

## AIR TEMPERATURE ESTIMATION FOR BATMAN CITY WITH SIMPLE AND MULTI-LINEAR REGRESSION MODELS UTILIZING METEOROLOGICAL PARAMETERS

*Emrullah ACAR<sup>1\*</sup>*


*Determination of air temperature has a significant role in numerous activities such as agriculture, animal husbandry, industry, highway, airlines and railway transportation. In this study, the monthly average of 67 meteorological parameters, which affects the temperature between 2012 and 2017, has taken from Batman Provincial Directorate of Meteorology and the monthly average air temperature of 2017 has been estimated using the meteorological data from 2012-2016. The estimation process has been carried out using two separate scenarios. In the first scenario, each parameter such as monthly average soil temperature, pressure, water vapour pressure, wind speed and relative humidity have been used in the simple linear regression model as input separately and the monthly average temperature has been estimated. In the second scenario, all 67 parameters have been employed in multi linear regression model as inputs and monthly average temperature has been estimated by this way. As a result, very low root mean square error (RMSE) values has been observed in the range of  $RMSE = [3.30- 10.55]$  while very high correlation coefficient ( $R^2$ ) values has been computed in the range of  $R^2 = [0.10- 0.99]$ .*

*Key words: Linear and Multi-Linear Regression; Estimation; Air Temperature; Meteorology;*

### 1. Introduction

Air temperature (AT) is one of the main meteorological observations and it plays an important role in an extensive range of implementations such as climate change, terrestrial hydrology and atmospheric sciences [1-2]. Moreover, it is very effective for the growth and development of plants. The optimum temperature value is needed for each plant to grow and plants are affected by extreme temperatures [3]. Thus, heat-based region maps studies were started for the first time in the USA, by using data from 4745 meteorological measurement centers from 1974-1995 years [4].

AT is specifically computed in thermometer shields that placed nearly 2 meter above floor at the meteorological stations by high precision. Nevertheless, the deploying of meteorological stations is generally sparse and unbalanced particularly in sparsely populated zones [5]. Therefore, meteorological satellites can provide uninterrupted data for atmospheric and superficial observations over wide coverage terrains [6].

<sup>1</sup> Faculty of Engineering and Architecture, Department of Electrical and Electronics Engineering, Batman University, Batman, TR, (acar\_emrullah@hotmail.com)  <https://orcid.org/0000-0002-1897-9830>

Looking at the studies in the literature, many studies have been conducted in AT estimation using meteorological data. These studies can be classified into two main clusters: simple and advanced statistical approaches [6-10] and machine learning approaches [11-13]. In this study, since simple and multi-linear regression models are used in AT estimation, the study can be evaluated within the scope of first group studies.

Considering the first group studies, Bechtel et al. has proposed an approach by implementing empirical models, which used time series of land surface temperature in limited time intervals to estimate air temperature [14]. Yang et al. has built multiple linear regression models for estimating maximum minimum and mean air temperatures in Northeast China with the aid of MODIS land surface temperature data [15]. Good suggested a dynamic multiple linear regression model in order to estimate daily minimum and maximum land air temperatures using in situ observations at meteorological stations [16]. Zhang et al. proposed the multiple linear regressions and the partial least squares regression to estimate daily air temperatures over the Tibetan Plateau thanks to MODIS LST data [17]. Xu et al. implemented two statistical methods for Estimating daily maximum air temperature in British Columbia via MODIS data [18].

In this study, the following steps have been carried out respectively. First, 67 meteorological parameters have been obtained from Batman Provincial Directorate of Meteorology, affecting the average monthly temperature between 2012 and 2017. Second, using the meteorological data from 2012-2016, the average air temperature value of Batman province for 2017 has been estimated. The estimation process has been executed in two main stages. In the first stage, the most important parameters such as monthly average soil temperature, pressure, water vapor pressure, wind speed and relative humidity have been employed as input in the simple linear regression model separately and the monthly average temperature of 2017 for Batman city has been estimated. In the second stage, all 67 parameters have been utilized as inputs of multi linear regression model and monthly average temperature has been estimated by this way.

## 2. Materials and Methods

### 2.1. Data Collection

The data employed in this study was obtained from Batman Provincial Directorate of Meteorology. The obtained data includes 67 meteorological data for each month among 2012-2017 years. The list of 67 employed meteorological parameters for each month is given in below.

