iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** V **Month of publication:** May 2022

DOI: https://doi.org/10.22214/ijraset.2022.42386

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue V May 2022- Available at www.ijraset.com

Taxi Demand Forecasting Utilizing Neighbourhood Influence Based on Ensemble Learning

Apoorva Thakur[1], Dr. Sandeep Monga[2]
[1, 2]*Computer Science and Engineering Department, Oriental Institute of Science and Technology, Bhopal*

Abstract*: This Time series forecasting (TSF) assists in making better strategic decisions under uncertain circumstances so that financial crisis can be avoided, wise investments can be made, under/over contracting of utility can be avoided, staffs can be scheduled appropriately, service providers can provide better service, mankind can get prepared for natural disasters and many more. However, the accuracy in forecasting plays a vital role and achieving such is a challenging task owing to the vagueness and nonlinearity associated with most of the real world time series. Therefore, improving the forecasting accuracy has become a keen area of interest among the forecasters from different domains of science and engineering. In this work, a time series forecasting model for Chicgo taxi trip data is proposed. The proposed model utilizes neighbourhood influence in terms of taxi demand to predict taxi demand.

Keywords: Time series forecasting, Regularization, Regression

I. INTRODUCTION

Decision making is an integral part of our day-to-day life. We make individual, social and organizational decisions based on our past knowledge and experience. However, we often find it difficult to make the correct decision under uncertain circumstances. The difficulty further increases as the degree of uncertainty increases. Under such uncertain situations, knowledge of the future assists in making better decisions. For example, knowledge of the future assists financial organizations to evade financial catastrophe [1], individuals to invest wisely in the stock market [1], electric power industry to implement the smart grid [2], energy companies to make appropriate purchasing and selling of electricity in the balancing market [3], organizations to schedule staff [4], service providers to provide better service [4], retailers to timely supply the exact product to the appropriate locations by keeping sufficient inventory levels, mankind to get prepared for natural disasters like drought, flood, earthquake, cyclone, etc. The most likely future of a phenomenon can be predicted by making a systematic analysis on its past observation (Time Series) and such a process is known as time series forecasting (TSF) [5]. Since TSF plays a crucial role in making decisions, improving its accuracy has always remained a keen area of interest among the forecasters. To achieve better forecasting accuracy, the choice of a particular forecasting method and model plays a crucial role [6]. A forecasting method is a procedure to predict the future values by systematically analyzing the past observations. The forecasting methods may be judgmental, univariate or multivariate. In judgmental methods, forecasts are made based on individual judgment, perception, or some other additional information. In univariate methods, forecasts are made based on the analysis of prediction series while in multivariate methods, forecasts are made based on the analysis of one or more additional series along with the prediction series. The univariate methods are especially useful when less knowledge is available relating to the underlying process producing data or when there is no suitable method available which reveals the relationship between the prediction series and other series [6]. Therefore, univariate TSF methods have been widely used in several areas of study including signal processing, retailing, computational finance, econometric, weather forecasting, astronomy and mostly in the domains of applied science and engineering involving temporal measurements. Looking at the wide range of applications, in this thesis efforts are made to study and improve the performance of univariate forecasting methods.

A forecasting model is a tool to reveal the relationship existing in past data and using which the future values are predicted. The forecasting models follow either model driven or data driven approach. In model driven forecasting, experts having the deep understanding of underlying data choose a model based on scientifically established relationships, followed by which forecasting is done. Since it is very difficult to identify the mathematical function or model for many data generating processes (DGP) producing the time series, such choice of pre-specified model may not be appropriate to capture the complex and dynamic behavior of time series data. On the other hand, data driven forecasting models use the data to determine the model parameters and then use the identified model to predict the future values. Since in the later approach the model is built upon data, it glorifies the chances to capture the underlying complex patterns of time series data. However, the data driven models require adequate amount of data and robust mining and/or learning techniques to learn the underlying patterns efficiently. In this internet era availability of data is not a problem.

