Passenger Volume Prediction

Ajay Kesarwani, Achraf Haddar,
Farah Ghaith, Amin Amini

Universität Passau, Passau, Germany
kesarw01@ads.uni-passau.de, haddar01@ads.uni-passau.de,
farah04@ads.uni-passau.de, aminim01@ads.uni-passau.de

Abstract. This paper studies the volume of bus passengers and predicts ticket sales for bus travel in Passau, Bavaria, Germany. We have conducted experiments based on the research question using data extracted from the Weather website and the Deutsche Bahn Region. We used random forest, ARIMA, and LSTM on the datasets to forecast future bus travellers. Using random forest we study the impact of three features temperature, is-weekend and holiday. We obtain 0.36, 0.36, 0.40, and 0.36 F1 scores on the dataset without using any features, with weather features, with the is-weekend feature, and with the holiday feature, respectively. The ARIMA model produces a 0.11 F1 score, whereas the LSTM experiment on each bus stop produces a mean macro F1 score between 0.2 and 0.4. We found that the output produced by the model was unsatisfactory, and one of the reasons could be that the data was imbalanced or a simple model was used.

Keywords: Random Forest, ARIMA, LSTM

1 Introduction

Passengers or travellers could be anyone who travels from one place to another either regularly or occasionally using public transport systems (i.e., Bus, Train, Metro, etc.). To improve the efficiency and effectiveness of the transportation system, roads, rail networks, and other infrastructure are built. Thereafter, the data are collected (i.e., the number of regular commuters, ticket sales per day, search results for routes from one location to another, traveller's routes, future demands, etc.) to predict passenger volume so that a well-planned design can be created to run buses, trains, etc., and better service can be provided on those routes.

The main problem is to provide better connectivity to the public so that people can travel comfortably from one place to another without any hassle and at the same time with the lowest possible costs (i.e., manpower, energy, money, etc.). It is possible to achieve this only by monitoring traffic routes, counting how many passengers got on and off buses at certain bus stops during a certain time of day, and observing external factors such as weather, gas prices, working days or weekends, booking app queries, etc. By analyzing this data, an efficient plan can be created where both passengers can get better service and the transport system could be improved for the city of Passau.

In this paper, we will study the bus transport system of Passau using data collected from the Deutscher Wetterdienst(DWD) [4] and Deutsche Bahn Regional, and, by analyzing the data, our focus will be to predict which routes, at what time intervals, and how many times the buses should run so that a maximum number of passengers can get the benefit.

There are a few research questions that we will be trying to find answers to with our experiments:

- **1. What effect does the weather have on passenger travel during a certain period of time?**
  Weather affects the transportation system. For instance, wet roads slow down vehicles, whether it is because of rain, snow, or other factors. Therefore, we should consider the weather when forecasting the passenger volume. In our experiment, we will collect the weather data for Passau from the DWD and monitor the passenger count on bad weather days. Our main goal would be to check whether passenger travel has any effects on weather factors, e.g., snow, wind, temperature, and precipitation. We will monitor hourly passenger count data from the Deutsche Bahn Region dataset and weather dataset. For instance, if it is the temperature, we will see how the passenger count changes for a particular hour of the day with the range of temperature values against the centigrade scale. e.g. We will collect passenger data for 10:00 hr from the entire year when the temperature range between 0-4 and will compare this data when the temperature range between 4-8 centigrade and so on. We would try to determine if there is any correlation with the passenger count if the temperature increases.

- **2. What is the impact of regular working days, weekends, and holidays on passenger counts during certain hours of the day?**
  Here, the Deutsche Bahn Region data will be used and divided into working days, weekends, and public holidays. Afterward, we will monitor how the passenger count changes hourly during working days compared to weekends, and the distribution of this data will be visualized with an appropriate graph.

- **3. For the above two questions, how do statistical models (i.e., ARIMA) and deep learning models (i.e., LSTM) predict passenger counts?**
  We will apply both ARIMA and LSTM models in the future prediction of passenger count as they are the most popular models used in many places, as mentioned in Section 2. Therefore, we would be using these two models on our above two research questions and trying to find which model gives a better performance based on the F1-Score value.

- **4. Could the location of the bus station be used to generate features such as population density?**

By considering the location of stations, clustering based on their location, and considering the population density of stations per area unit as population density, we want to study whether the location of stations can be a feature to increase the accuracy of the selected models in predicting future passenger volume.

This paper aims to answer these research questions using the two most popular statistical and deep learning models: Autoregressive Integrated Moving Average (ARIMA) model [6] and Long Short-Term Memory (LSTM) model[18]. The ARIMA model is used in most places due to the non-stationary property of the data, and again, the LSTM model also gives a better solution for many situations, for instance, in cases where a longer period of time is used for training the features of the data, or in cases where there is no linear relationship between features. Therefore, we will use both methods in our experiments and do a comparative study between them. Firstly, we will start by understanding the collected data from Deutsche Bahn Regional (i.e., seven CSV files) and weather data freely available from the DWD website, then we will pre-process this data so that passenger counts per week or month are aggregated to understand the change in behavior of passenger counts in any particular period. Thereafter, we will analyze the different features using exploratory data analysis that could answer our research questions and make our prediction more accurate. Afterward, we will use statistical and deep learning models to forecast passenger volumes.