1. Monthly Average Air Pressure (hPa)
2. Monthly Maximum Air Pressure (hPa)
3. Monthly Minimum Air Pressure (hPa)
4. Monthly Average Temperature (° C)
5. Average Number of Days with Daily Average Temperature 5 ° C and above
6. Monthly Average of Maximum Daily Temperatures (° C)
7. Monthly Average of Daily Minimum Temperatures (° C)
8. Monthly Maximum Temperature (° C)
9. Monthly Average above Ground Minimum Temperature (° C)
10. Monthly Maximum above Ground Minimum Temperature (° C)
11. Monthly Minimum above Ground Minimum Temperature (° C)

12. Monthly Average Water Vapor Pressure (hPa)
13. Monthly Average Relative Humidity (%)
14. Monthly Average Wind Speed (m ÷ sec)
15. Monthly Strong Windy Days Avg.
16. Sum of Blowing Numbers of Wind in Monthly N Direction
17. Monthly N Direction of Blowing (%)
18. Average Wind Speed in Monthly N Direction (m ÷ sec)
19. Sum of Blowing Numbers of Wind in Monthly NNE Direction
20. Monthly NNE's Blowing Rate (%)
21. Average Wind Speed in Monthly NNE Direction (m ÷ sec)
22. Sum of Blowing Numbers of Wind in Monthly NE Direction
23. Monthly NE Direction of Blowing (%)
24. Average Wind Speed in Monthly NE Direction (m ÷ sec)
25. Sum of Blowing Numbers of Wind in Monthly ENE Direction
26. Monthly ENE Direction of Blowing (%)
27. Average Wind Speed in Monthly ENE Direction (m ÷ sec)
28. Sum of Blowing Numbers of Wind in Monthly E Direction
29. Monthly E Direction of Blowing (%)
30. Average Wind Speed in Monthly E Direction (m ÷ sec)
31. Sum of Blowing Numbers of Wind in Monthly ESE Direction
32. Blowing Rate in Monthly ESE Direction (%)
33. Average Wind Speed in Monthly ESE Direction (m ÷ sec)
34. Sum of Blowing Numbers of Wind in Monthly SE Direction
35. Blowing Rate in Monthly SE Direction (%)
36. Average Wind Speed in Monthly SE Direction (m ÷ sec)
37. Sum of Blowing Numbers of Wind in Monthly SSE Direction
38. Blowing Rate in Monthly SSE Direction (%)
39. Average Wind Speed in Monthly SSE Direction (m ÷ sec)
40. Sum of Blowing Numbers of Wind in Monthly S Direction
41. Blowing Rate in Monthly S Direction (%)
42. Average Wind Speed in Monthly S Direction (m ÷ sec)
43. Sum of Blowing Numbers of Wind in Monthly SSW Direction
44. Blowing Rate in Monthly SSW Direction (%)
45. Average Wind Speed in Monthly SSW Direction (m ÷ sec)
46. Sum of Blowing Numbers of Wind in Monthly SW Direction
47. Blowing Rate in Monthly SW Direction (%)
48. Average Wind Speed in Monthly SW Direction (m ÷ sec)
49. Sum of Blowing Numbers of Wind in Monthly WSW Direction
50. Blowing Rate in Monthly WSW Direction (%)
51. Average Wind Speed in Monthly WSW Direction (m ÷ sec)
52. Sum of Blowing Numbers of Wind in Monthly W Direction
53. Monthly W Direction of Blowing (%)
54. Average Wind Speed in Monthly W Direction (m ÷ sec)
55. Sum of Blowing Numbers of Wind in Monthly WNW Direction

56. Monthly WNW Direction of Blowing (%)
57. Average Wind Speed in Monthly WNW Direction (m ÷ sec)
58. Sum of Blowing Numbers of Wind in Monthly NW Direction
59. Monthly NW Direction of Blowing (%)
60. Average Wind Speed in Monthly NW Direction (m ÷ sec)
61. Sum of Blowing Numbers of Wind in Monthly NNW Direction
62. Monthly NNW Direction of Blowing (%)
63. Average Wind Speed in Monthly NNW Direction (m ÷ sec)
64. Average Number of Days with Monthly Daily Average Wind 2.5 (m ÷ s) and above Avg.
65. Monthly Average 5 cm Soil Temperature (° C)
66. Maximum 5 cm per month Soil Temperature (° C)
67. Minimum 5 cm per month Soil Temperature (° C)

## 2.2. Regression Analysis

Regression can be defined as the dependent variable or variables being a function of the independent variable. Thanks to the regression analysis, information about the existence of the relationship between the variables can be reached. By expressing the relationship between the variables with equations, we can estimate the variables we do not know with the help of the variables we know. Regression can be examined as simple and multiple regression analysis according to the number of independent variables. In addition, according to the function used, it can be divided into two groups as linear and non-linear regression analysis [19]. In this study, simple linear and multi-linear regression models are used.

