Avrupa Bilim ve Teknoloji Dergisi
Özel Sayı, 34, S. 458-462, Mart 2022

Araştırma Makalesi

www.ejosat.com ISSN:2148-2683

European Journal of Science and Technology
Special Issue 34, pp. 458-462, March 2022

Research Article

Prediction of Road Visibility Based on Meteorological Parameters by Machine Learning Methods

Tahsin Baykal*[1], Fatih Ergezer[2] Ekinhan Eriskin[3] and Serdal Terzi[2]

[1*] 1 Suleyman Demirel University, Graduate School of Natural and Applied Sciences, 32260, Isparta, Turkey, (ORCID: 0000-0001-6218-0826), tahsinbaykal@gmail.com

[2] Suleyman Demirel University, Engineering Faculty, Department of Civil Engineering, 32260, Isparta, Turkey, (ORCID: 0000-0001-8034-5743, 0000-0002-4776-824X), fatihergezer@sdu.edu.tr, serdalterzi@sdu.edu.tr

[3] Suleyman Demirel University, Property Protection and Security Department, 32260, Isparta, Turkey, (ORCID: 0000-0002-0087-0933), ekinhan.eriskin@gmail.com

(2nd International Conference on Applied Engineering and Natural Sciences ICAENS 2022, March 10-13, 2022)

(**DOI:** 10.31590/ejosat.1082868)

ATIF/REFERENCE: Baykal, T., Ergezer, F., Eriskin, E. & Terzi, S. (2022). Prediction of Road Visibility Based on Meteorological Parameters by Machine Learning Methods. *European Journal of Science and Technology*, (34), 458-462.

Abstract

One of the important parameters in ensuring traffic safety is road visibility. Road visibility depends on the geometric design of the road, lighting conditions, as well as the climatic conditions in the area where the road passes. Visibility depends on meteorological parameters such as temperature, humidity, wind speed, pressure, fog, precipitation type. In this study, it is aimed to predict road visibility to ensure traffic safety. Machine learning methods were used for road visibility estimation. Machine learning models were developed with Random Forest, Extra Tree and Gradient Boosting methods. In the models, 96453 meteorological data sets such as temperature, humidity, wind speed, pressure, precipitation types, visibility were used between 2006 and 2016 in Szeged, Hungary. Developed models were evaluated with coefficient of determination (R^2) and Root mean squared error (RMSE). As a result of the evaluation, the random forest method gave the best result.

Keywords: road safety, visibility, machine learning, intelligent transport system, meteorological parameters

Makine Öğrenmesi Yöntemleri ile Meteorolojik Parametrelere Dayalı Yol Görünürlüğü Tahmini

Özet

Trafik güvenliğini sağlamada en önemli parametrelerden birisi yol görünürlüğüdür. Yol görünürlüğü yolun geçtiği bölgedeki iklim koşullarının yanı sıra, yolun geometric tasarımına ve aydınlatma koşullarına bağlıdır. Görünürlük, sıcaklık, nem, rüzgar hızı, basınç, sis, yağış tipi gibi meteorolojik parametrelere bağlıdır. Bu çalışmada, trafik güvenliğini sağlamak için yol görünürlük tahminlemesi yapılmıştır. Yol görünürlük tahmini için Makine öğrenme metotları kullanılmıştır. Makine öğrenme metotları Random Forest, Extra Tree ve Gradient Boosting yöntemleri ile geliştirilmiştir. Modelde, 2006-2016 yılları arasında Macaristan'ın Szeged şehrindeki sıcaklık, nem, rüzgar hızı, basınç, yağış tipi, görünürlük gibi 96453 meteorolojik data seti kullamılmıştır. Geliştirilen modeller belirtme katsayısı (R^2) ve Karesel ortalama hata (KOH) ile değerlendirilmiştir. Değerlendirme sonucunda random forest methodu en iyi sonucu vermiştir.