2 Related Work

In recent years, many studies have been carried out on passenger flow forecasting. If we look at this problem from the perspective of the methodology model, the existing literature has applied parametric and non-parametric Time Series forecasting models to the traffic system. Junji Ji compares the GRU NN with ARIMA to Forecast Bus Trip Demand [11], Anila Cyril uses ARIMA to Model and Forecast Bus Passenger Demand [2], Yongxue Tian uses LSTM NN to predict Short-term Traffic Flow [20]. Forecasting the Origin destination flow of Bus transit has been employed in various research papers during the last few years. Researchers focused on employing different approaches to efficiently estimate inter-station passenger flow. Toque F. et al. proposed a method that uses LSTM RNN to forecast origin-destination matrices in subway networks. LSTM has shown the capability to provide accurate forecasting. However, LSTM models were not able to capture the passenger flow pattern from the long past since it was weakly utilized in the forecast process [22] [21]. These studies mainly consider the temporal influence on passenger flow but do not take other factors into account, such as weather conditions, etc. Other studies have looked at this issue from a data model perspective. In addition to the impact of time and travel history on passengers, these studies have tried to consider the effects of other environmental and social factors. Among these studies, it seems that weather has a higher percentage than other factors. With the use of a deep LSTM neural network, Lijuan Liu studied the impacts of weather on forecasting Metro Passenger Flows [13] and Tang et al. [19] studied the impact of rain on forecasting short-term passenger flow. This paper examines which forecasting methods offer the best predictions of Passenger Volume with respect to environmental factors such as weather. There are numerous stochastic models available in time series forecasting. Among the most widely used methods is ARIMA. In this paper, we compare ARIMA and LSTM models based on their performance on environmental factors. ARIMA is chosen because passengers' data is non-stationary, while LSTM is chosen because it is used in preserving and training the features of given data over a longer period of time.

3 Data Acquisition and Analysis

3.1 Data Collection:

The data has been collected by Deutsche Bahn (DB) Regional. DB is one of the largest transportation companies in Germany. DB Regional provides a regional transport network either by using trains or buses [1]. The dataset is aggregated to reflect ticket sales per hour per bus stop. In addition, to enforce the prediction, we have an additional document containing information about the route definitions and locations, including longitude and latitude for bus stops. In our study, we will use this dataset to predict ticket sales for bus travel in the region of Passau, Bavaria, Germany.
In order to study the impact of various weather conditions on the volume of passengers, we collected weather data from the Deutscher Wetterdienst(DWD) [4] which is freely available through the ftp server of DWD's Climate Data Center (CDC) [5]in the form of text files. The DWD has different stations for collecting different measurement parameters on an hourly, daily and monthly basis. For our study temperature, precipitation, and wind speed data related to station Fürstenzell has been used in the hourly frequency due to its proximity to Passau. In the collected Data missing values are marked as -999.0. We updated missing values with the value of the first available data before.

3.2 Data Analysis

The dataset consists mainly of seven csv files provided by DB Regional. The first file represents the on-demand travel dataset. The second file contains the regular travel dataset. The day of ticket sales, the exact hour of departure, the bus stop name, and the number of passengers are the features present in both of these csv files. Both csv files describe the number of passengers at 50 different bus stops. The on-demand travel dataset has ticket sales from January 1st, 2019 to December 31st, 2019 with a total number of rows equal to 438000. The regular travel dataset has ticket sales from the beginning of 2019 until the 15th of December of the same year. It contains 420000 rows in total. Both datasets have 12 different folds, each consisting of 3 weeks for training and 1 week for forecasting. For instance, the first three weeks in January are filled, however, the fourth week has Nans values in the Passengers column. Table 1 describes various statistics about the number of Passengers for regular and on-demand routes datasets. In addition, the third and fourth files, which are an extended version of the previous ones, contain additional features, which are the route reference as well as the target bus stop. They also include the precise departure date in the year-month-day hour-minute-second format. The fifth file is used to describe the bus stops, it contains several attributes such as a unique identifier, a cleaned-up name for fusion with other data, exact location, original name, longitude, latitude, and an abbreviation for a name if available. Moreover, the sixth CSV file is used to define the regular routes, it contains different features such as the original name of the stop, the reference of the route, and the cleaned-up name. The last csv file is used to describe queries posted to the "Wohin du willst" service in 2019, which allows users to specify departure and destination exactly via free text search for places. The file contains the exact timestamp of the query creation, the desired travel time, the place of departure, and the destination.

Dataset	Count	Mean	Std	Min	25%	50%	75%	Max
Regular routes	420000	0.514226	1.721261	0.0	0.0	0.0	0.0	47.0
On-demand routes	420000	0.105400	0.510013	0.0	0.0	0.0	0.0	26.0

Table 1. Passengers target feature statistics

To better comprehend the structure of the dataset as well as create and choose the best features that make data modeling easier, exploratory data analysis is used in this study. In our study, we can perform multiple analysis implementations, which should be interpretable. The different plots should answer most of our questions, which could be summarized as follows: [17]:

- **What is the distribution of the target?**
  Figure 1 illustrates the density distribution for passengers in regular and on-demand travel datasets. The passenger's density is skewed close to 0 for both datasets. In fact, about 60% of the Passengers column has 0 as a value for the regular travel dataset, and 70% for the on-demand travel dataset. Therefore, we transform the objective of this study from predicting exactly the passenger's value using a regression approach to a classification solution.

- **How do the sales vary overall for each bus stop and from one month to another?**
  Figures 2 and 3 illustrate the passenger volume variation throughout 2019 for regular travel as well as on-demand travel datasets respectively. For the regular travel dataset, all months seem to have almost the same passenger volume except the volume for the "Passau, HBF" point of departure which

