLP-based model

The prediction results of the developed model for the IETT line in Figure 7.a, the ÖHO line in Figure 7.b, the motor/boat line in Figure 7.c and the IETT tunnel line in Figure 7.d are shown. As can be seen in the Figure 7, the MLP-based model successfully predicted the patterns in the training and test data.

4. Conclusions and Future Studies

In this study, a comparative analysis of popular machine learning algorithms such as kNN, LR, RF, SVM, XGBoost and MLP for passenger flow prediction is presented. The experimental results for IETT, ÖHO, motor/boat and IETT tunnel lines have been extensively tested using MSE, RMSE, MAE and R^2.

For the IETT line, the MLP has more successful than the other models compared. After MLP, SVM, kNN, RF, XGBoost and LR have been successful, respectively. For the ÖHO line, the MLP has more successful than the other models compared. After MLP, RF, kNN, SVM, LR and XGBoost have been successful, respectively. For the motor/boat line, the MLP has more successful than the other models compared. After MLP, SVM, XGBoost, LR, RF and kNN have been successful, respectively. For the IETT tunnel line, the MLP has more successful than the other models compared. After MLP, kNN, RF, SVM, XGBoost and LR have been successful, respectively.

Experimental results show that these machine learning methods can be used in passenger flow prediction problems. Among the compared algorithms, MLP achieved successful results in all of the transportation lines. MLP is a neural network model developed based on biological neural network structures. The MLP consists of interconnected processing units, similar to the functioning of neurons. MLP's ability to detect non-linear, linear or non-linear distributed data makes it perform well on most datasets. XGBoost is a machine learning model that uses a gradient boosting

220

framework. XGBoost is a decision-tree and gradient-boosting based machine learning model. It works successfully on non-structured data such as images, text and audio. kNN may be inefficient in terms of performance on small datasets. SVM is successful when having a limited set of points. SVM is good at outliers as it will only use the most relevant points to find support vectors. For this reason, SVM have successful results in this study. LR is expected to be successful when the dataset is truly linear, especially when there are many features with a very low signal-to-noise ratio. However, RF may fail to model linear combinations of many features.

All methods compared in this study had successful results. All methods had R^2 values above 0.90 for the IETT line, above 0.94 for the ÖHO line, above 0.88 for the motor/boat line, and above 0.84 for the IETT tunnel line. Experimental results showed that the developed MLP-based model gives better results than the compared models for all transfer lines used in the prediction of the number of passengers. The prediction of the number of passengers is an important factor for the urbanization phenomenon and city management. Transportation planning is also important in terms of avoiding disruptions in transportation and reducing the traffic load. The developed model can be applied to real-world problems by using effective passenger predicting in the field of transportation planning. In future studies, longer-term predictions can be made using passenger data over a larger time period. In addition, the results can be evaluated by applying different models such as deep learning.

In this study, traditional machine learning methods and MLP, which is a neural network-based model, are compared in practice. Here, it is aimed to benefit from the prominent features of neural networks in the time series prediction problem. The ability of a neural network to process data in detail stems from its ability to reveal hidden patterns between input and output data. An important advantage of neural networks is that they have the ability to learn and generalize information. MLP is tolerant of missing values and can model complex relationships such as nonlinear trends. It can also support multiple inputs.

One of the important limitations of this study is that it only considers the number of transfer passenger volume prediction. For this reason, different external factors such as transfer time, rush hours and holiday days could be examined for passenger prediction model in the future. Secondly, ML algorithms such as kNN, LR, RF, SVM, XGBoost and MLP methods was employed during the short-term prediction process. In the further study, a state of art deep neural network algorithm could be developed to improve the prediction result for the number of transferring passengers.

Conflicts of Interest*: The authors declare no conflicts of interest.

References

Abeyrathna, K. D., Rasca, S., Markvica, K., & Granmo, O.C. (2021). Public Transport Passenger Count Forecasting in Pandemic Scenarios Using Regression Tsetlin Machine. Case Study of Agder, Norway. In Smart Transportation Systems 2021 (pp. 27–37). Springer.

Boukerche, A., & Wang, J. (2020). Machine Learning-based traffic prediction models for Intelligent Transportation Systems. Computer Networks, 181, 107530.

