ASEC 2025 Conference

The 6th International Electronic Conference on Applied Sciences

09–11 December 2025 | Online

MDPI

Predictive Modelling of Ionospheric Total Electron Content over the Philippines using Machine Learning Methods

Vincent Louie Maglambayan[1], Jazzie Jao[1,2], Edgar Vallar[1]

1 Department of Physics, De La Salle University Manila, Manila 1004, Philippines

2 Department of Software Technology, De La Salle University Manila, Manila 1004, Philippines

INTRODUCTION & AIM

Vertical total electron content (VTEC) is the measure of the electron density in the Earth's ionosphere. It is expressed in total electron content units (TECU) which is equivalent to 10^{16} electrons per square meter. GNSS observations are used to measure this amount. With the increasing prevalence of machine learning algorithms, in other fields, ionospheric researchers have started applying these algorithms to model and predict ionospheric trends in specific local sectors of the ionosphere. However, none have been done so in the Philippine region. Traditional empirical and physics-based models exist but these are insufficient in looking into possibly interesting local features in the area due to its proximity to the equator. This study aims to asses the performance of three machine learning techniques: support vector regression (SVR), random forest (RF) and gradient boosting (GB), in forecasting ionospheric total electron content using Philippine GNSS data with geomagnetic and solar data as predictors. The importance of these predictors will also be evaluated.

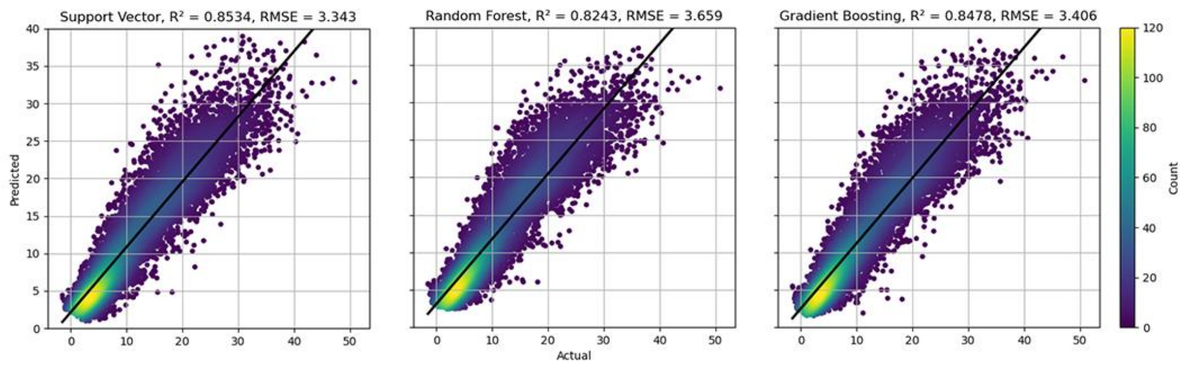
METHOD

The data was obtained from PIMO receiver station located in Manila, Philippines. The station contains satellite pseudorange data that can be converted into VTEC. Geomagnetic and solar activity data were taken from https://omniweb.gsfc.nasa.gov/form/dx1.html which is used as predictors for VTEC. The following values were used as predictors: sine and cosine of time (year, DOY, hour), scalar B, Bz, SW proton density, SW plasma speed, flow pressure, Kp-index, dst-index, f10.7-index and Lyman Alpha. Data observed from January 1, 2010 to December 31, 2019 were used as training data while the period from January 1, 2020 to December 31, 2020 were used as test data with a temporal resolution of one hour. The ionosphere was assumed as a thin spherical shell of height 450 km and an elevation mask of 25° was applied. The machine learning algorithms were coded using Python's sklearn module. The predictors were first normalized before it was fed into the algorithm for training. Hyperparameter tuning was done. The following shows the optimal parameters for each method. Any parameter not found in the table was set to their default values.

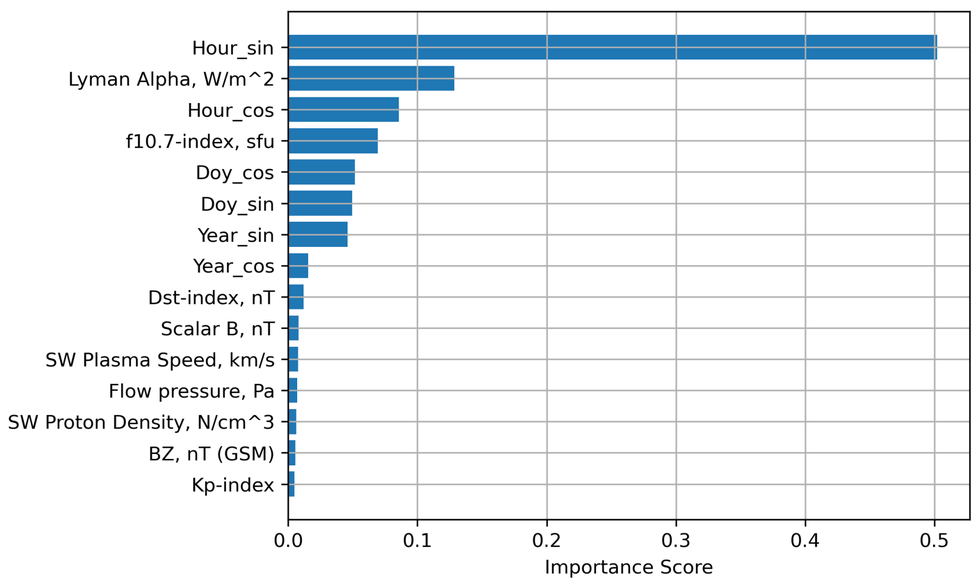
Model	Parameter	Value
SVR	*gamma*	auto
	C	2.3
	epsilon	0.001
RF	*n_estimators*	424
	min_samples_leaf	3
	max_features	sqrt
	random state	42
GB	*learning_rate*	0.028
	n_estimators	300
	subsample	0.8
	criterion	squared error
	min_samples_leaf	4
	max_depth	9
	random_state	42
	max_features	sqrt

RESULTS & DISCUSSION

The support vector machine algorithm performed the best out of three with a R^2 coefficient of 0.8534 and RMSE of 3.343 TECU. This is followed by gradient boosting with R^2 coefficient of 0.8478 and RMSE of 3.406 TECU. Random forest performed the least among the three with an R^2 of 0.8243 and RMSE of 3.659. The following table shows the plots of the predicted VTEC of each model against the actual VTEC measured by PIMO. The line of best was is also included. It should be noted that in terms of runtime, support vector regression performed the slowest and random forest performed the fastest



Using permutation-based importance values, the most important features across all three models. It is determined that the most important feature is the sine of the hour in Universal Time across all three models. This is followed by Lyman alpha solar index and cosine of the hour coming in third. It is can also be seen that time holds a lot of significance as six of then ten most important features are time-based. The non-time features that hold the most significance are the Lyman alpha solar index and the f10.7 solar flux index which is consistent in other literatures.



CONCLUSION

SVR gave the best performance among the three. However, it should be noted that all three models yielded similar R^2s and RMSEs and thus, all three models are interchangeable. However, SVR runtime was significantly slower than the other two models. The most important feature was evaluated to be the hours of the day. These are followed by the Lyman alpha solar index and the f10.7 solar flux index. The researchers will make use of these results in producing a two-dimensional