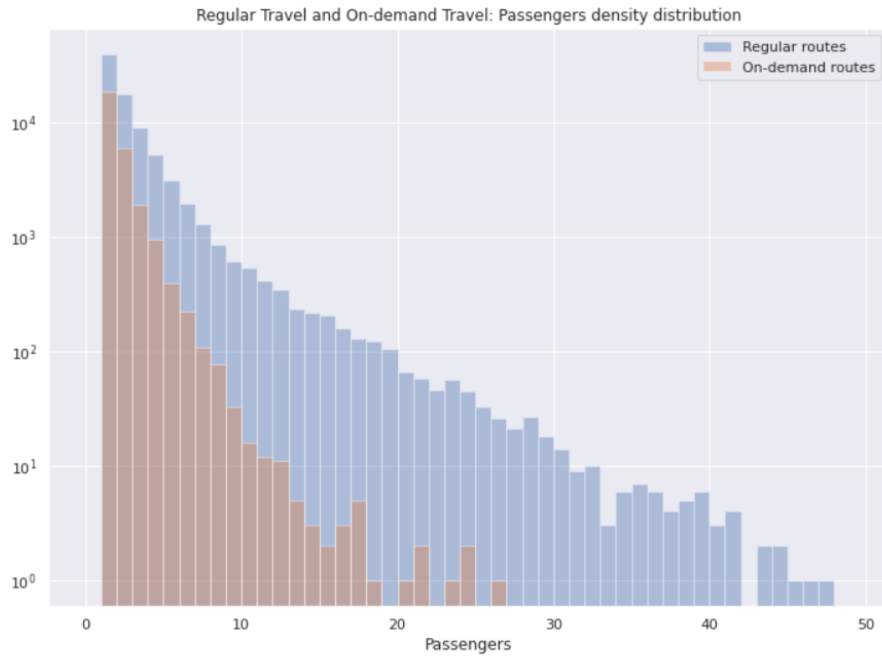

Fig. 1. Passenger density distribution for regular and on-demand routes

has a higher number of passengers. In addition, the variation in the number of passengers from one month to another for a specific bus stop is quite low. For the on-demand dataset, the passenger volume for the "Tirschenreuth, Marktplatz" bus station is slightly higher than other bus stops. Also, by interpreting the variations in the different graphs, we can assume that the total number of passengers is not constant from one month to another for each bus stop.

- **How do the sales behave for a specific month?**
  Figures 4 and 5 illustrate the variation of the average number of passengers grouped by all the bus stops during 2019 for both Regular and On-demand travel datasets. The figures demonstrate that more people buy tickets during working days; however, the amount decreases during the weekends. In fact, each month, the average number drops almost 4 times as the passenger volume decreases. For instance, on the 5th and 6th (Saturday and Sunday) of January, the average number of passengers falls from 0.6 to 0.1 and increases again on the 7th. The same applies to the On-demand travel dataset. Therefore, we could use a binary column that takes 1 if the day is a weekend as a feature for our training.
- **What is the impact of regular working days, weekends, and holidays on passenger counts during certain hours of the day?**