©IJRASET: All Rights are Reserved

Additionally, due to the emergence of fields like statistics, data mining and machine learning (ML), data driven forecasting models have gained tremendous attention in TSF. Traditionally, data driven statistical models like autoregressive integrated moving average (ARIMA), exponential smoothing have been widely used in TSF. This is because of the ease in understanding and implementing these models. The statistical models assume linear correlation structure in the time series and hence fail to capture the nonlinear patterns efficiently [6]. To add to this owe, most of the time series resulted from real world phenomena are nonlinear and have temporal as well as spatial variability [6]. Therefore, nonlinear statistical models like autoregressive conditional heteroskedasticity (ARCH) and generalized ARCH (GARCH) have been used to model nonlinear time series. The ARCH and GARCH family has thousands of models with each model capable of capturing a specific type of nonlinear pattern efficiently [5, 7]. Thus, it becomes complicated to select a suitable model without understanding the characteristics of time series. This limits the applicability of these models in general forecasting problems. On the other hand, data driven machine learning (ML) models have the capability to automatically extract unknown, novel, valid, and potentially useful patterns existing in data. In ML models, instead of fitting data to a pre-specified model, the data itself is used to identify the model and its parameters. This enhances the chances to capture the underlying complex relationships existing in the time series. Because of these advantageous characteristics, over the past few decades, a number of TSF methods based on ML models [1, 8-9] have been developed either to improve the forecasting accuracy or to reduce the computational complexity. Although the ML models like ANN have shown promising results in the empirical forecasting competitions like M3 and NN3 and can provide better result, ML models have not yet outperformed the most promising statistical models [9]. This demands a more systematic study on ML techniques for TSF.

II. LITERATURE REVIEW

Mohamday, Nazem and Tohme, Sadia (2011) used seasonal Time Series Models to forecast electrical power consumption in Fallujah City This research deals with using seasonal time series models to study and analysis of the monthly data on consumption of electricity in Fallujah city for the period (2005-2010), whereas these models are distinct with high accuracy and flexible in time series analysis. The results of the application show that the proper and efficient model for representing time series data is the multiplicative seasonal model of order: SARIMA(1, 1, 1)×(0, 1, 1)12.

Shah et al. (2014) analysed growth and trend in the area, production and yield of major crops of Khyber Pakhtunkhwa by using a time series data from 1980-81 to 2011-12 of major crops (wheat, maize, rice, and sugarcane). The compound growth rates, as well as trend analysis, indicated that the area under wheat crop has decreased over time due to shifting to other Rabi crops. The production of wheat during 1981-85 to 2011-12 was increased due to the corresponding increase in per hectare yield of the wheat crop in Khyber Pakhtunkhwa. The results showed that area, production, and yield of maize increased over time due to increased area under hybrid and improved open-pollinated maize varieties. The area under rice crop has decreased whereas their production increased due to the corresponding increase in per hectare yield of rice crop. It was revealed from the data that area, production, and yield of sugarcane crop was increased at a rate of 0.24 percent, 0.85 percent, and 0.60 percent per annum, respectively.

Rajaraman and Datta (2003) intended to estimate historical agricultural outputs for five states of India– Rajasthan, Karnataka, Punjab, Andhra Pradesh, and Uttar Pradesh – through the univariate ARIMA models. The researchers found similarity between the ARIMA methodology process for the states Andhra Pradesh, Punjab and Rajasthan in keeping the autoregressive and moving average terms that suit the best univariate process. Mulu and Tilahun (2009) conducted a study to examine the trends and generate a prediction model for health and its related indicators of Ethiopia from the period 1987 to 2000. The predictors of the established trends were obtained using ARIMA models developed in STATA. The study revealed that the mortality indicator viz Maternal Mortality Ratio exhibited a statistically significant reduction during the study period. Indicators like Total fertility rate, postnatal care, Physicians/100000 population, and skilled birth attendance were found to have a significant association with the Maternal Mortality Ration Trend. The research concluded that the existing trend required to accelerate the development of indicators to help reach the MDGs target with particular emphasis on Maternal health and provision of clean and safe water supply.

Tohme (2012) in their research to determine the number of malignant tumours infected patients applied the Time Series Analysis forecasting process incorporating the Box & Jenkins method. The research was conducted among the victims of malignant tumours in the province of Anbar using the dataset for the period (2006-2010). The result of the research revealed that ARIMA (2, 1, 0) was the best-suited model for the forecasting of patients affected by malignant tumours for two consecutive years on a monthly basis.

ARIMA models were applied by Balasubramanian and Dhanavanthan (2002) to acquire cyclical crop yield predictions. The research was based on historical data sets used in the assessment of area, production and yield rate of India and Tamil Nadu paddy food grains. The data were collected for three consecutive seasons viz; Kuruvaui, Samba, and Kodai along with information on Kharif and Rabi season seasons for both India and Tamil Nadu during the season 1966-99. The research concluded that seasonal ARIMA models were effective in retrieving an accurate forecast for the crop.