Anahtar Kelimeler: Yol güvenliği, görünürlük, makine öğrenmesi, akıllı ulaşım sistemleri, meteorolojik parametreler.

1. Introduction

Road visibility on highways is also an important parameter to prevent accidents. Road visibility conditions vary according to the geometric design of the road, lighting conditions, and the current climatic conditions of the region where the road passes. Low visibility conditions can increase the probability of traffic accidents and thus endanger traffic safety (Babari et al., 2012). According to the U.S. Department of Transportation Federal Highway Administration (2022), parameters such as fog, heavy precipitation, and wind that create low visibility conditions increase the risk of traffic accidents. Every year, 38700 traffic accidents occur due to foggy weather conditions and cause many deaths and injuries. Visibility is affected by meteorological conditions such as temperature, humidity, wind speed, pressure, fog, rain, snow, dust, and smoke (Jonnalagadda and Hashemi, 2020; Kwon, 2004). In the literature studies, the use of image processing algorithms for visibility estimation depending on meteorological conditions (Yang et al., 2018; Negru and Nedevschi, 2013), the use of neural network algorithms (Zhao et al., 2021), the use of Deep Learning approach (Chaabani et al., 2018; Yufeng et al., 2022), the use of Machine Learning approach (Lakshmi et al., 2017; Cornejo-Bueno et al., 2021; Ortega et al., 2019) related studies were examined.

In the study, 96453 atmospheric data sets such as temperature, humidity, wind speed, pressure, precipitation types, visibility were used in the city of Szeged, Hungary, between 2006 and 2016. Visibility estimation was made by using Random Forest, ExtraTree, Gradient Boosting algorithms, and Machine Learning approach depending on meteorological conditions on highways.

2. Material and Method

2.1. Dataset

96453 datasets of meteorological parameters obtained between 2006-2016 belonging to the city of Szeged in Hungary were used within the scope of the study (URL-1, 2022).

2.2. Machine Learning

In this study, Random Forest, Extra Tree, and Gradient boosting were used in machine learning methods. The algorithms used are described below.

2.2.1. Random Forest

Tree algorithms, similar to human consideration, start with a single node. They take the form of a tree structure as the algorithm progresses. Tree algorithms are used to solve regression and classification problems (Uyanık et al., 2021). The random forest (RF) algorithm proposed by Ho (1995) splits the data set into decision trees. Analyzes according to decision trees. Evaluates each decision tree and gives a common forecast result (Uyanık et al., 2021). The estimation process of random forest regression is shown in Figure 1.

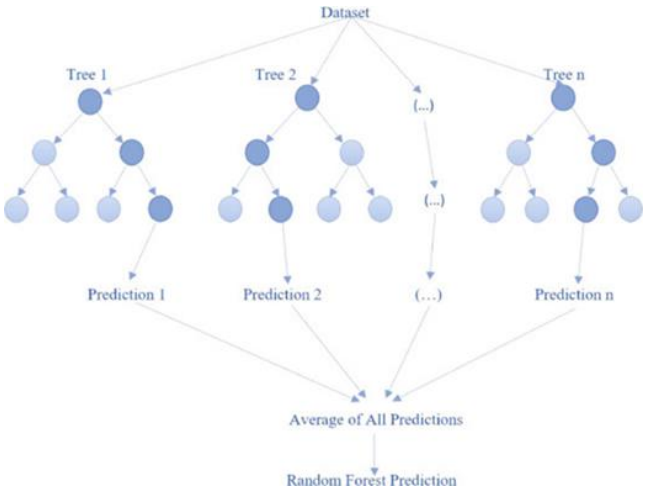

Fig. 1 The estimation process of random forest regression (Uyanık et al., 2021)

2.2.2. Extra Tree:

Geurts et al. (2006), the Extra Tree Regression (ETR) approach is an approach originally derived from the RF model. The ETR algorithm creates a collection of unpruned decisions or regression trees (Hammed et al., 2021). The two main differences between ETR and other tree-based methods are that ETR splits nodes by choosing random breakpoints and grows the tree using the whole learning sample. Tree estimates are combined with the arithmetic mean in regression problems to produce the final estimate (Jhon et al., 2016).