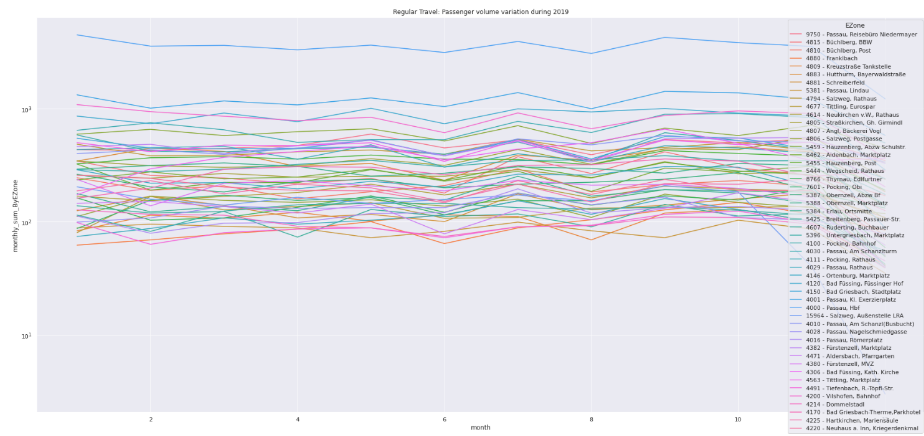

Fig. 2. Passenger volume variation during 2019 for regular routes

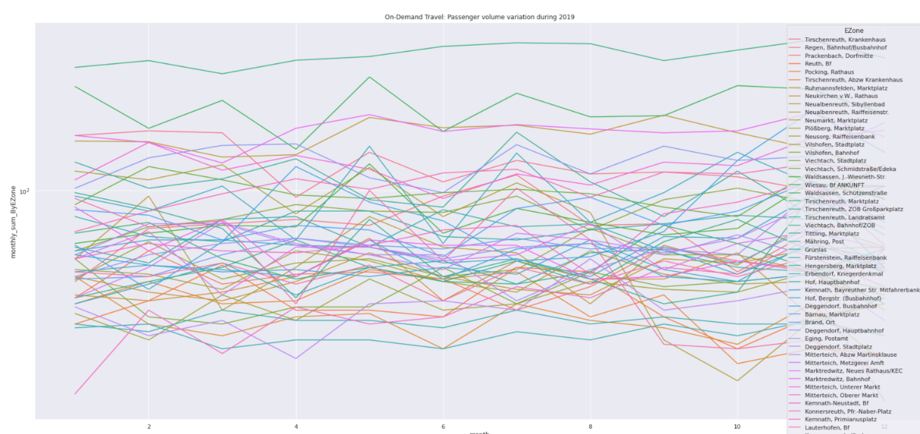

Fig. 3. Passenger volume variation during 2019 for on-demand routes

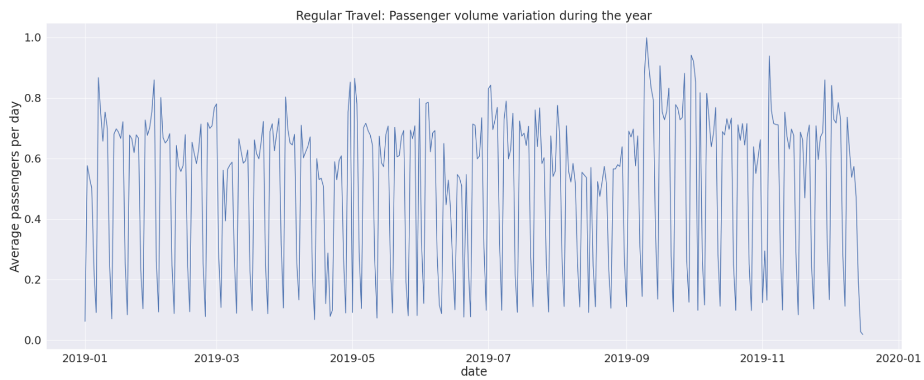

Fig. 4. Passenger average volume during 2019 for regular routes

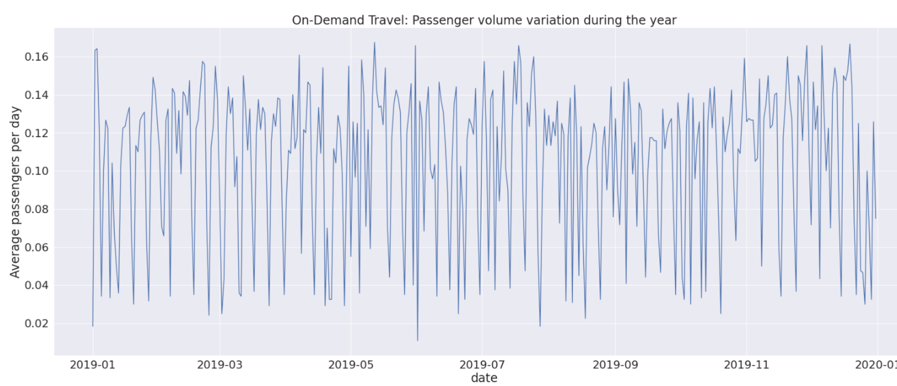

Fig. 5. Passenger average volume during 2019 for on-demand routes

Figures 6 and 7 illustrate the passenger's mean number per hour for workdays, weekends, and holidays for regular travel as well as on-demand travel datasets respectively. For regular traffic, the passenger mean number follows a Gaussian distribution, varying between 0.1 and 1.6 on workdays having the largest values around 8h and 13h. The mean of passenger takes almost a constant number between 6h and 17h which can describe the regular movement of people on workdays in fact roughly 90% of the travels are between 6h and 17h and takes null values at the late night and early morning contrary to weekends and holidays where the passengers mean number varies between 0.1 and 0.5 with an extra small number (around 0.1) of passengers at late night and a maximum reached around 9h which can explain the low number of ticket sales weekends and holidays. For the on-demand traffic, the passenger mean number varies between 0 and 0.34 during workdays reaching a maximum of 0.34 at 8h and then decreasing by the end of the day until reaching 0 late at night. For weekends the mean number of passengers didn't pass 0.2 having in general low values which is the same case for on-demand traffic in holidays where the passenger number varies strongly between the parts of the day but with low values also.

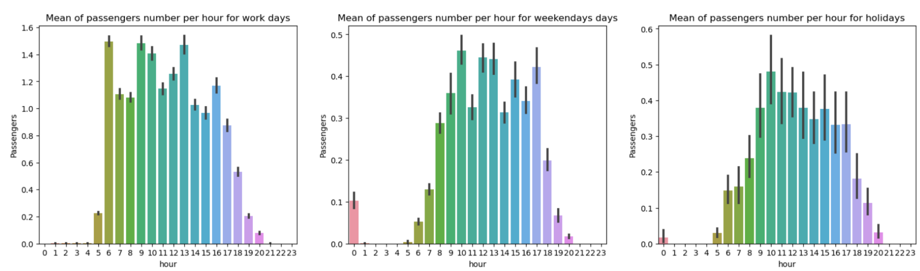

Fig. 6. Passenger mean variation between weekends, workdays, and holidays for the regular traffic

– **Does the exact longitude and latitude of the bus stop affect the passenger volume?**
To see the impact of the longitude and latitude of the position of the bus stop on the target variable, we used the KMeans clustering algorithm to convert the two geographic properties into more meaningful region labels. We select four as the number of clusters using the elbow method [16]. The left side of figure 8 describes the position of different bus locations on a geographic map. We use colors to differentiate between the four clusters. The right side of figure 8 confirms that cluster 1, i.e., the locations with the blue markers have the highest number of passengers. This can be confirmed by the fact that the bus stops in this cluster are closer to the center of Passau. Therefore,

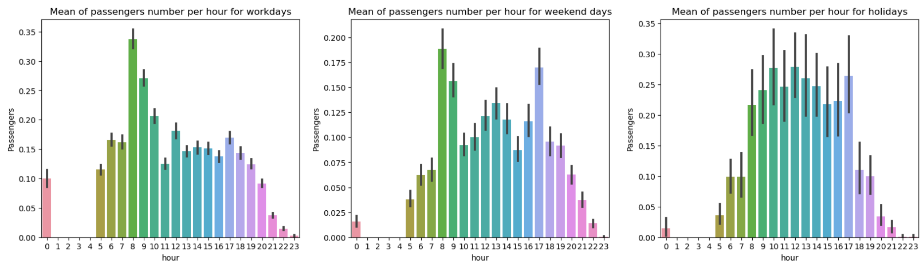

Fig. 7. Passenger mean variation between weekends, workdays, and holidays for the on-demand traffic

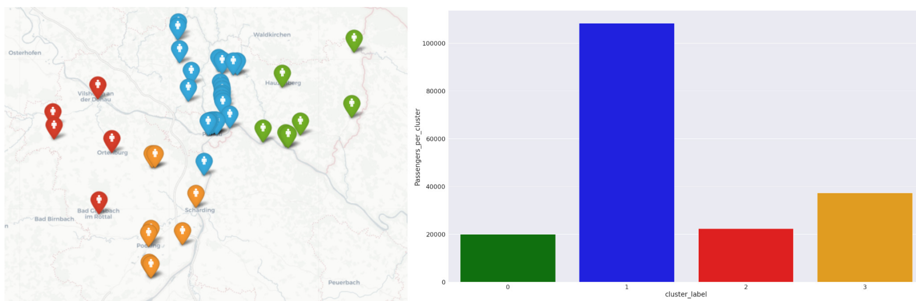

Fig. 8. Influence of longitude and latitude towards the number of passengers

the exact location of the bus stops can be used to help the model make better predictions.

- **Do people prefer to travel in extreme hot or cold weather?**
  Figures 9 and 10 show the volume of passengers at different temperatures for regular travel and on-demand travel dataset. For this purpose, the volume of passengers has been aggregated for different stations.

4 Pre-processing

- Data Discretization: The data is already skewed left as there are many bus stops with zero passengers; therefore, we will convert the regression problem into a classification problem by simply discretizing the target variable (number of passengers), where the real values will be replaced with their categorized or quantized ones 0,1,2 and 3, or more.