Kandiannan et al (2002) implemented the time series data analysis using the meteorological and yield data to forecast turmeric yield using based on 20 years data set. The data set used for the research belonged to the tenure (1979-80 to 1988-89) for the development of the forecast model and consecutive 10 years (1989-90 to 1998-99) used as a base for testing the model. The researchers were successful in developing a forecast for dry turmeric production and yield with 89% forecasting criterion value of viz co-efficient of determination. Nochai and Nochai (2006) attempted to develop the forecasting model for the palm oil price in Thailand for the period of five years (2000-2005), the oil types were separated into three main types; pure oil price, wholesale price and farm price. The researchers intended to fit the best ARIMA models for the forecast using the forecasting criterion Mean Absolute Percentage Error (MAPE). The study concluded that final three ARIMA model suited best for the three oil category viz; ARIMA (2,1,0) for the farm oil price, ARIMA (1,0,1) for wholesale oil price and ARIMA (3,0,0) for the pure oil price.

Peng, et al (2008) conducted a study to develop a short-term forecasting system to predict property crime for a city in China. ARIMA methodology was fitted using the 50 weeks historical time series data, the model was designed to predict the occurrence of crime before a period of one week. The finalized model was related to the SES and HES. The study result revealed that the ARIMA model displayed good forecasting and fitting accuracy that other techniques like exponential smoothing. This study can help law enforcement agencies and government bodies to predict decision-making and crime reduction. Rahman et al. (2013) conducted a study to compare growth patterns and ARIMA models to accurately forecast the production of chickpea, pigeon pea and field pea pulse in Bangladesh. The researcher applied the differencing order of 1 to stationarize the data set. The study comprised of two models namely ARIMA and growth models to find the best- fitted forecasting models. The results revealed that the best models were ARIMA (1, 1 and 1), ARIMA (0, 1 and 0) and ARIMA (1, 1 and 3) for pigeon pea, chickpea and field pea pulse production, respectively. Among the deterministic type growth models, the cubic model was the best for pigeon pea, chickpea, and field pea pulse production forecasting.

III. DATASET DESCRIPTION

A. Chicago Taxi Trips Records

This dataset has been collected from the official data portal handled by the Department of Business Affairs & Consumer Protection of the city of Chicago. It contains information about taxi flow between different community areas, consisting of around 195 Millionrows of pickup and drop off locations per 15 minutes for the past four years (2016-19). The average number of records per year is around 49 Million.

IV. PROPOSED MODEL

The proposed model aims at following (Refer Fig. 1):

1) A systematic hotspot analysis will be done on historical taxi demand data to capture the patterns and perform spatio-temporal analysis.
2) Based on spatio-temporal analysis, appropriate spatial and temporal resolution will be identified to train the models.
3) Selection of appropriate regression model will be performed by suitable sensitivity analysis.
4) Robustness of proposed model and suitability in real world will be tested by performing sensitivity analysis.

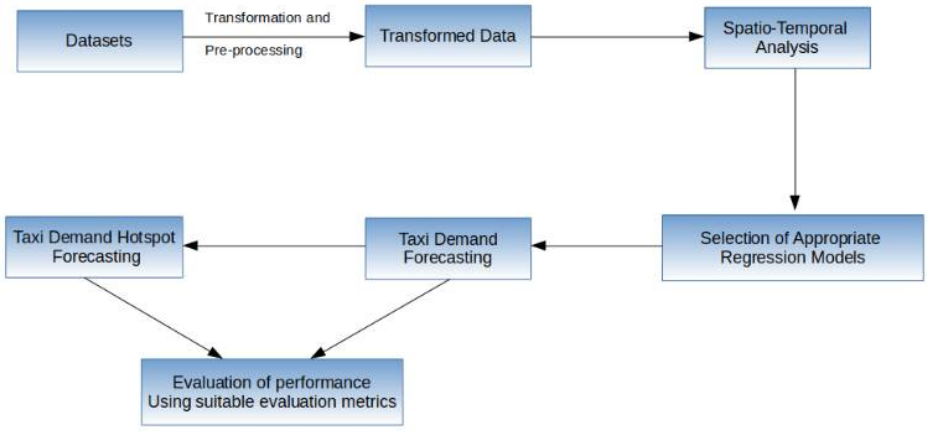


Figure 1: Block diagram representing flow of proposed work.

Algorithm 1 Demand Count Data transformation algorithm.

```
procedure DemandCountDataTransformation(intervalDemand)
    for CA in CA_List do
        for row in intervalDemand do
            if row[PickupCommunityArea] == CA then
                intervalDictionary[row.intervals] = row[Demand]
            CA_List[CA_Count] = intervalDictionary[values]
        CA_Count = CA_Count + 1
    demandCountMatrix = Transpose(CA_List)
    return demandCountMatrix
```

