Comparison of Linear Regression and Simple Linear Regression for critical temperature of semiconductor

Dr.R.Venkatesh Babu, Dr.G.Ayyappan, Dr.A. Kumaravel

Dean Academic, BIHER, BIST, Bharath University, Chennai

Associate Professor, Department of Information Technology, BIHER, BIST, Bharath University, Chennai.

Professor and Head, Department of Information Technology, BIHER, BIST, Bharath University, Chennai.

Abstract

The regression analysis plays a vital role in forecasting, estimating and predicting the material science domain. In this research work measures a statistical model to estimate the critical temperature of superconductor. This critical temperature formulated by using the superconductor's chemical formula. The statistical model has given several measurements like Correlation Co efficient, Mean Absolute Error (MAE), Root Mean Squared Error (RMSR), Relative Absolute Error (RAE), and Root Relative Squared Error (RRSE). These measurements extracted based on atomic mass (AM), atomic radius (AR), valence(V), thermal conductivity (TC), and electron affinity (EA) contribute the most to the model's predictive accuracy. This research work focuses the comparisons of various measurements namely Correlation Co efficient, Mean absolute error, Root Mean squared error, Relative absolute error, Root relative squared error and also time taken to build the model of leading regression algorithms like Linear and Simple Linear regression models for superconductor.

Keywords: Conductivity, Simple Linear Regression, Atomic Radius, Linear Regression, valence, Electron affinity, Correlation Coefficient and Atomic Mass.

I INTRODUCTION

In this section presents the introduction of the Superconducting materials. The materials that conduct current with zero resistance - have an important practical use. The well-known usage is in the Magnetic Resonance Imaging systems (MRI) has commonly used by health care sectors for clear and detail view of an internal body imaging. Another Important usage is the superconducting coils. It applied to maintain high magnetic fields in the Large Hadron Collider at CERN, Furthermore, superconductors could revolutionize the energy industry as frictionless (zero resistance) superconducting wires and electrical system may transport and deliver electricity with no energy loss[2].

In this study has taken an entirely data-driven approach to create a statistical model that predicts Tc used on its chemical formula. The superconductor data comes from the Superconducting Material Database maintained by Japan's National Institute for Materials Science (NIMS) at http://supercon.nims.go.jp/index_en.html. After some data preprocessing, 21,263 superconductors are used.[1]

The paper is structured as follows. First, in Sect. 2, the background and related works are discussed. Second, in Sect. 3, the materials and methods are used in this research work. Second, in Sect. 4, the results and discussions are presented. Finally, in Sec.5, Conclusions of this research work.

II LITERATURE SURVEY

In this section presents the background work of this research work. To our knowledge, Valentin et al. [3] and our work are the only papers that focus on statistical models to predict Tc for a broad class of materials. The few researches like Owolabi et al. [4], Owolabi and Olatunji [5] proposes on estimating Tc for Fe and MgB2 based superconductors respectively.

The proposed system features (or predictors) based on the superconductor's primary attributes that could be helpful in predicting Tc.[1] For example, consider Nb Pd 0.8 0.2 with Tc = 1.98 K. We can derive a feature based on the average thermal conductivities of the elements. Niobium and palladium's thermal conductivity coefficients are 54 and 71 W/(m×K) respectively. The mean thermal conductivity is (54 71)/2 62. + = 5 W/ (m×K). We can treat the mean thermal conductivity variable as a feature to predict Tc. In total, we define and extract 81 features from each superconductor.[1]

In the without presence of any theory-based prediction models, simple empirical rules based on experimental results have guided researchers in synthesizing superconducting materials for many years. For example, Tc is related to the number of available valence electrons per atom.[6] It is now well known that many of the simple empirical rules are violated;[7] We tried several statistical models but we eventually settled on two: A multiple regression model which serves as a benchmark and a gradient boosted models as the vital model for the predication which is used in our software. In choosing the attributes, and also use the conclusion to select few attributes. This proposed work dropped the boiling point variable, and as a replacement for the utilization of the combination heat variable which has no missing values, and is highly correlated with the boiling point variable. This work proposed some experience and insight creating some initial models for predicting Tc of elements only. [4,8,9&10].

III MATERIALS AND METHODS

In this section presents the materials and methods of this research work. The data set borrowed from https://archive.ics.uci.edu/ml/datasets/Superconductivty+Data. In this research work uses the one of the famous machine learning tool wake3.8.3. The dataset applies the condition based on the test mode 10 old cross validations.