Bozanic, D., Tešić, D., Marinković, D., & Milić, A. (2021). Modeling of neuro-fuzzy system as a support in decision-making processes. Reports in Mechanical Engineering, 2(1), 222-234.

Ge, M., Junfeng, Z., Jinfei, W., Huiting, H., Xinghua, S., & Hongye, W. (2021). ARIMA-FSVR Hybrid Method for High-Speed Railway Passenger Traffic Forecasting. Mathematical Problems in Engineering, 2021.

Gummadi, R., & Edara, S. R. (2018). Analysis of Passenger Flow Prediction of Transit Buses Along a Route Based on Time Series. In S. C. Satapathy, J. M. R. S. Tavares, V. Bhateja, & J. R. Mohanty (Eds.), Information and Decision Sciences (pp. 31–37). Springer. https://doi.org/10.1007/978-981-10-7563-6_4

Guo, X., Grushka-Cockayne, Y., & De Reyck, B. (2021). Forecasting Airport Transfer Passenger Flow Using Real-Time Data and Machine Learning. Manufacturing & Service Operations Management. https://doi.org/10.1287/msom.2021.0975

Hayadi, B. H., Kim, J.-M., Hulliyah, K., & Sukmana, H. T. (2021). Predicting Airline Passenger Satisfaction with Classification Algorithms. International Journal of Informatics and Information Systems, 4(1), 82–94.

In Public Transportation in Istanbul. (2021). https://iett.istanbul/en/main/pages/public-transportation-in-istanbul/316, Accessed 13 January 2022.

Jackson, M. D., Leung, C. K., Mbacke, M. D. B., & Cuzzocrea, A. (2021). A Bayesian framework for supporting predictive analytics over big transportation data. 2021 IEEE 45th Annual Computers, Software, and Applications Conference (COMPSAC), 332–337.

Kamandanipour, K., Yakhchali, S. H., & Tavakkoli-Moghaddam, R. (2022). Learning-based dynamic ticket pricing for passenger railway service providers. Engineering Optimization, 10(1), 1–15.

Li, W., Sui, L., Zhou, M., & Dong, H. (2021). Short-term passenger flow forecast for urban rail transit based on multi-source data. EURASIP Journal on Wireless Communications and Networking, 2021(1), 1–13.

Li, X., Zhang, Y., Du, M., & Yang, J. (2020). The forecasting of passenger demand under hybrid ridesharing service modes: A combined model based on WT-FCBF-LSTM. Sustainable Cities and Society, 62, 102419.

Liu, W., Tan, Q., & Wu, W. (2020a). Forecast and early warning of regional bus passenger flow based on machine learning. Mathematical Problems in Engineering, 2020, 6625435, https://doi.org/10.1155/2020/6625435.

Liu, W., Tan, Q., & Wu, W. (2020b). Forecast and Early Warning of Regional Bus Passenger Flow Based on Machine Learning. Mathematical Problems in Engineering, 2020, e6625435. https://doi.org/10.1155/2020/6625435.

Messinis, S., & Vosniakos, G. C. (2020). An agent-based flexible manufacturing system controller with Petri-net enabled algebraic deadlock avoidance. Reports in Mechanical Engineering, 1(1), 77-92.

Milenković, M., Švadlenka, L., Melichar, V., Bojović, N., & Avramović, Z. (2018). Sarima Modelling Approach For Railway Passenger Flow Forecasting. Transport, 33(5), 1113–1120.

Müller-Hannemann, M., Rückert, R., Schiewe, A., & Schöbel, A. (2022). Estimating the robustness of public transport schedules using machine learning. Transportation Research Part C: Emerging Technologies, 137, 103566. https://doi.org/10.1016/j.trc.2022.103566

Ni, M., He, Q., & Gao, J. (2017). Forecasting the Subway Passenger Flow Under Event Occurrences With Social Media. IEEE Transactions on Intelligent Transportation Systems, 18(6), 1623–1632. https://doi.org/10.1109/TITS.2016.2611644

Pamucar, D., Deveci, M., Canıtez, F., & Bozanic, D. (2020). A fuzzy Full Consistency Method-Dombi-Bonferroni model for prioritizing transportation demand management measures. Applied Soft Computing, 87, 105952. https://doi.org/10.1016/j.asoc.2019.105952

Rajendran, S., Srinivas, S., & Grimshaw, T. (2021). Predicting demand for air taxi urban aviation services using machine learning algorithms. Journal of Air Transport Management, 92, 102043. https://doi.org/10.1016/j.jairtraman.2021.102043.