In Chicago city, there are 77 community areas. The taxi demand in these areas is heterogeneous; that is, the demand is considerably high in some areas while it falls short in other areas. This fluctuation in demand gives rise to prompt prediction of spatial characteristics of taxi demand to identify the hotspots and thus redirect the taxis in the areas where the demand is high. Another problem being the dynamic nature of taxi demand concerning the temporal dimension. If we consider hours of the day, demand may be high during working hours whereas comparatively lower during early morning and late-night (Refer Fig. 2a). Another perspective can be to look into the variability of demand between weekdays and weekends and across different months of the year (Refer to Figs. 2B and 2c). This accounts for the problem of the prediction of temporal characteristics of taxi demand. Therefore, to address the spatio-temporal heterogeneity, the 77 community areas are considered for spatial resolution. Furthermore, the prediction in each area is done on an hourly basis and then aggregated monthly, which is deliberated as the temporal resolution of demand prediction (Refer Fig. 2b). We are dealing with monthly prediction since the yearly aggregation will exclude the seasonality effect, i.e., the variation of demand due to different seasons. Also, to capture the seasonality effect due to differences in demand over weekdays and weekends, the demand pattern is analyzed separately. This spatio-temporal resolution of prediction results may aid in identifying hotspots. Since the demand varies with both space and time, it induces dynamism in the pattern of hotspots. Hence, hotspots will also vary from month to month and year to year. In this paper, we have proposed various dynamic demand prediction models based on the application of various statistical and ensemble regression models to forecast the taxi demand. These models can be further used to study the demand pattern inculcating the temporal feature and spatial characteristics and eventually to identify the dynamic pattern of hotspots.

Algorithm 2 NPBM algorithm.

```
procedure NPBM(t_CA, all_CA)
    for CA_i in all_CA do
        for CA_j in t_CA do
            if boundary(CA_i, CA_j == True) and CA_i != CA_j then
                i_CA_i^t ← CA_i
    for target_t in t_CA do
        predict_t^target ← regressor(i_CA_i^t)
    return predict_t^target, i_CA^t
```

For a given community area, the taxi demand may also be affected by the distance of the drop-off destination from the pickup location. If the destination is distant, the passenger may prefer to travel by other convenient transport like metro and buses. Thus, neighborhood proximity may play an essential role in determining the demand in a given community area. To instill this idea of neighborhood influence, we proposed a proximity-based prediction model.

The official data portal of Chicago city comprises taxi trip records from the year 2013 to the current year, from which we extracted the recent 4-year data from 2016-19 for this study. Out of the 23 attributes present in this dataset as discussed in the Table I, we considered attributes Trip Start Timestamp and Pickup Community Area for the transformation phase. Initially, we dropped all those rows which had missing values for Pickup Community Area. This raw data consisted of taxi flow records between community areas for every interval of 15 minutes which we aggregated on an hourly basis into a data frame interval Demand.

We then converted the obtained data into a demand count matrix (demandCountMatrix) whose columns represented the 77 community areas, whereas rows represented the count of taxi demand for each interval of a day for the given year. To implement the targeted transformation, we first stored each interval of the day into a dictionary (intervalDictionary) as keys and initialized its corresponding values as 0. Then, for every community area, we iterated over the interval Demand and mapped the taxi demand into the matching interval of dictionary value and consequently obtained the final demand count matrix of size t x c, where t is the total number of intervals in a day multiplied by total days in a year, and c is the total number of community areas (in this case, 77). The intervals for which there was no value for taxi demand remained 0 as initialized. These steps can be summed up by the Demand Count Data Transformation Algorithm 1. CA list is the list of all community areas and CA Count is the taxi demand in that CA during the given interval.

Forecasting of taxi demand is done by capitalizing the taxi booking data generated in the past, consisting of booking records of community areas to predict the number of taxi demands that might happen in the future in those particular areas. This forecasting problem is, in fact, a multivariate regression problem where the selection of independent variables is quite decisive. The pivotal decision is whether to consider all the community areas as independent variables or consider an intelligent subset that is highly correlated to the target variable. We proposed a model based on the interdependence of the target community area with its surrounding area. The model is termed as Neighbourhood Proximity Based Model (NPBM), which is based on the concept that the taxi demand in a given community area can be forecasted precisely if we consider the areas located in its vicinity as an influencing factor. This spatial characteristics based model is discussed in Section VII. In a broader sense, the algorithm for taxi demand forecasting models works by taking each community area once as a target variable and applying the models mentioned above for determining the vector of independent variables for it. The prediction of taxi demand count in the target community area is eventually made by employing various statistical and ensemble state-of-the-art regression models as base regressors.