### 2.2.1 Simple Linear Regression Analysis

In simple linear regression (SLR) analysis, the regression model in which the relationship between one independent variable 'x' and the dependent variable 'y' is expressed by a linear function as shown in Equation 1.

$$y = b_0 + b_1x + \epsilon \quad (1)$$

Here,  $b_0$  and  $b_1$  are unknown constants and represent intercept and slope, respectively. They are also called regression coefficients.  $\epsilon$  represents the error term. SLR is a method of estimating a standard distributed numerical variable whose value is unknown from a standard distributed numerical variable and which has a relationship between them [18-19].

### 2.2.2 Multiple Linear Regression Analysis

In multiple linear regressions, the relationship between multiple independent variables ( $x_1, x_2, \dots, x_n$ ) and one dependent variable, "y", is obtained. The relationship between each independent variable and the dependent variable can be determined with the aid of below equation.

$$y = a + b_1x_1 + b_2x_2 + b_3x_3 + \dots + b_nx_n \quad (2)$$

In this way, instead of the linear 'b' coefficient, 'n' linear regression coefficient will be transferred. The method used to achieve its function in both simple and multiple regression method is the least squares method. Moreover, MLR can also be fitted in other methods, such as least absolute deviations regression, or the least squares cost function as in lasso (L1-norm penalty) and in ridge regression (L2-norm penalty). [20-22].

### 2.3. Block diagram of the proposed Approach

The architecture of the proposed system is shown in Fig. 1. In the first phase, some important parameters were utilized in the simple linear regression model as input separately. In the second phase, all parameters were in multiple linear regression models as inputs and monthly average temperature was estimated by this way.

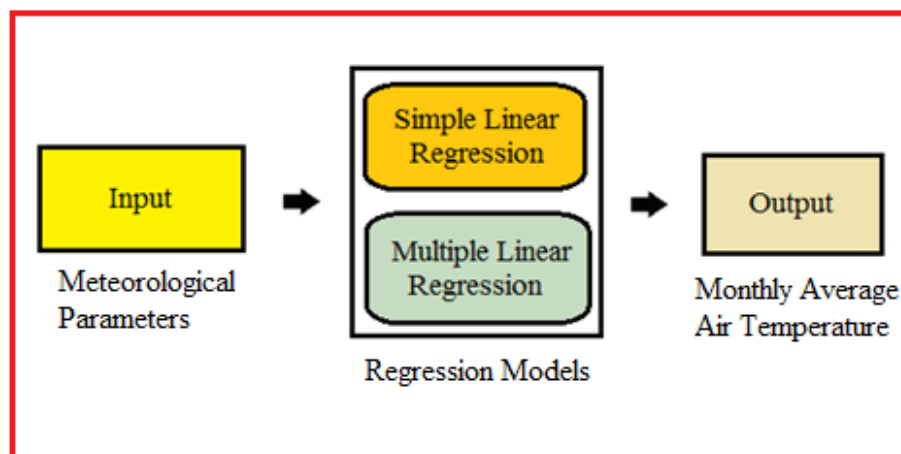


Figure 1. Architecture of the Proposed System.

For performance criteria, different error metrics [Root Mean Square Error (RMSE) and coefficient of determination ( $R^2$ )] were calculated for each scenario. The equation of those error metrics are given in Eq. (2) and Eq. (3).