Linear Regression: Class for using linear regression for prediction. And also applies the parameters like Attribute selection method is M5 method, batch size is 100, etc.

SimpleLinearRegression: Learns the simple linear regression model. This measurement considers based on the batch size and debug, etc.,

Dataset Description: There are two files:(1) train.csv contains 81 features extracted from 21263 superconductors along with the critical temperature in the 82nd column,(2) unique_m.csv contains the chemical formula broken up for all the 21263 superconductors from the train.csv file. The last two columns have the critical temperature and chemical formula. The original data comes from https://archive.ics.uci.edu/ml/datasets/Superconductivty+Data which is public. The goal here is to predict the critical temperature based on the features extracted.

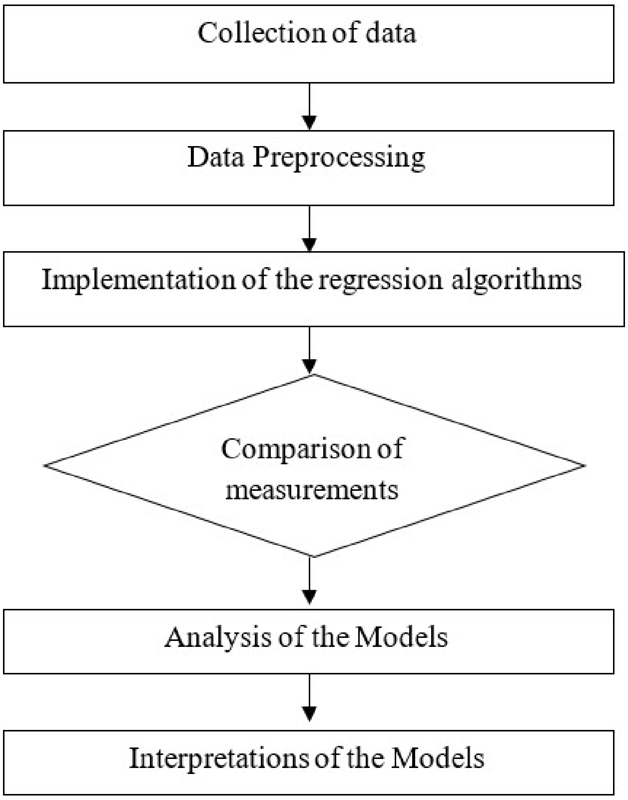

Figure 1: Proposed system

The borrowed dataset contains 82 attributes and 21263 instances. The temperature predication or estimation computed based on the 82th attribute namely critical_temp. There are 8 attributes shown in Table 1.

Table 1: List of Attributes

S.No	Name of the attribute	Measurement	Description of the attribute
1	Atomic Mass	Atomic mass units (AMU)	Total proton and neutron rest masses
2	First Ionization Energy	Kilo-Joules per mole (kJ/mol)	Energy required to remove a valence electron
3	Atomic Radius	Picometer (pm)	Calculated atomic radius
4	Density	Kilograms per meters cubed (kg/m3)	Density at standard temperature and pressure
5	Electron Affinity	Kilo-Joules per mole (kJ/mol)	Energy required to add an electron to a neutral atom
6	Fusion Heat	Kilo-Joules per mole (kJ/mol)	Energy to change from solid to liquid without temperature change
7	Thermal Conductivity	Watts per meter-Kelvin (W/(m K))	Thermal conductivity coefficient κ
8	Valence	No units	Typical number of chemical bonds formed by the element

IV RESULTS AND DISCUSSIONS

In this section presents the results and interpretations of this research work. In this research work implements the there are two leading regression algorithms one is Linear regression and another one is simple linear regression algorithm. The below table clearly demonstrates the several measurements indicates the which one is best model.

Table 2: Measurements of Linear Regression Vs Simple Linear Regression

S.No	Name of the Model	Time Taken to build the Model (In seconds)	Correlation Co efficient	Mean absolute error	Root Mean squared error	Relative absolute error	Root relative squared error	Total Number of Instances
1	Linear Regression	6.55	0.86	13.42	17.68	45.75%	51.63%	21263
2	Simple Linear Regression	0.22	0.72	18.24	23.72	62.19%	69.26%	21263

The above table represents the comparison of the time to build the respective model. Linear regression model has taken 6.55 seconds and Simple Linear Regression has 0.22 seconds.