Table 1: Table showing R2 values for NPBM of the year 2019

Months		Linear					Ensemble				
		OLS	LASSO	LARS	BR	RIDGE	ET	RF	BAGGING	STACKING	VOTING
January	Weekdays	0.4156	0.4106	0.2328	0.4160	0.4156	**0.7192**	0.5689	0.7158	0.7135	0.7158
	Weekends	0.5496	0.5125	0.4339	0.5512	0.5496	**0.6789**	0.5445	0.6570	0.6542	0.6570
February	Weekdays	0.5050	0.4833	0.3478	0.5051	0.5050	**0.6834**	0.6218	0.6643	0.6617	0.6642
	Weekends	0.5556	0.4977	0.3531	0.5200	0.5556	0.6810	0.5786	0.6903	0.6877	**0.6902**
March	Weekdays	0.2686	0.2835	0.1966	0.2703	0.2686	**0.7198**	0.4899	0.6477	0.6473	0.6474
	Weekends	0.5536	0.6637	0.3319	0.5548	0.5536	0.7156	0.6060	**0.7196**	0.7177	0.7188
April	Weekdays	0.3266	0.3509	0.0947	0.3278	0.3266	0.6958	0.5881	0.6743	0.6742	**0.6748**
	Weekends	0.1948	0.2030	0.2084	0.1964	0.1948	**0.7513**	0.5819	0.6816	0.6774	0.6802
May	Weekdays	0.4561	0.4388	0.2030	0.4561	0.4561	**0.6335**	0.5790	0.6311	0.6291	0.6311
	Weekends	0.5514	0.5849	0.4175	0.4148	0.4143	**0.6769**	0.5622	0.6309	0.6288	0.6305
June	Weekdays	0.2562	0.2562	0.1499	0.2567	0.2562	**0.7077**	0.6080	0.7005	0.6999	0.7011
	Weekends	0.3322	0.3636	0.1452	0.3335	0.3322	0.6491	0.5607	**0.6821**	0.6797	0.6809
July	Weekdays	0.4355	0.4221	0.2718	0.4356	0.4355	0.6148	0.5951	**0.6043**	0.6009	0.6040
	Weekends	0.5085	0.4595	0.4974	0.5088	0.5085	**0.6292**	0.5167	0.6189	0.6161	0.6183
August	Weekdays	0.4248	0.4082	0.3000	0.4249	0.4248	**0.6683**	0.6036	0.6468	0.6443	0.6470
	Weekends	0.5279	0.4626	0.4712	0.5278	0.5279	0.6034	0.5146	**0.6258**	0.6232	0.6257
September	Weekdays	0.3732	0.3560	0.1616	0.3731	0.3732	**0.6386**	0.6291	0.6274	0.6264	0.6275
	Weekends	0.5668	0.4985	0.5585	0.5667	0.5668	**0.6622**	0.5448	0.6544	0.6527	0.6544
October	Weekdays	0.3436	0.3475	0.1286	0.3440	0.3436	**0.6715**	0.5691	0.6669	0.6660	0.6674
	Weekends	0.1716	0.1991	0.1442	0.1749	0.1716	0.6284	0.4240	0.6536	0.6498	**0.6542**
November	Weekdays	0.4190	0.4114	0.2994	0.4194	0.4190	0.6128	0.5923	**0.6289**	0.6278	0.6285
	Weekends	0.3857	0.3977	0.3988	0.3866	0.3857	**0.6187**	0.4731	0.5984	0.5963	0.5985
December	Weekdays	0.5288	0.4354	0.3974	0.4322	0.4321	0.6756	0.5853	0.6895	0.6889	**0.6899**
	Weekends	0.3345	0.4911	0.3783	0.4727	0.4768	**0.6757**	0.4223	0.5893	0.5833	0.5901

As discussed in Section I, there is a need to consider neighborhood influence for forecasting the taxi demand of a given community area. To achieve this goal, we proposed a neighborhood proximity-based regression model where the selection of independent variables corresponding to a target variable are selected directly based its immediate neighbors (Refer Algorithm 2). Let t CA be the target variable representing the community area whose demand is to be predicted temporally, then i CA will be a vector of independent variables consisting of community areas sharing boundaries with t CA.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

We plotted a heatmap as shown in Fig. 3 to find the correlation of a given community area with other areas. It can be inferred from the plot that the areas located in the immediate neighborhood of a given community area are more correlated to it than those far away which strengthens the use of such a model for independent variable selection. The model is tested on the taxi demand of the year 2019 and R2 value of the this model for mentioned base regressors are shown in Table II. The taxi demands are predicted on an hourly basis to meet the real-world requirements, but results are aggregated monthly. The taxi demands are filtered based on weekdays and weekends to capture the variation in spatio- temporal patterns of taxi demands due to change in routines of people on weekends. It is to be noted that results of linear models and ensemble models are separated in order to illustrate that ensemble models will provide better predictions (Refer values shown in bold in Table II) to the problem as compared to linear models. The reason being ensemble models provides better predictions as the samples are operated on diverse models. The comparison of average R2 values of linear and ensemble models when tested for taxi demand of the year 2018 for each month can be done clearly with the help of Fig. 4. It can be inferred that ensemble models outperforms the linear models.