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

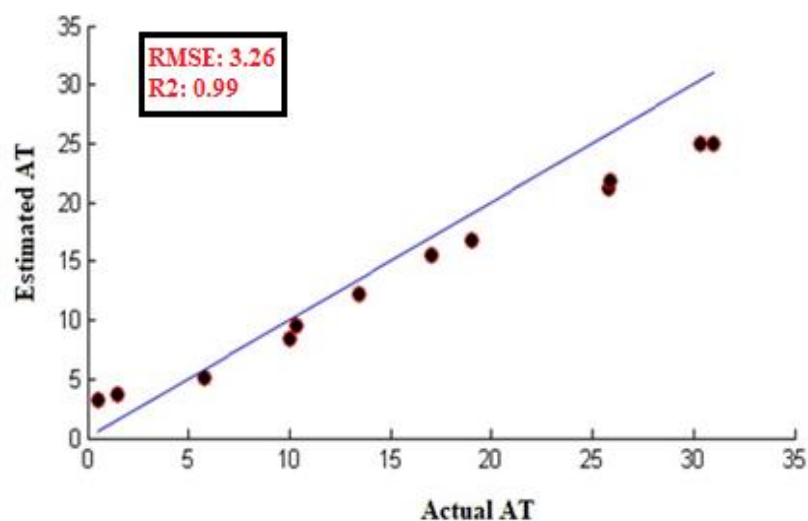
$$R^2 = \left( \frac{\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{m \sum x_i^2 - (\sum x_i)^2} \sqrt{m \sum y_i^2 - (\sum y_i)^2}} \right) \quad (3)$$

Here,  $y$ ,  $y_i$  and  $m$  show the actual value, the estimated value and the number of observations, respectively.

### 3. Results

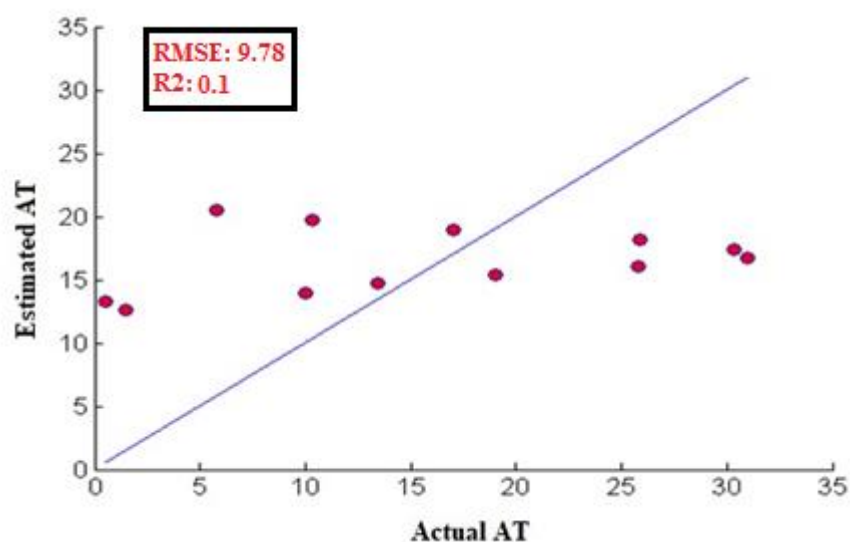
In this section, the parameters affecting the air temperature between 2012 and 2016 taken from Batman Meteorology Provincial Directorate were used individually and collectively as a training set of regression models. After that, the average monthly air temperature (AT) of Batman city in 2017 was estimated with the help of these data sets.

As a result, by using these parameters separately or collectively, all desired graphs,  $R^2$  and RMSE values were obtained in below figures, respectively.



**Figure 2. Effect of monthly average soil temperature on air temperature (AT) prediction**

As seen in the Fig. 2, an almost perfect regression curve ( $R^2=0.993$ ) was obtained with a very low error (RMSE=3.26) in the estimation of the monthly average air temperature using the monthly average soil temperature parameter, which can also be considered as the parameter with the highest impact among all parameters for the AT estimation.



**Figure 3. Effect of monthly average air pressure on air temperature (AT) prediction**

In the Fig. 3, a bad regression curve ( $R^2=0.1$ ) was obtained with a high error rate (RMSE=9.78) for estimation of the monthly average air temperature using the monthly average air pressure parameter.