2.2.3. Gradient boosting:

The Gradient Boosting algorithm introduced by Friedman (2001) creates separate functions from the prediction made in the first iteration and the differences between the actual and the predictions values. Prediction success is boosted if functions are attempted to synchronize with each refresh (Uyanık et al., 2021).

3. Results and Discussion

In this study, hours, summary, precipitation type, temperature, apparent temperature, humidity, wind speed, wind bearing, pressure, and daily summary were used for visibility estimation with machine learning (ML). Statistical parameters of the data used are given in Table 1.

Table 1. Statistical parameters of the data

	Count	Mean	Std	Min	Max
Hours	96453	11.50	6.92	0.00	23.00
Summary	96453	3.10	0.98	1.00	10.00
Precip Type	96453	1.12	0.34	1.00	3.00
Temperature	96453	11.93	9.55	-21.82	39.91
Ap. Temperature	96453	10.86	10.70	-27.72	39.34
Humidity	96453	0.73	0.20	0.00	1.00
Wind Speed	96453	10.81	6.91	0.00	63.85
Wind Bearing	96453	187.51	107.4	0.00	359.00
Pressure	96453	1003.2	116.9	0.00	1046.4
Daily Summary	96453	3.18	0.47	1.00	7.00
Visibility	96453	10.35	4.19	0.00	16.10

First, correlation matrix and pair plot were used to examine the relationships of each data with the others. The correlation matrix of the data set is given in Figure 2 and the pair plot is given in Figure 3.