V. CONCLUSIONS

Taxi demand forecasting is a challenging task, as the taxi demands may have variable spatio-temporal patterns. In this work, a spatio-temporal analysis of taxi demands in Chicago will be performed. Based on such analysis appropriate spatial and temporal resolutions will be obtained. According to temporal resolution, taxi demand data might be split for training the regression models. After suitable testing of base models, predicted data can be be used for predicting hotspots. The predicted hotspots will be matched with actual hotspots to find the accuracy of prediction. In this work, a novel model termed as neighborhood proximity-based regression model is proposed and for ensemble models it can be inferred from the results that the proposed model achieves better accuracy as compared to non ensemble models.

REFERENCES

[1] T. Kim, S. Sharda, X. Zhou, and R. M. Pendyala, “A stepwise interpretable machine learning framework using linear regression (lr) and long short-term memory (lstm): City-wide demand-side prediction of yellow taxi and for-hire vehicle (fhv) service,” Transportation Research Part C: Emerging Technologies, vol. 120, p. 102786, 2020. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0968090X20306963

[2] Q. Liu, C. Ding, and P. Chen, “A panel analysis of the effect of the urban environment on the spatiotemporal pattern of taxi demand,” Travel Behaviou and Society, vol. 18, pp. 29–36, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2214367X19300936

[3] C. Yang and E. J. Gonzales, Modeling Taxi Demand and Supply in New York City Using Large-Scale Taxi GPS Data. Cham: Springer International Publishing, 2017, pp. 405–425. [Online].

[4] H. Luo, J. Cai, K. Zhang, R. Xie, and L. Zheng, “A multi-task deep learning model for short-term taxi demand forecasting considering spatiotemporal dependences,” Journal of Traffic and Transportation Engineering (English Edition), 2020. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S209575641830521X

[5] B. Hu, S. Zhang, Y. Ding, M. Zhang, X. Dong, and H. Sun, “Research on the coupling degree of regional taxi demand and social development from the perspective of job-housing travels,” Physica A: Statistical Mechanics and its Applications, vol. 564, p. 125493, 2021. [Online]. Available:https://www.sciencedirect.com/science/article/pii/S0378437120307913

[6] S. Faghih, A. Shah, Z. Wang, A. Safikhani, and C. Kamga, “Taxi and mobility: Modeling taxi demand using arma and linear regression,” Procedi Computer Science, vol. 177, pp. 186 – 195, 2020, the 11th International Conference on Emerging Ubiquitous Systems and Pervasive Networks (EUSPN 2020) / The 10th International Conference on Current and Future Trends of Information and Communication Technologies in Healthcare (ICTH 2020) Affiliated Workshops. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1877050920322948

[7] Z. Liu, H. Chen, Y. Li, and Q. Zhang, “Taxi demand prediction based on a combination forecasting model in hotspots,” Journal of Advanced Transportation, vol. 2020, p. 13, 2020.

[8] I. Markou, F. Rodrigues, and F. C. Pereira, “Multi-step ahead prediction of taxi demand using time-series and textual data,” Transportation Researc Procedia, vol. 41, pp. 540 – 544, 2019, urban Mobility - Shaping the Future Together mobil.TUM 2018 - International Scientific Conference on Mobilit and Transport Conference Proceedings. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S2352146519305113

[9] C. Antoniades, D. Fadavi, and A. F. Amon, “Fare and duration prediction : A study of new york city taxi rides.”

[10] A. Safikhani, C. Kamga, S. Mudigonda, S. S. Faghih, and B. Moghimi, “Spatio-temporal modeling of yellow taxi demands in new york city using generalized star models,” International Journal of Forecasting, vol. 36, no. 3, pp. 1138–1148, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0169207018301468

[11] J. Xu, R. Rahmatizadeh, L. Boloni, and D. Turgut, “Real-time prediction of taxi demand using recurrent neural networks,” IEEE Transactions on Intelligent Transportation Systems, vol. 19, no. 8, pp. 2572–2581, Aug 2018.