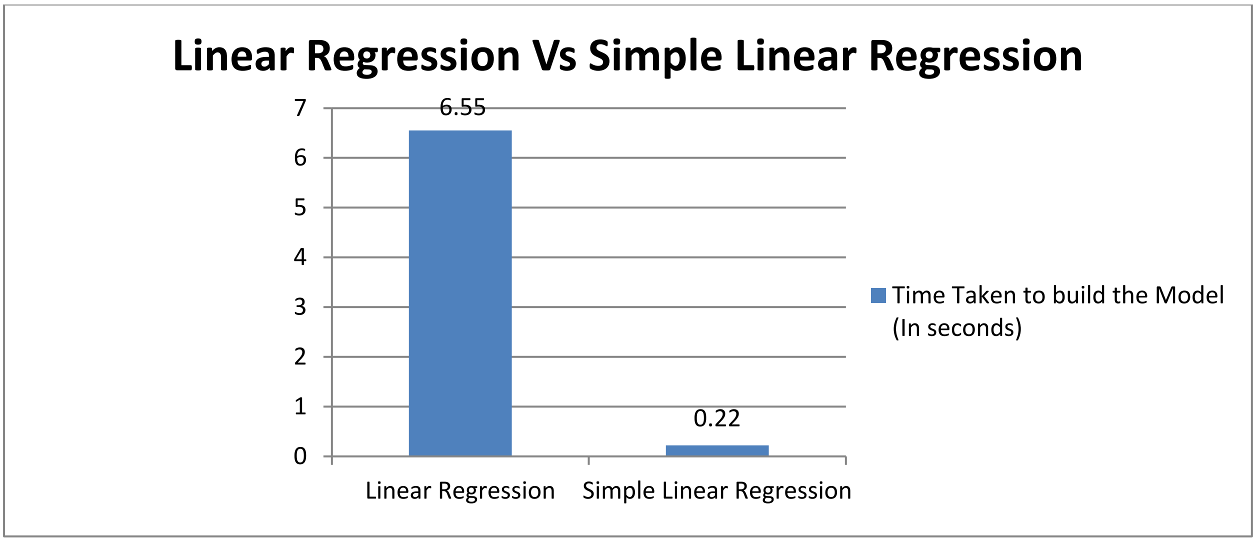

Figure 2: Graphical representation of Time taken to build the models

The above table represents the comparison of the time to build the respective model. Linear regression model has taken more time consumption compare with Simple Linear Regression.

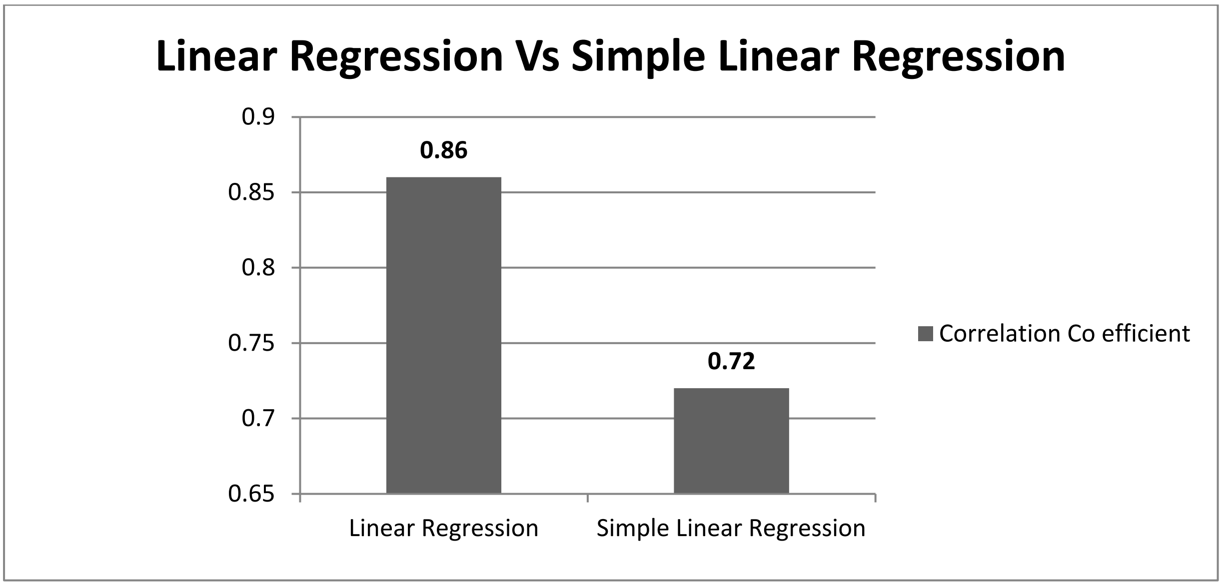

Figure 3: Graphical representation of Correlation Co efficient

The above diagram represents the comparison of the Correlation Coefficient. Linear regression model has 0.86 and Simple Linear Regression has 0.72.