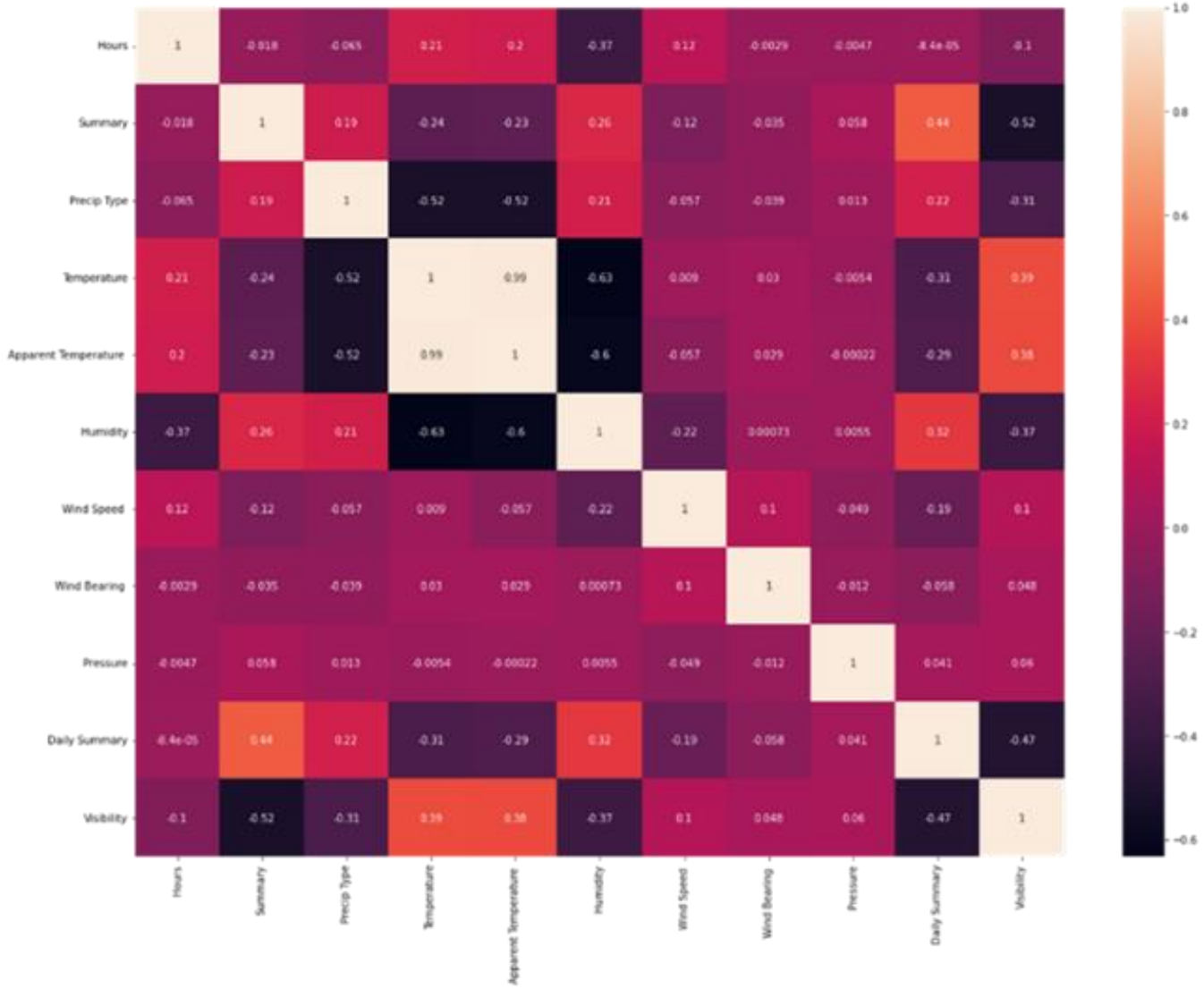

Fig. 2 Correlation matrix of the data set

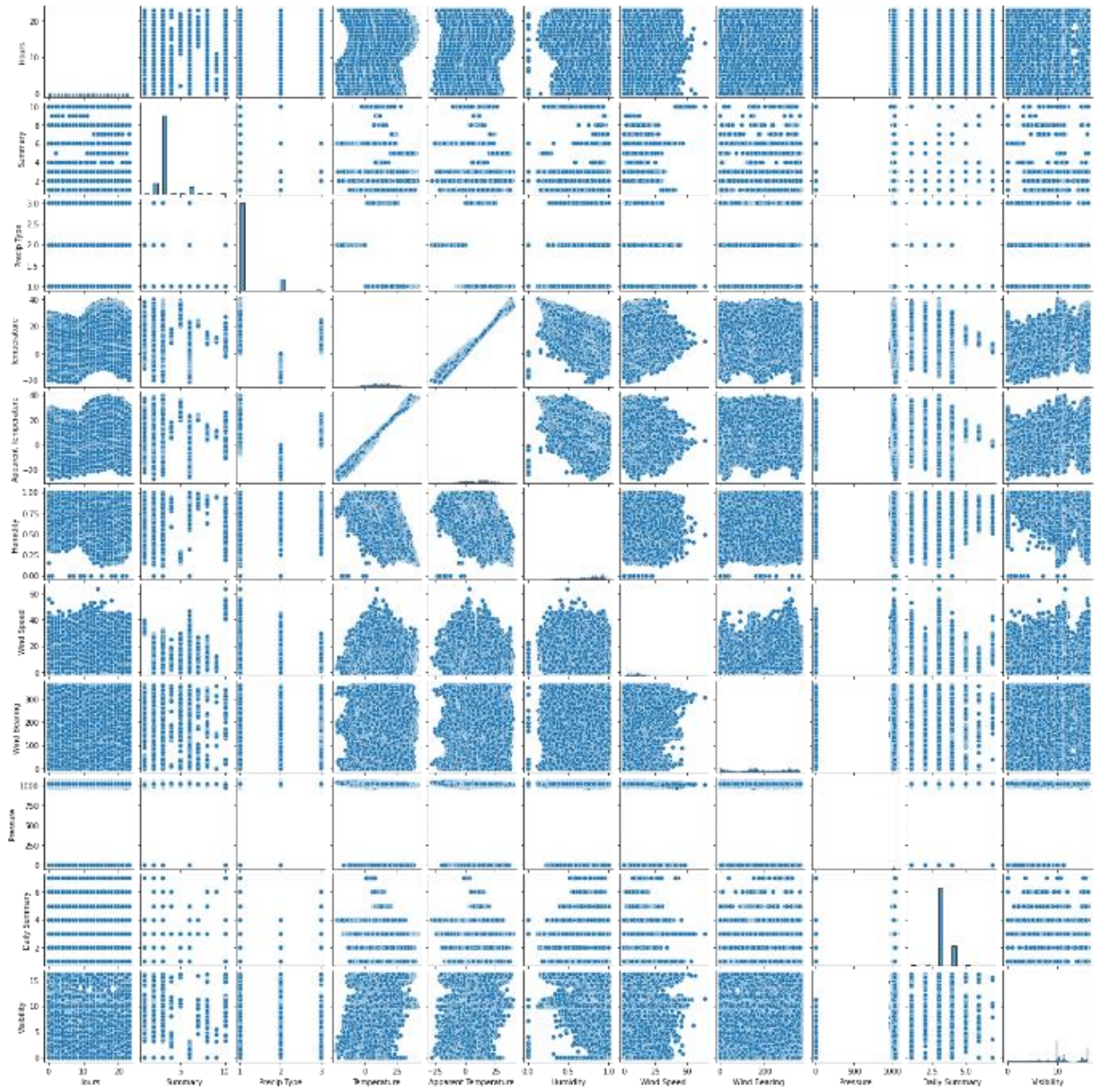

Fig. 3 Pair plot of the data set

When Figures 1 and 2 are examined, it is seen that the highest positive correlation with visibility is given by temperature with 0.39. It was determined that there was a highly negative relationship between summary and visibility.

Then the data set is randomly divided into 80% training and 20% test set. ML models were developed by using Random Forest, Extra Tree, and Gradient Boosting algorithms for the separated data sets. The Python programming language Scikit-learn 1.0.2 library was used to run the ML models. The developed model results were evaluated with a coefficient of determination (R^2) and Root mean squared error (RMSE). R^2 and RMSE are given in Eq. 1 and 2.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - y_i')^2}{\sum_{i=1}^{n}(y_i - y_m)^2} \quad (1)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - y')^2} \quad (2)$$

Where n is the size of the dataset, y_i are the actual values, y_m is the mean of the actual values, and y_i' is the predicted values.

The results of the models are given in Table 2.

Table 2. Results of the models

Algorithms	Train R^2	Test R^2	Train RMSE	Test RMSE
Random Forest	0.96	0.70	0.85	2.30
ExtraTree	1.00	0.69	0.00	2.33
Gradient Boosting	0.66	0.66	2.44	2.46

When Table 2 is examined, it is seen that the random forest algorithm gives the best result. The R^2 values for the training and test sets of this algorithm are 0.96 and 0.70, respectively.

4. Conclusions and Recommendations

Road visibility is an important factor in ensuring traffic safety. Road visibility is dependent on climatic conditions as well as road geometry and lighting conditions. In this study, road visibility based on meteorological parameters was estimated by machine learning method. In the estimation of road visibility, 96453 meteorological data sets such as temperature, humidity, wind speed, pressure, precipitation types, visibility were used between 2006 and 2016 in Szeged, Hungary. Machine learning models were developed with Random Forest, ExtraTree and Gradient Boosting algorithms. When the developed models were evaluated, the Random Forest algorithm gave the highest R2 value for the test set. The R^2 value of this algorithm is 0.70.