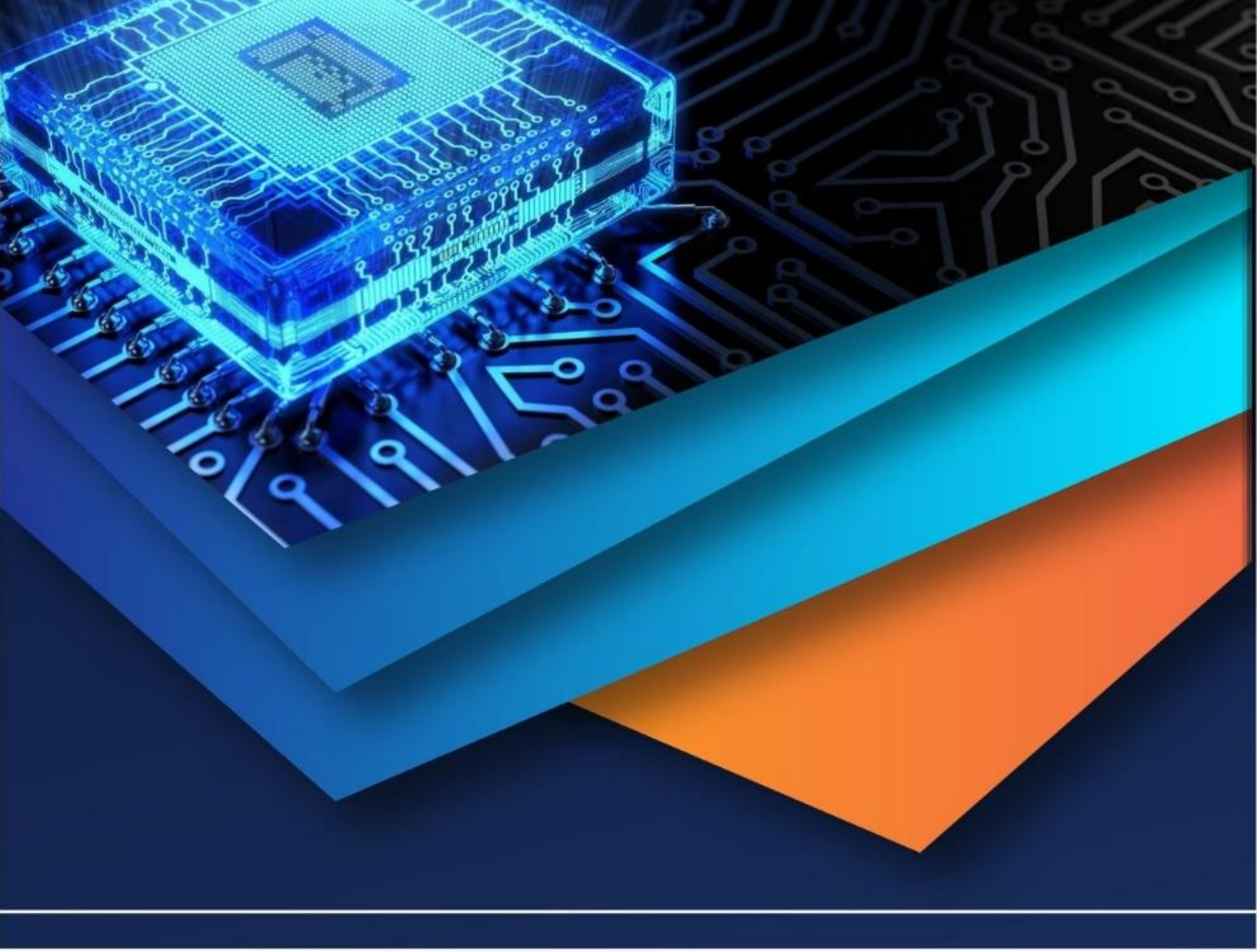
[12] T. Liu, W. Wu, Y. Zhu, and W. Tong, “Predicting taxi demands via an attention-based convolutional recurrent neural network,” Knowledge- Based Systems, vol. 206, p. 106294, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S095070512030469X

[13] P. Shu, Y. Sun, Y. Zhao, and G. Xu, “Spatial-temporal taxi demand prediction using lstm-cnn,” in 2020 IEEE 16th International Conference on Automation Science and Engineering (CASE), 2020, pp. 1226–1230.

[14] J. Ye, L. Sun, B. Du, Y. Fu, X. Tong, and H. Xiong, Co- Prediction of Multiple Transportation Demands Based on Deep Spatio-Temporal Neural Network. New York, NY, USA: Association for Computing Machinery, 2019, pp. 305 – 313. [Online]. Available: https://doi.org/10.1145/3292500.3330887