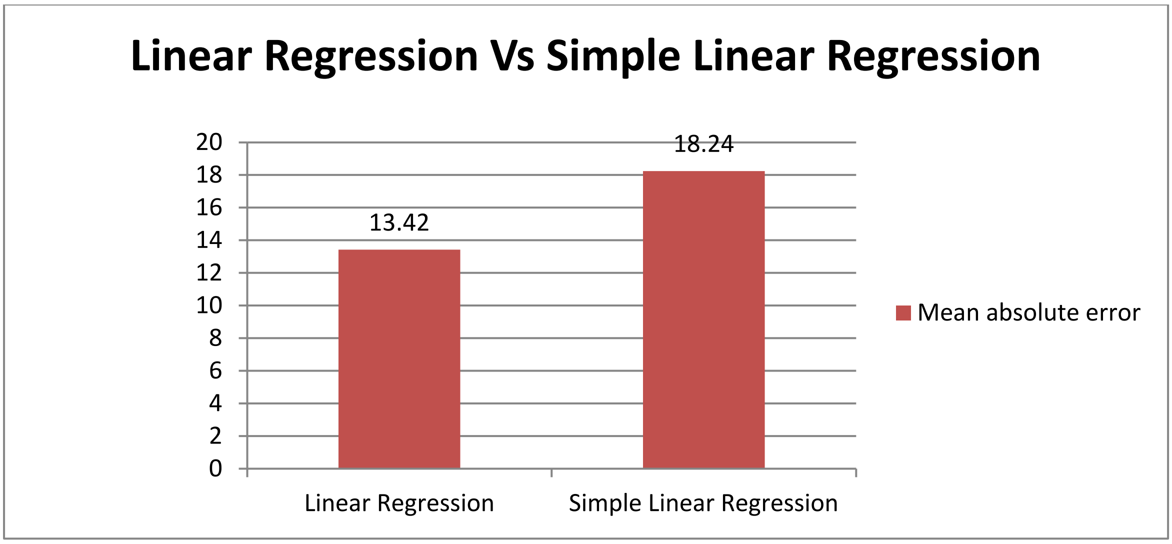

Figure 4: Graphical representation of Mean Absolute Error

The above diagram represents the comparison of the Correlation Coefficient. Linear regression model has 13.42 and Simple Linear Regression has 18.24.