5. Acknowledge

The first author (T. Baykal) was supported by the Council of Higher Education's 100/2000 doctoral scholarship.

References

Babari, R., Hautière, N., Dumont, É., Paparoditis, N., & Misener, J. (2012). Visibility monitoring using conventional roadside cameras–Emerging applications. Transportation research part C: emerging technologies, 22, 17-28.

U.S. Department of TransportationFederal Highway Administration, (2022). Low visibility. Available: https://ops.fhwa.dot.gov/weather/weather_events/low_visibility.htm

Jonnalagadda, J., & Hashemi, M. (2020, August). Forecasting atmospheric visibility using auto regressive recurrent neural network. In 2020 IEEE 21st International Conference on Information Reuse and Integration for Data Science (IRI) (pp. 209-215). IEEE.

Kwon, T. M. (2004). Atmospheric visibility measurements using video cameras: Relative visibility.

Yang, L., Muresan, R., Al-Dweik, A., & Hadjileontiadis, L. J. (2018). Image-based visibility estimation algorithm for intelligent transportation systems. IEEE Access, 6, 76728-76740.

Negru, M., & Nedevschi, S. (2013, September). Image based fog detection and visibility estimation for driving assistance systems. In 2013 IEEE 9th International Conference on Intelligent Computer Communication and Processing (ICCP) (pp. 163-168). IEEE.

Zhao, J. I. N., Kangjun, Q. I. U., & Miaomiao, Z. H. A. N. G. (2021). Investigation of Visibility Estimation Based on BP Neural Network. Journal of Atmospheric and Environmental Optics, 16(5), 415.

Chaabani, H., Werghi, N., Kamoun, F., Taha, B., & Outay, F. (2018). Estimating meteorological visibility range under foggy weather conditions: A deep learning approach. Procedia Computer Science, 141, 478-483.

Yufeng, W., Du Jiamin, Y. Z., Yuehui, S., & Dengxin, H. (2022). Atmospheric visibility prediction by using the DBN deep learning model and principal component analysis.

Lakshmi, C. R., Rao, D. T., & Rao, G. S. (2017, September). Fog detection and visibility enhancement under partial machine learning approach. In 2017 IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI) (pp. 1192-1194). IEEE.

Cornejo-Bueno, S., Casillas-Pérez, D., Cornejo-Bueno, L., Chidean, M. I., Caamaño, A. J., Cerro-Prada, E., ... & Salcedo-Sanz, S. (2021). Statistical Analysis and Machine Learning Prediction of Fog-Caused Low-Visibility Events at A-8 Motor-Road in Spain. Atmosphere, 12(6), 679.

Ortega, L., Otero, L. D., & Otero, C. (2019, April). Application of machine learning algorithms for visibility classification. In 2019 IEEE International Systems Conference (SysCon) (pp. 1-5). IEEE.

URL-1, (2022). kaggle website. [online]. Available. https://www.kaggle.com/budincsevity/szeged-weather

Uyanık, T., Karatuğ, Ç., & Arslanoğlu, Y. (2021). Machine learning based visibility estimation to ensure safer navigation in strait of Istanbul. Applied Ocean Research, 112, 102693.

Ho, T. K. (1995, August). Random decision forests. In Proceedings of 3rd international conference on document analysis and recognition (Vol. 1, pp. 278-282). IEEE.15

Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. Machine learning, 63(1), 3-42.

Hammed, M. M., AlOmar, M. K., Khaleel, F., & Al-Ansari, N. (2021). An Extra Tree Regression Model for Discharge Coefficient Prediction: Novel, Practical Applications in the Hydraulic Sector and Future Research Directions. Mathematical Problems in Engineering, 2021.

John, V., Liu, Z., Guo, C., Mita, S., & Kidono, K. (2015, November). Real-time lane estimation using deep features and extra trees regression. In Image and Video Technology (pp. 721-733). Springer, Cham.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. Annals of statistics, 1189-1232.