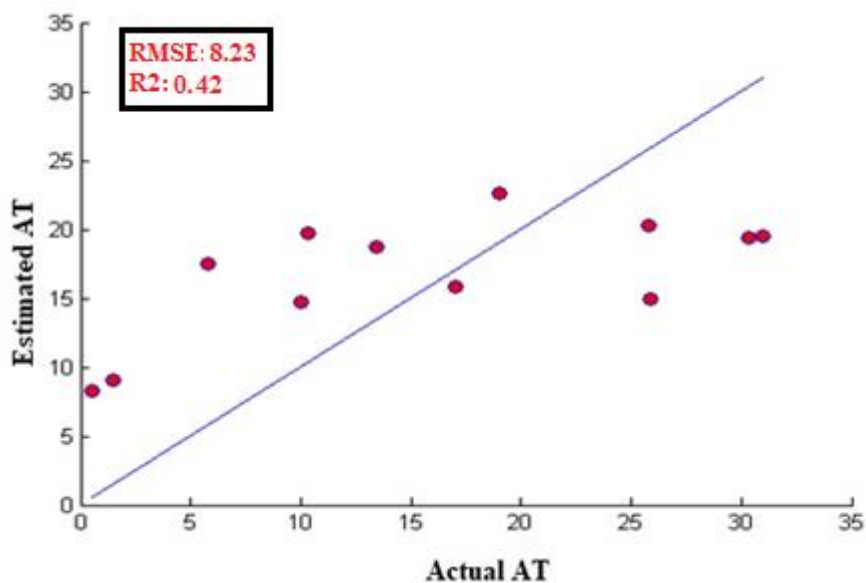


Figure 4. Effect of monthly average water vapour pressure on air temperature (AT) prediction

In the Fig. 4, a coefficient of correlation ( $R^2=0.42$ ) was acquired with an error rate (RMSE=8.23) for estimation of the monthly average air temperature using the monthly average water vapour pressure parameter.

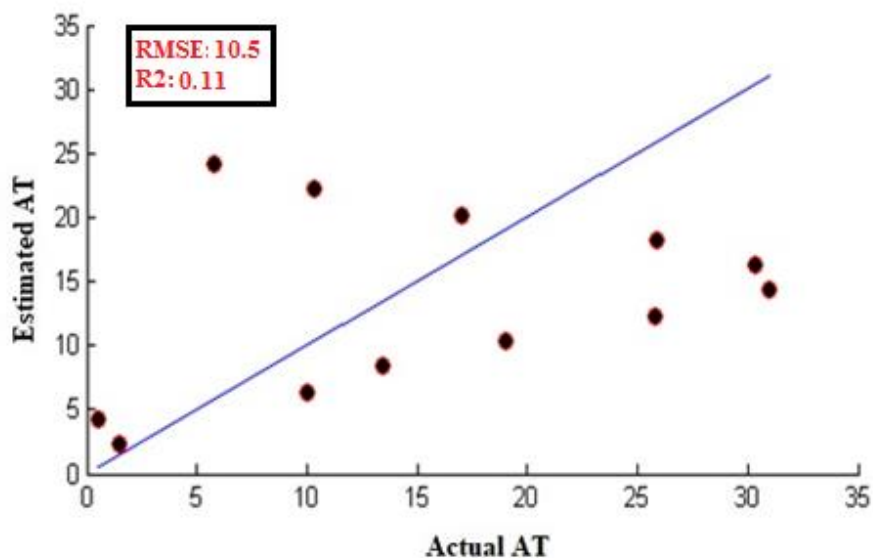


Figure 5. Effect of monthly average wind speed on air temperature (AT) prediction

As indicated in the Fig. 5, a coefficient of correlation ( $R^2=0.11$ ) was computed with an error rate (RMSE=10.53) to estimate the monthly average air temperature using the monthly average wind speed parameter, which might also be evaluated as the parameter with the lowest impact among all parameters for the AT estimation.

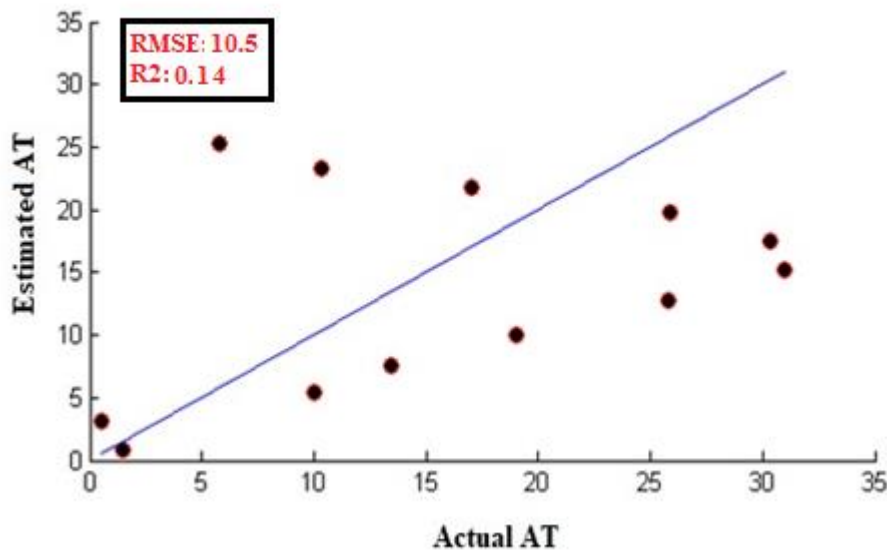


Figure 6. Effect of monthly average relative humidity on air temperature (AT) prediction