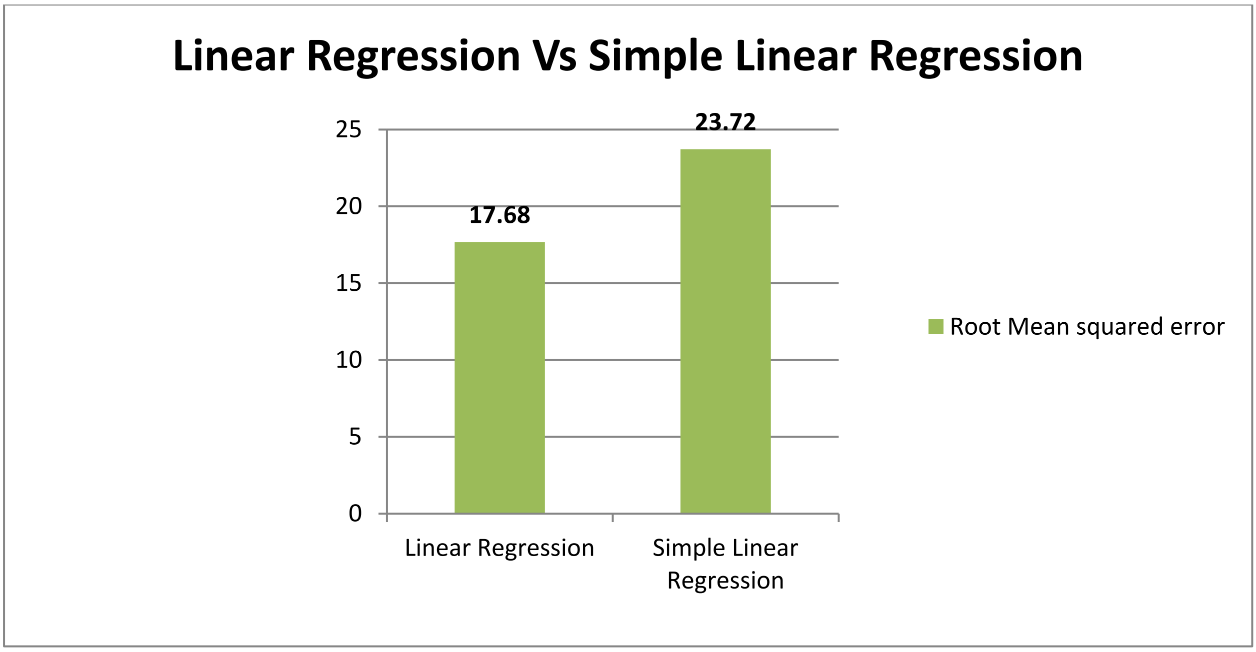

Figure 5: Graphical representation of Root Mean Squared Error

The above diagram represents the comparison of the Correlation Coefficient. Linear regression model has 17.68 and Simple Linear Regression has 23.72.

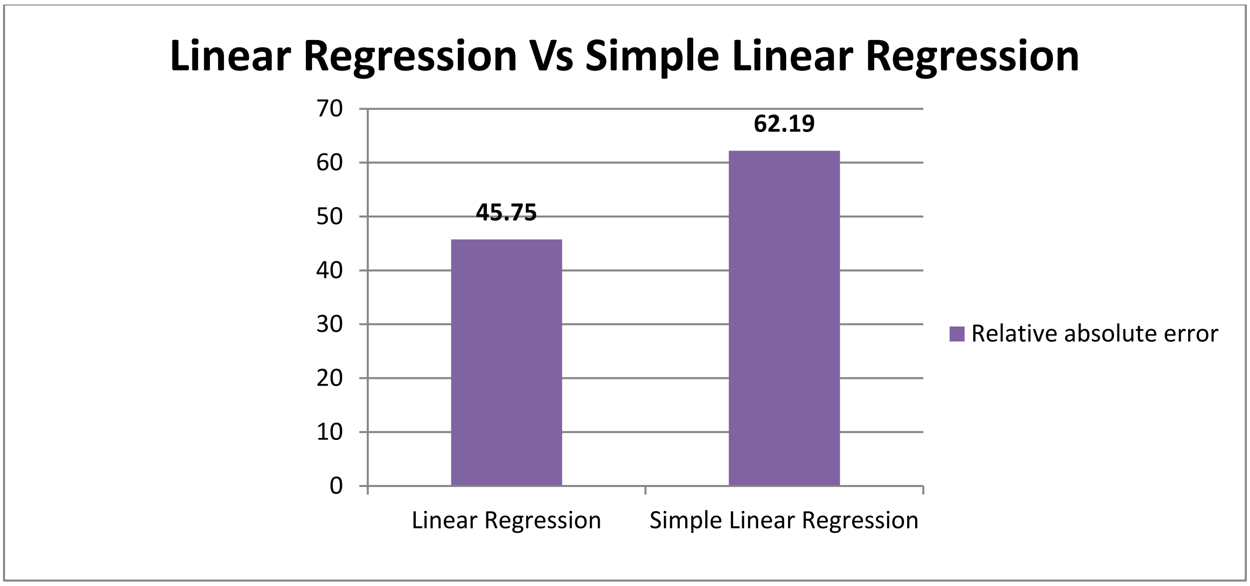

Figure 6: Graphical representation of Relative Absolute Error

The above diagram represents the comparison of the Relative Absolute Error. Linear regression model has 45.75% and Simple Linear Regression has 62.19%.