5 Feature Engineering

- The dominant part of the variability in passenger numbers is related to time, such as the hour of the day, the day of the week, weekdays and weekends, and the public holiday. For example, most days of the week have the same

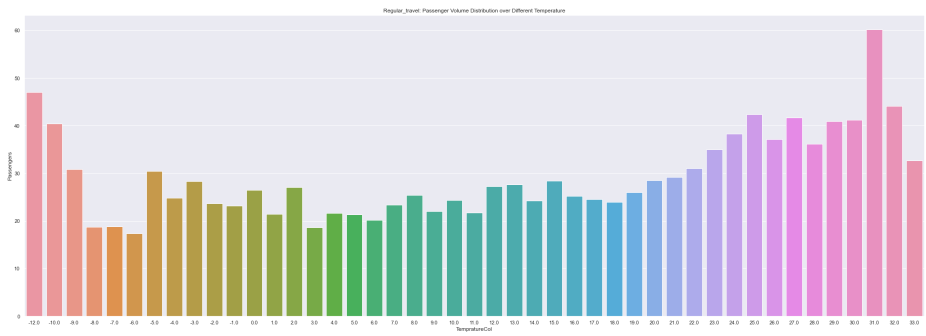

Fig. 9. Regular Travel: sum of Passengers in different Temperatures

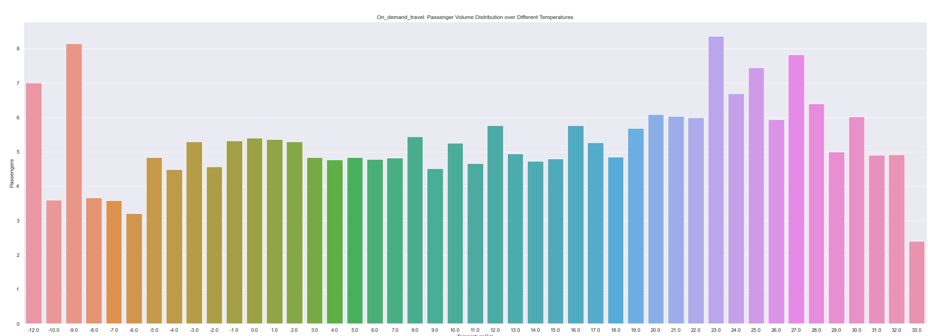

Fig. 10. on-demand Travel: sum of Passengers in different Temperatures

daily cycle, the night hour of Friday and Saturday are busier than usual, and Sunday has the lowest number of passengers, so to model these variations. We will add a binary feature that takes '1' if it's a weekend or "0" otherwise.
- Another additional factor is school holidays because passengers number drop significantly during peak hours due to the absence of students. We will cover this factor by adding another binary tag that takes "1" for holidays and "0" for non-holidays.

6 Methodology

Time series forecasting is a technical method used to predict a future event or trend by using the past during a predetermined period of time. Given its benefits, time series forecasting is used in various fields such as Economics [10], Finance [8], Energy [3] [14], and many more. There are two main categories for time series forecasting models. The first are parametric models, which use a fixed set of parameters that summarize the data [7]. The main objective of parametric techniques is to assess the time series model's parameters, which describe a stochastic process. The ARIMA model is one of the most popular parametric models. The second is non-parametric models which use the structure and state of the data to make future predictions without relying on any specific fixed set of parameters. LSTM is the most popular non-parametric model [7].

6.1 Modeling

ARIMA ARIMA is a stochastic model developed by Box and Jenkins. ARIMA is an extension of Auto Regressive Moving Average (ARMA) by adding an integration component. ARMA combines two main techniques. The first method is Auto Regressive which aims to predict future values by using historical data with the condition that the data should be stationary. The second approach uses the errors generated by the predictions of the first technique on the past given time series in order to do a smoothing process. Since ARIMA is a parametric model, this approach is based on three main settings (p, d, q) [6]:

- AR—p: It refers to the usage of historical values in the time series regression equation by taking advantage of the dependence between a current observation and observations from a preceding time.
- I—d: creates a stationary time series by using differences of observations. A series' current values are subtracted from its prior values d times
- MA—q: a model that makes use of the relationship between an observation and the residual error from a moving average model when applied to lag-timed data. q indicates how many previous error terms should be included in the model.

LSTM is a specification of a recurrent neural network that is commonly used for time series predictions. It uses the actual input and the memorized past knowledge to predict the future; it is suitable for applications that need long-term memories, therefore it solves the long-term dependency problem related to general RNNs. Each cell consists of a cell state unit and three gates used to control information in the memory model. The cell state represents the memory of the system. The forget gate takes a value between 0 and 1, and the sigmoid function is used to process information from the previous hidden state as well as information from the current input to decide what information should be discarded or saved. The closer to 1 means to save, and the closer to 0 means to forget. The input gate controls which new information will be saved in the cell state using a sigmoid in the first stage and then a tanh function. The cell state's output information is determined by the output gate. [18]

7 Implementation/Experiments

We have conducted several experiments to find the answer to our research question, as mentioned in Section 1. First, we studied the result without adding any features (i.e., weather data) by using a simple baseline model named random forest [9], and later we implemented LSTM [18] and ARIMA [6] models and compared their results. We first train a model using passenger datasets based on categories, such as without weekends, with weekends, and only holidays. We got different outputs for different training datasets and studied the results to deduce meaningful information from them. We also added other features, such as the weather feature, where we added temperature information to train the model to see whether it affects the output of the model when compare to the baseline model. In the end, we try to conclude our experiments by discussing how one model is better than another for predicting passenger volume information.

1. **First research question: What effect does the weather have on passenger travel during a certain period of time?**
   Second research question: What is the impact of regular working days, weekends, and holidays on passenger counts during certain hours of the day? :
   In our experiment, we used the random forest as a baseline model to answer our first and second questions. Random forest [9]is a machine learning algorithm that combines the output of multiple decision trees into a single result and uses it for both regression and classification problems. In fact, we have tried to train a random forest classifier using dataset with and without weather, is-weekend, and is-holidays features to evaluate their impact towards the training process.

   For the first case, we have trained the regular travel weekday data, which includes features such as date, hour, Ezone, and passenger data. This data is divided into training windows of size 21 and test windows of size 7. These steps are followed in all subsequent cases. Then we built a random forest model with class weights because the data is imbalanced (a passenger count value of 0 is much higher than any other passenger count value). In this experiment, we trained the model without weather, is-weekend, and is-holiday features. We obtained a macro-average f1-score value of 0.36. In the second experiment, we added the temperature feature to the existing dataset, and we got a macro-average f1-scores of 0.35. For the third and fourth experiments, we added respectively the is-weekend and is-holiday features and we obtained a macro-average f1-score equal to 0.40 and 0.36 respectively.