In the Fig. 6, a low coefficient of correlation ( $R^2=0.14$ ) was obtained with an bad error rate (RMSE=10.55) for estimating the monthly average air temperature using the monthly average relative humidity parameter.

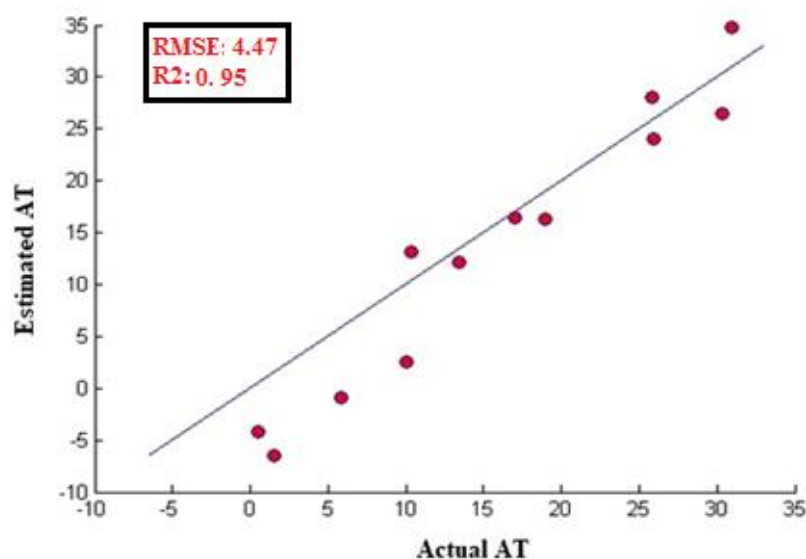


Figure 7. Effect of all parameters on air temperature (AT) prediction

Finally, in the Fig. 7, a very high coefficient of correlation ( $R^2=0.95$ ) was observed with a very low error rate (RMSE=4.47) in order to estimate the monthly average air temperature using the all 67 meteorological parameters. It can be said that when all features were combined, a very good correlation might be observed for AT estimation.



#### 4. Conclusion

In this study, monthly average temperature of Batman province was estimated by simple and multi-linear regression methods using some important meteorological parameters.

In the estimation phase, all necessary data between 2012 and 2017 were collected and the estimation process was carried out in two stages. In the first stage, some of the important data (monthly average soil temperature, pressure, water vapour pressure, wind speed and relative humidity) were used separately as input parameters in simple regression analysis. In the second stage, all meteorological data was used as the input of the multiple liner regression model. Later, with linear regression analysis, estimation results were obtained as numerical values and graphs, and the error rates of the analyzes and the closeness of the estimation values to actual values were examined. When the numerical values of the graphics are examined carefully, it is understood that my study gives good result.

In the future, it is aimed to collect more data and apply different machine learning techniques to achieve even better results.