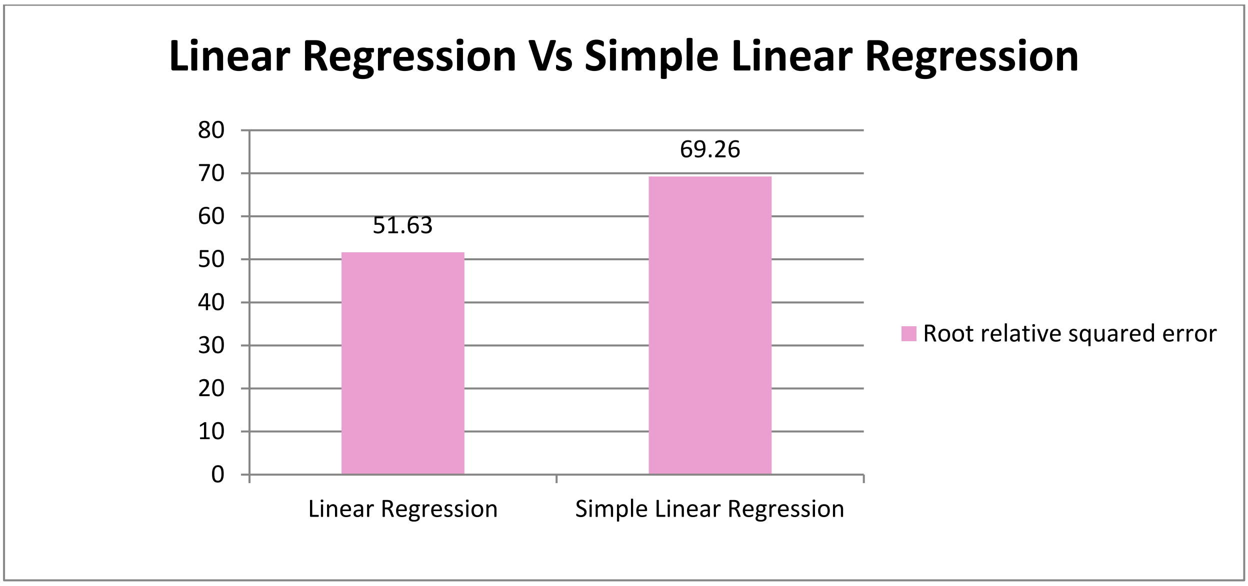

Figure 7: Graphical representation of Root Relative Squared Error

The above diagram represents the comparison of the Correlation Coefficient. Linear regression model has 51.63% and Simple Linear Regression has 69.26%.

V CONCLUSION

This research work concludes that the statistical model using only the superconductors' chemical formula and Linear regression model has taken 6.55 seconds and Simple Linear Regression has 0.22 seconds. Linear regression model has 0.86 and Simple Linear Regression has 0.72. Linear regression model has 13.42 and Simple Linear Regression has 18.24. Linear regression model has 17.68 and Simple Linear Regression has 23.72. Linear regression model has 51.63% and Simple Linear Regression has 69.26%. So, this research work recommends that Linear regression model has produced high efficiency.

References

[1] Hamidieh, Kam, A data-driven statistical model for predicting the critical temperature of a superconductor, Computational Materials Science, Volume 154, November 2018, Pages 346-354,

[2] T. Chen, T. He, M. Benesty, V. Khotilovich, Y. Tang, xgboost: Extreme Gradient Boosting. R package version 0.6.4.1, 2018.

[3] Curtarolo, I. Takeuchi, Machine learning modeling of superconducting critical temperature, 2017. <https://arxiv.org/abs/1709.02727>.

[4] T. Owolabi, A. Akande, S. Olatunji, Prediction of superconducting transition temperatures for fe-based superconductors using support vector machine. 35 (2014) 12–26.

[5] T.O. Owolabi, K. Akande, S. Olatunji, Estimation of superconducting transition temperature tc for superconductors of the doped MgB2 system from the crystal lattice parameters using support vector regression, J. Supercond. Novel Magn. 28 2015) 75–81.

[6] B.T. Matthias, Empirical relation between superconductivity and the number of electrons per atom, Phys. Rev. 97 (1955) 74–76.

[7] K. Conder, A second life of the matthias's rules, Supercond. Sci. Technol. 29 (8) (2016).

[8] T. Chen, C. Guestrin, Xgboost: A scalable tree boosting system, 2016, <https:// arxiv.org/abs/1603.02754>.

[9] T. Chen, T. He, M. Benesty, V. Khotilovich, Y. Tang, xgboost: Extreme Gradient Boosting. R package version 0.6.4.1, 2018.

[10] S. Nishiyama, H. Fujita, M. Hoshi, X. Miao, T. Terao, X. Yang, T. Miyazaki, H. Goto, T. Kagayama, K. Shimizu, H. Yamaoka, H. Ishii, Y.-F. Liao, Y. Kubozono, Preparation and characterization of a new graphite superconductor: Ca Sr C 0.5 0.5 6, Sci. Rep. 7 (7436) (2017).

[11] G.Ayyappan., K.Sivakumar, SMO classification for cervical cancer dataset by applying various kernels, e-ISSN:0976-5166, Vol.10,No.1, Feb-Mar 2019.

[12] G.Ayyappan, K.Sivakumar,A Novel Approach of Ensemble models by using EDM, -ISSN:0976-5166, Vol.8,No.6, Dec-Jan 2018.