The table 2 presents the key metrics for each experiment, providing a clear and comprehensive overview of the results. It is clear that "is-weekend" feature gives a rough improvement regarding the performance of the model. Therefore, is-weekend could be considered as an important feature. On the other hand, our findings do not show any significant impact on the training performance of the model when incorporating the "weather" and "is-holiday" features. As such, these features may be disregarded.

Experiments	**Accuracy**	**Macro Average F1-score**	**Weighted Average F1-score**
Without weather, is-weekend, and is-holiday features	0.81	0.36	0.77
With weather feature	0.77	0.36	0.75
With is-weekend feature	0.82	0.40	0.78
With is-holiday feature	0.81	0.36	0.77

Table 2. Performance evaluation for random forest experiments

2. **Third research question: For the above two questions, how do statistical models (i.e., ARIMA) and deep learning models (i.e., LSTM) predict passenger counts?**

ARIMA model We have chosen randomly a bus stop and its respectiv dataset for the second month to train and test an ARIMA [6]model. In fact, the data was taken from the regular travel and was divided into 21 days for training and 7 days for testing. In order to train the model efficiently, we have added lags features such as the average value and the standard deviation for the last 24 and 48 hours. Using adfuller which is a statistical test used to determine whether a time series is stationary or non-stationary, we have found that the data is not stationary since the p-value is lower than 0.05. We have used the Auto Arima model to find the optimal parameters (p, d, and q). We have obtained a low macro f1-score equal to 0.11. The results of the ARIMA approach did not meet the desired standards and did not yield the expected outcomes. As a result, it has been decided to explore a more advanced method to attain better results. The next step is to proceed with testing another, more sophisticated approach which is training LSTM model.

LSTM model We have implemented 50 LSTM[18] models, one per bus stop. First of all, we categorize the input data as unique zone-wise so that we have zone-wise hourly passenger count values in a separate dataset. This data is then normalized, and each model is trained 12 times, in which we use three weeks of data as training and predict the output for the fourth week until the end of the year. The structures of the 50 models are the same: a sequential model with 2 layers, an LSTM layer with one unit, relu as an activation function, and then a dense layer with four units and a softmax activation function because our target classes are ranging from 0 to 3. Each model is then compiled using a categorical cross-entropy loss function and an Adam optimizer with a learning rate of 0.001 and a macro-F1-score as an evaluation metric. In each step (3 weeks of training and 1 week of testing), the models are fitted with a batch size of 10 and a number of epochs of 50, and a validation rate of 0.2. Those are some results of this implementation.

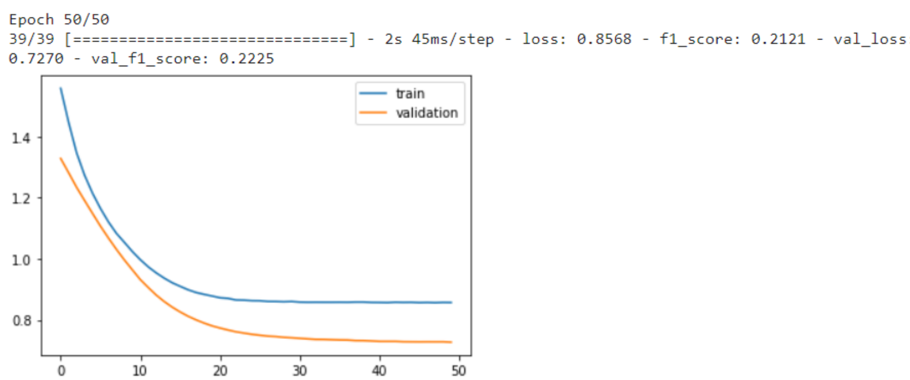

Fig. 11. Training and validation loss for one bus stop after 3 weeks of training (1st month)

The results of the remaining months, which are not visualized here, are quite similar to the figure 11 and figure 12.
The variation of the training and validation loss shows that the changing trend of the two variables is basically the same; they are decreasing as the number of epochs increases to reach a minimum of 0.65 for the validation error and 0.8 for the training error with a corresponding maximum f1-score of 0.225.
As the size of the training data set (3 weeks) in each step is considerably small, we have thought about using an accumulative size of training dataset

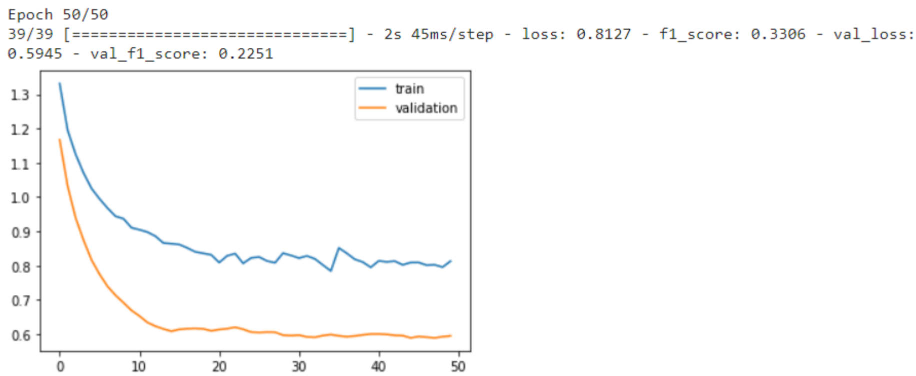

Fig. 12. Training and validation loss for one bus stop after 3 weeks of training(2nd month)

over the year by making use of all the previous data before a date t (the hourly data of 2019 before the first test date t) to train the model and predict the hourly passengers count for the fourth week of each month beginning from date t.