[15] U. Vanichrujee, T. Horanont, W. Pattara-atikom, T. Theeramunkong, and T. Shinozaki, "Taxi demand prediction using ensemble model based on rnns and xgboost," in 2018 International Conference on Embedded Systems and Intelligent Technology International Conference on Information and Communication Technology for Embedded Systems (ICESIT-ICICTES), 2018, pp. 1–6.
[16] Z. Liu, H. Chen, X. Sun, and H. Chen, "Data-driven real-time online taxi-hailing demand forecasting based on machine learning method," Applied Sciences, vol. 10, no. 19, 2020. [Online]. Available: https://www.mdpi.com/2076-3417/10/19/6681
[17] Z. Chen, B. Zhao, Y. Wang, Z. Duan, and X. Zhao, "Multitask learning and gcn-based taxi demand prediction for a traffic road network," Sensors, vol. 20, no. 13, 2020. [Online]. Available: https://www.mdpi.com/1424-8220/20/13/3776
[18] T. L. Quy, W. Nejdl, M. Spiliopoulou, and E. Ntoutsi, "A neighborhood- augmented lstm model for taxi-passenger demand prediction," in Multiple-Aspect Analysis of Semantic Trajectories, K. Tserpes, C. Renso, and S. Matwin, Eds. Cham: Springer International Publishing, 2020, pp. 100–116.
[19] X. Guo, "Prediction of taxi demand based on cnn-bilstm-attention neural network," in Neural Information Processing, H. Yang, K. Pasupa, A. C.- S. Leung, J. T. Kwok, J. H. Chan, and I. King, Eds. Cham: Springer International Publishing, 2020, pp. 331–342.
[20] P. Rodrigues, A. Martins, S. Kalakou, and F. Moura, "Spatiotemporal variation of taxi demand," Transportation Research Procedia, vol. 47, pp. 664 – 671, 2020, 22nd EURO Working Group on Transportation Meeting, EWGT 2019, 18th - 20th September 2019, Barcelona, Spain. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S2352146520303446
[21] Q. S. G. G. Xinmin Liu, Lu Sun, "Multitask learning and gcn-based taxi demand prediction for a traffic road network," Sensors, vol. 20, no. 13, 2020. [Online]. Available: https://www.mdpi.com/1424-8220/20/13/3776
[22] Y. Zhou, Y. Wu, J. Wu, L. Chen, and J. Li, "Refined taxi demand prediction with st-vec," in 2018 26th International Conference on Geoinformatics, 2018, pp. 1–6.
[23] D. Faial, F. Bernardini, E. M. Meza, L. Miranda, and J. Viterbo, "A methodology for taxi demand prediction using stream learning," in 2020 International Conference on Systems, Signals and Image Processing (IWSSIP), 2020, pp. 417–422.
[24] N. Davis, G. Raina, and K. Jagannathan, "A multi-level clustering approach for forecasting taxi travel demand," in 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC), 2016, pp. 223–228.
[25] T. Liu, W. Wu, Y. Zhu, and W. Tong, "Predicting taxi demands via an attention-based convolutional recurrent neural network," Knowledge- Based Systems, vol. 206, p. 106294, 2020. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S095070512030469X
[26] Mohamday, Nazem and tTohme, Sadia (2011) "Using Seasonal Time Series Models to Forecast Electrical Power Consumption in Fallujah City". Anbar University Journal of Economic and Administrative Sciences, Vol.8 No.7, pp.21-41.
[27] Shah, S. M. Ullah, A., & Hadi, F. (2014). Ecological characteristics of weed flora in the wheat crop of Mastuj valley, District Chitral, Khyber Pakhtunkhwa, Pakistan. Pak. Journal of Weed Science Res, 20(4), 479-487.
[28] Rajaraman, I., & Datta, A. (2003). Univariate forecasting of state-level agricultural production. Economic and Political weekly, 1800-1803.
[29] Mulu W.A and Tilahun H.N (2009), Modeling trends of health related indicators in Ethiopia (1995-2008); A Time series study : Health Research policy and Systems 2009,7:29, 2008 Ghana Millenium Development Goals report (April 2010).
[30] Tohme, (2012),'Using Analysis of Time Series to Forecast numbers of the patients with Malignant Tumors in Anbar province'. Anbar university Journal of Economic and administrative Sciences. Vol(4), No.(8),371-392.
[31] Balasubramanian, P., and P. Dhanavanthan. (2002) "Seasonal modeling and forecasting of crop production." Statistics and Applications 4(2), 107-118.
[32] Kandiannan, K., Chandaragiri, K. K., Sankaran, N., Balasubramanian, T. N., & Kailasam, C. (2002). Crop–weather model for turmeric yield forecasting for Coimbatore district, Tamil Nadu, India. Agricultural and Forest Meteorology, 112(3-4), 133-137.
[33] Peng Chen, Yuvan, H (2008). Forecasting crime using the ARIMA model, Fifth International Conference on Fuzzy systems and knowledge Discovery, October 18-20, 2008, Jian, Shandong, China, pp: 627-630

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)