Reitmann, S., & Schultz, M. (2022). An Adaptive Framework for Optimization and Prediction of Air Traffic Management (Sub-) Systems with Machine Learning. Aerospace, 9(2), 77, 1-15.

Rodríguez-Sanz, Á., de Marcos, A. F., Pérez-Castán, J. A., Comendador, F. G., Valdés, R. A., & Loreiro, Á. P. (2021). Queue behavioural patterns for passengers at airport terminals: A machine learning approach. Journal of Air Transport Management, 90, 101940. https://doi.org/10.1016/j.jairtraman.2020.101940

Roos, J., Gavin, G., & Bonnevay, S. (2017). A dynamic Bayesian network approach to forecast short-term urban rail passenger flows with incomplete data. Transportation Research Procedia, 26, 53–61.

Statista Demographics. (2020). https://www.statista.com/statistics/1101883/largest-european-cities/, Accessed 17 February 2022.

Sun, Y., Leng, B., & Guan, W. (2015). A novel wavelet-SVM short-time passenger flow prediction in Beijing subway system. Neurocomputing, 166, 109–121.

Tom Mitchell. (2006). The Discipline of Machine Learning. Pittsburgh, PA. http://ra.adm.cs.cmu.edu/anon/usr0/ftp/anon/ml/CMU-ML-06-108.pdf

Toqué, F., Khouadjia, M., Come, E., Trepanier, M., & Oukhellou, L. (2017). Short amp; long term forecasting of multimodal transport passenger flows with machine learning methods. 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC), 560–566. https://doi.org/10.1109/ITSC.2017.8317939

Traffic congestion ranking | TomTom Traffic Index. (2021). https://www.tomtom.com/en_gb/traffic-index/ranking/

TUIK. (2021). https://data.tuik.gov.tr/Bulten/Index?p=Adrese-Dayal%C4%B1-N%C3%BCfus-Kay%C4%B1t-Sistemi-Sonu%C3%A7lar%C4%B1-2020-37210&dil=1, Accessed 19 February 2022.

Wang, B., Wu, P., Chen, Q., & Ni, S. (2021). Prediction and Analysis of Train Passenger Load Factor of High-Speed Railway Based on LightGBM Algorithm. Journal of Advanced Transportation, 2021, ID 9963394, https://doi.org/10.1155/2021/9963394.

Wood, J., Yu, Z., & Gayah, V. V. (2022). Development and evaluation of frameworks for real-time bus passenger occupancy prediction. International Journal of Transportation Science and Technology. https://doi.org/10.1016/j.ijtst.2022.03.005

Xie, G., Wang, S., & Lai, K. K. (2014). Short-term forecasting of air passenger by using hybrid seasonal decomposition and least squares support vector regression approaches. Journal of Air Transport Management, 37, 20–26.

Yang, X., Xue, Q., Yang, X., Yin, H., Qu, Y., Li, X., & Wu, J. (2021). A novel prediction model for the inbound passenger flow of urban rail transit. Information Sciences, 566, 347–363.

Ye, Y., Chen, L., & Xue, F. (2019). Passenger Flow Prediction in Bus Transportation System using ARIMA Models with Big Data. 2019 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC), 436–443. https://doi.org/10.1109/CyberC.2019.00081

Zhang, J., Shen, D., Tu, L., Zhang, F., Xu, C., Wang, Y., Tian, C., Li, X., Huang, B., & Li, Z. (2017). A Real-Time Passenger Flow Estimation and Prediction Method for Urban Bus Transit Systems. IEEE Transactions on Intelligent Transportation Systems, 18(11), 3168–3178.

Zheng, Z., Ling, X., Wang, P., Xiao, J., & Zhang, F. (2021). Hybrid model for predicting anomalous large passenger flow in urban metros. IET Intelligent Transport Systems, 14(14), 1987–1996.