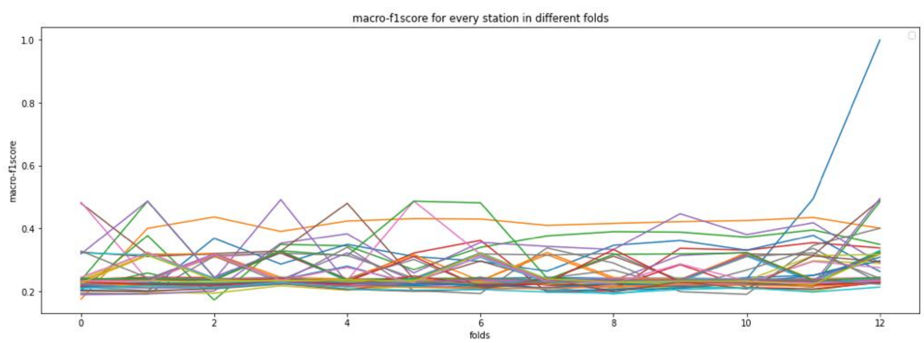

Fig. 13. macro-f1score of prediction for 13 different folds for every stations with accumulative training

As the structure of the model is Simple the results are not expected to be good. For our first experiment, We used a simple model because it allows faster training and testing as it has fewer parameters to adjust and monitor which can save time and resources, also it helps to prevent overfitting which can occur when the model is more complex.

Busstop	Min(f1score)	Max(f1score)	Mean(f1score)
15964	0.22	1	0.31
4000	0.24	0.44	0.4
4001	0.17	0.4	0.32
4016	0.22	0.49	0.26
4028	0.22	0.48	0.27
4029	0.22	0.486	0.30
4030	0.18	0.19	0.26
4100	0.19	0.24	0.21
4200	0.21	0.44	0.33
4214	0.21	0.49	0.31

Table 3. F1 score measures for LSTM trained on different busstops

Our main goal is to make the model converge and verify that is learning from the data. From pictures 11 and 12 we can conclude that the model converges and is not overfitted in view of the fact that the training and validation errors are decreasing as the number of epochs increases and converge roughly to the same value which means that the model is learning from the data and maybe it needs more time to learn and need to be more complex to capture more underlying patterns. Those results can be a good starting point for further experimentation and optimization.

8 Evaluation

To estimate model performance on new data, two criteria are used: Accuracy is defined as the ratio of the number of correct predictions to the total number of input samples. The closer accuracy is to one, the better. Accuracy is great, but gives a false sense of achieving high accuracy when the data is imbalanced so, we will also use the F1-score, which is defined as the harmonic mean of precision and recall. Since it is a multi-class classification problem, we will use the one vs. rest method to compute the F-1 score for each class. This approach scores the success of each class individually as if each class had a different classifier. Therefore, to calculate the F-1 score for class 'a', we first need to calculate the precision and recall values for that class. The precision measures the closeness between predictions, and the Recall represents the measure of how well the model can detect positive samples[12]:

$$Precision(class = a) = \frac{TP(class = a)}{TP(class = a) + FP(classa = a)} \tag{1}$$

$$Recall(class = a) = \frac{TP(class = a)}{TP(class = a) + FN(classa = a)} \tag{2}$$

$$F1 - score = \frac{2 * Precision(class = a) * Recall(class = a)}{Precision(class = a) + Recall(class = a)} \tag{3}$$

$$Accuracy = \frac{Number\ of\ correct\ predictions}{Total\ number\ of\ predictions} \tag{4}$$

The higher the F1 score, the better the performance of our model. The F1 score is a combination of both precision and recall, so a classifier gets a high F1-score only if both precision and recall are high.

9 Discussion and Future Work

Despite our efforts to accurately forecast passenger volume at bus stops, our approach has produced unsatisfactory results. The limitations of our methods and the limitations of the data used have resulted in predictions that are not in line with actual passenger volume.

9.1 ARIMA limitations

Since our dataset is highly imbalanced (containing too many 0), we thought about an approach to handle it. We skipped the oversampling and undersampling options because they can lead to loss of information and distorted temporal relationships between data points. Another technique is to use defined class weights; however, Arima does not provide this option. Therefore, the model was trained on imbalanced data, which may be one of the causes behind the poor predictions.
In addition, Arima can be sensitive to the choice of hyperparameters, especially when the time series has a complex trend and seasonality structure since ARIMA may not be able to capture it accurately leading to poor predictions.

9.2 LSTM limitations

Some of the causes behind getting poor results from the LSTM model could be the fact that we didn't initialize the weights properly. In fact, the performance of LSTMs can be sensitive to the initialization of the weights since improper initialization of the weights can lead to issues such as vanishing gradients, exploding gradients, and slow convergence. Therefore, it is a good idea to experiment with different initialization methods to find the best one for our problem.
Since LSTMs require a lot of hyperparameter tuning, including the learning rate, the number of layers and hidden units, the activation function, and the optimization function used, we tried to use a simple model that contains one LSTM layer with 1 unit and a ReLU activation followed by a dense layer with 4 units and a softmax activation function. In fact, we focused on achieving convergence rather than tuning hyperparameters as well as using large and complex models to attain higher performance. As a future work, we can test further LSTM architectures and hypertune properly all the hyperparameters using the most common techniques such as grid search, random search and bayesian optimization.

9.3 Conclusion

In this paper, we studied the passenger data of Passau to forecast future predictions. We experimented with three approaches, i.e., Random Forest, ARIMA, and LSTM models to answer our four research questions regarding the effect of weather (i.e., temperature) on passengers, the impact of passenger count on weekdays, weekends, and holidays during certain hours of the day, and how the location of a bus stop could be used to generate features such as population density. To analyze feature importance i.e., the impact of weather, weekend, and holidays we used the random forest, where we did four experiments on a given dataset, first without using any features such as weather, is-weekend, or holiday, then with the weather feature, then with the is-weekend feature, and finally with the is-holiday feature where we got the macro average score of 0.36, 0.36, 0.40 and 0.36 respectively. The ARIMA model was trained and tested on a given dataset that was determined to be stationary, and we obtained a low macro f1-score of 0.11. We did not achieve the desired standard result due to an imbalanced dataset (i.e. more passenger counts than others), a lack of class weights, and hyper-parameters sensitivity, so we use a more advanced method called LSTM. In LSTM we implemented 50 LSTM models, one per bus stop, and the result of the implementation was visualized by the figure using training and validation loss per bus stop. LSTM produced better results with a mean macro F1 score between 0.2 and 0.4 for the bus stops. The results of our experiment need to improve, as the F1-score values are below 0.5.