#### References

- [1] S. Solomon, Intergovernmental Panel on Climate Change. Climate Change 2007: 8e Physical Science Basis: Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, NY, USA, 2007.
- [2] S. Stisen, I. Sandholt, A. Nørgaard, R. Fensholt, and L. Eklundh, "Estimation of diurnal air temperature using MSG SEVIRI data in west Africa," *Remote Sensing of Environment*, vol. 110, no. 2, pp. 262–274, 2007.
- [3] Akyüz, A, Kumaş, K, Ayan, M, Güngör, A. (2020). Antalya İli Meteorolojik Verileri Yardımıyla Hava Sıcaklığının Yapay Sinir Ağları Metodu ile Tahmini. *Gümüşhane Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 10 (1), 146-154. DOI: 10.17714/gumusfenbil.511481.
- [4] Sri Sankari G. and Valarmathi, Dr.A., 2017. Weather Forecasting with Back Propagation of Neural Network using MATLAB. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, 2(2), 2456-3307.
- [5] C. J. Willmott and S. M. Robeson, "Climatologically aided interpolation (CAI) of terrestrial air temperature," *International Journal of Climatology*, vol. 15, no. 2, pp. 221–229, 1995.
- [6] Liu, H., Zhou, Q., Zhang, S., & Deng, X. (2019). Estimation of Summer Air Temperature over China Using Himawari-8 AHI and Numerical Weather Prediction Data. *Advances in Meteorology*, 2019.
- [7] H. Nieto, I. Sandholt, I. Aguado, E. Chuvieco, and S. Stisen, "Air temperature estimation with MSG-SEVIRI data: calibration and validation of the TVX algorithm for the Iberian Peninsula," *Remote Sensing of Environment*, vol. 115, no. 1, pp. 107–116, 2011.
- [8] A. Benali, A. C. Carvalho, J. P. Nunes, N. Carvalhais, and A. Santos, "Estimating air surface temperature in Portugal using MODIS LST data," *Remote Sensing of Environment*, vol. 124, no. 9, pp. 108–121, 2012.
- [9] P. Noi, J. Degener, and M. Kappas, "Comparison of multiple linear regression, cubist regression, and random forest algorithms to estimate daily air surface temperature from dynamic combinations of MODIS LST data," *Remote Sensing*, vol. 9, no. 5, p. 398, 2017.
- [10] Catal, C , Ece, K , Arslan, B , Akbulut, A . "Benchmarking of Regression Algorithms and Time Series Analysis Techniques for Sales Forecasting". *Balkan Journal of Electrical and Computer Engineering* 7 (2019 ): 20-26.
- [11] Akyüz, A. Ö., Kumaş, K., Ayan, M., & Güngör, A. Antalya İli Meteorolojik Verileri Yardımıyla Hava Sıcaklığının Yapay Sinir Ağları Metodu ile Tahmini. *Gümüşhane Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 10(1), 146-154.
- [12] Bilgili, M., & Sahin, B. (2009). Prediction of long-term monthly temperature and rainfall in Turkey. *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, 32(1), 60-71.

- [13] Akinci, T , Noğay, H . (2019). Application Of Decision Tree Methods For Wind Speed Estimation. *European Journal of Technique (EJT)* , 9 (1) , 74-83 . DOI: 10.36222/ejt.558914.
- [14] Bechtel, B., Zakšek, K., Oßenbrügge, J., Kaveckis, G., & Böhner, J. (2017). Towards a satellite based monitoring of urban air temperatures. *Sustainable cities and society*, 34, 22-31.
- [15] Yang, Y. Z., Cai, W. H., & Yang, J. (2017). Evaluation of MODIS land surface temperature data to estimate near-surface air temperature in Northeast China. *Remote Sensing*, 9(5), 410.
- [16] Good, E. (2015). Daily minimum and maximum surface air temperatures from geostationary satellite data. *Journal of Geophysical Research: Atmospheres*, 120(6), 2306-2324.
- [17] Zhang, H., Zhang, F., Ye, M., Che, T., & Zhang, G. (2016). Estimating daily air temperatures over the Tibetan Plateau by dynamically integrating MODIS LST data. *Journal of Geophysical Research: Atmospheres*, 121(19), 11-425.
- [18] Xu, Y., Knudby, A., & Ho, H. C. (2014). Estimating daily maximum air temperature from MODIS in British Columbia, Canada. *International Journal of Remote Sensing*, 35(24), 8108-8121.
- [19] Tabachnick, B. G., Fidell, L. S., & Ullman, J. B. (2007). *Using multivariate statistics* (Vol. 5). Boston, MA: Pearson.
- [20] Sykes, A. O. (1993). An introduction to regression analysis.
- [21] Uyanık, G. K., & Güler, N. (2013). A study on multiple linear regression analysis. *Procedia-Social and Behavioral Sciences*, 106(1), 234-240.
- [22] Yan, X., & Su, X. (2009). *Linear regression analysis: theory and computing*. World Scientific.
- [23] Abraham, A., Cherukuri, A. K., Melin, P., & Gandhi, N. (Eds.). (2019). *Intelligent Systems Design and Applications: 18th International Conference on Intelligent Systems Design and Applications (ISDA 2018) Held in Vellore, India, December 6-8, 2018, Volume 1 (Vol. 940)*. Springer.
- [24] Allison, P. D. (1999). *Multiple regression: A primer*. Pine Forge Press.