9.4 Future work

The challenges and limitations encountered in our approach highlight the need for continued research and development in this field to improve the accuracy and reliability of passenger volume predictions. Further investigation into alternative techniques, incorporating additional relevant data sources, and refining the preprocessing and modeling steps will be necessary to produce better results in the future. In fact, it is important to carefully consider the features to use when training the model. External factors can be useful to make capturing patterns easier for the model. We can state economic indicators, local events, or road closures, which can have an impact on passenger volume and therefore can be considered.
Answering our fourth question can be considered for future work. In fact, the location of bus stops can be used to obtain the population around it using, for example, Open Street Map API which provides a RESTful API that can be used to retrieve population density data. The API provides access to geospatial data, including population density information [15]. Therefore we can analyze the relation between the density and the number of passengers. In fact, the population density can indicate the potential demand for public transportation.
In our experiments we used LSTM which is a complex RNN model but more sample RNN models can be tested such as GRU and GURNN that can yields to better results.

References

[1] Deutsche Bahn. *DB Bahn Regional*. URL: https://www.deutschebahn.com/en/group/business_units/business__unit__db__regional-6929384.

[2] Anila Cyril, Raviraj H. Mulangi, and Varghese George. "Modelling and Forecasting Bus Passenger Demand using Time Series Method". In: *2018 7th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*. 2018, pp. 460–466. DOI: 10.1109/ICRITO.2018.8748443.

[3] Chirag Deb et al. "A review on time series forecasting techniques for building energy consumption". In: *Renewable and Sustainable Energy Reviews* 74 (2017), pp. 902–924. ISSN: 1364-0321. DOI: https://doi.org/10.1016/j.rser.2017.02.085. URL: https://www.sciencedirect.com/science/article/pii/S1364032117303155.

[4] *Deutscher Wetterdienst*. https://www.dwd.de/DE/Home/home_node.html.

[5] *Deutscher Wetterdienst Open Data Service*. https://www.dwd.de/EN/ourservices/opendata/opendata.html.

[6] Marilena Din. "ARIMA by Box Jenkins Methodology for Estimation and Forecasting Models in Higher Education". In: Jan. 2015. DOI: 10.13140/RG.2.1.1259.6888.

[7] Anjali Gautam and Vrijendra Singh. "Parametric Versus Non-Parametric Time Series Forecasting Methods: A Review". In: *Journal of Engineering Science and Technology Review* 13 (June 2020), pp. 165–171. DOI: 10.25103/jestr.133.18.

[8] Ma Hiransha et al. "NSE stock market prediction using deep-learning models". In: *Procedia computer science* 132 (2018), pp. 1351–1362.

[9] Tin Kam Ho. "Random decision forests". In: *Proceedings of 3rd international conference on document analysis and recognition*. Vol. 1. IEEE. 1995, pp. 278–282.

[10] Sheikh Mohammad Idrees, M. Afshar Alam, and Parul Agarwal. "A Prediction Approach for Stock Market Volatility Based on Time Series Data". In: *IEEE Access* 7 (2019), pp. 17287–17298. DOI: 10.1109/ACCESS.2019.2895252.

[11] Junyi Ji and Jun Hou. "Forecast on Bus Trip Demand Based on ARIMA Models and Gated Recurrent Unit Neural Networks". In: *2017 International Conference on Computer Systems, Electronics and Control (ICCSEC)*. 2017, pp. 105–108. DOI: 10.1109/ICCSEC.2017.8446813.

[12] Hung-Yi Lin. "Efficient classifiers for multi-class classification problems". In: *Decision Support Systems* 53 (June 2012), pp. 473–481. DOI: 10.1016/j.dss.2012.02.014.

[13] Lijuan Liu, Rung-Ching Chen, and Shunzhi Zhu. "Impacts of Weather on Short-Term Metro Passenger Flow Forecasting Using a Deep LSTM Neural Network". In: *Applied Sciences* 10 (Apr. 2020), p. 2962. DOI: 10.3390/app10082962.

[14] Hung Nguyen and Christian K. Hansen. “Short-term electricity load forecasting with Time Series Analysis”. In: *2017 IEEE International Conference on Prognostics and Health Management (ICPHM)*. 2017, pp. 214–221. DOI: 10.1109/ICPHM.2017.7998331.

[15] OpenStreetMap contributors. *Planet dump retrieved from https://planet.osm.org*. https://www.openstreetmap.org. 2017.

[16] Graciela Racolte et al. “Spherical K-Means and Elbow Method Optimizations With Fisher Statistics for 3D Stochastic DFN From Virtual Outcrop Models”. In: *IEEE Access* 10 (Jan. 2022), pp. 1–1. DOI: 10.1109/ACCESS.2022.3182332.

[17] Robert Shumway and David Stoffer. “Time Series Regression and Exploratory Data Analysis”. In: (Jan. 2011). DOI: 10.1007/978-1-4419-7865-3_2.

[18] Ralf Staudemeyer and Eric Morris. “Understanding LSTM – a tutorial into Long Short-Term Memory Recurrent Neural Networks”. In: (Sept. 2019).

[19] Liyang Tang et al. “Forecasting Short-Term Passenger Flow: An Empirical Study on Shenzhen Metro”. In: *IEEE Transactions on Intelligent Transportation Systems* PP (Nov. 2018), pp. 1–10. DOI: 10.1109/TITS.2018.2879497.

[20] Yongxue Tian and Li Pan. “Predicting Short-Term Traffic Flow by Long Short-Term Memory Recurrent Neural Network”. In: Dec. 2015, pp. 153–158. DOI: 10.1109/SmartCity.2015.63.

[21] Florian Toqué et al. “Forecasting dynamic public transport origin-destination matrices with long-short term memory recurrent neural networks”. In: *2016 IEEE 19th international conference on intelligent transportation systems (ITSC)*. IEEE. 2016, pp. 1071–1076.

[22] Dan Yang et al. “Urban rail transit passenger flow forecast based on LSTM with enhanced long-term features”. In: *IET Intelligent Transport Systems* 13.10 (2019), pp. 1475–1482.