Blue Mars Initiative: Developing Machine Learning Models to Forecast Mesoscale Martian Weather Conditions

Jared Frazier and Dr. David Butler
Department of Computer Science, MTSU, Murfreesboro TN

OBJECTIVES AND MOTIVATION

The Blue Mars Initiative is a student research group at MTSU where each member independently researches a topic that may be relevant to the hypothetical colonization of Mars. My role is to….

1) Implement multiple linear regression (MLR), Multilayer Perceptron (MLP) Neural Network (NN), K-Nearest Neighbors (KNN), Random Forest (RF), and Support Vector Regression (SVR) for a single-step time series forecast of the mean ambient air temperature in a mesoscale region of Mars known as the Gale Crater.
2) Compare the performance of initial models using Root Mean Square Error (RMSE).

MATERIALS AND METHODS

The dataset was scraped from the National Aeronautics and Space Administration Planetary Data System. Machine learning models were selected based off previous studies for short term time-series weather forecasting on Earth [1], [2].

MATERIALS AND METHODS CONTINUED

Three days of temperature, pressure, and humidity data were used to predict the mean ambient air temperature in the Gale Crater a single time step ahead (one day in advance). Data was min-max normalized and ~4% was imputed using KNN.

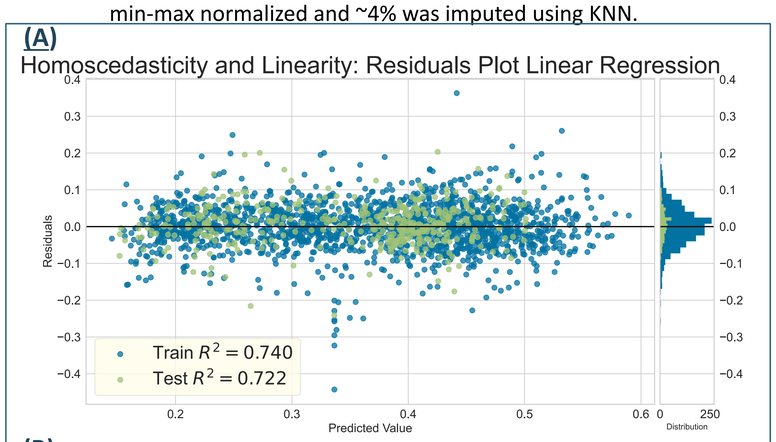

Fig. 1. Assumptions for Linear Regression. (A) Linearity and homoscedasticity, **(B)** no/little multicollinearity, and **(C)** Gaussian residuals (violation).

RESULTS

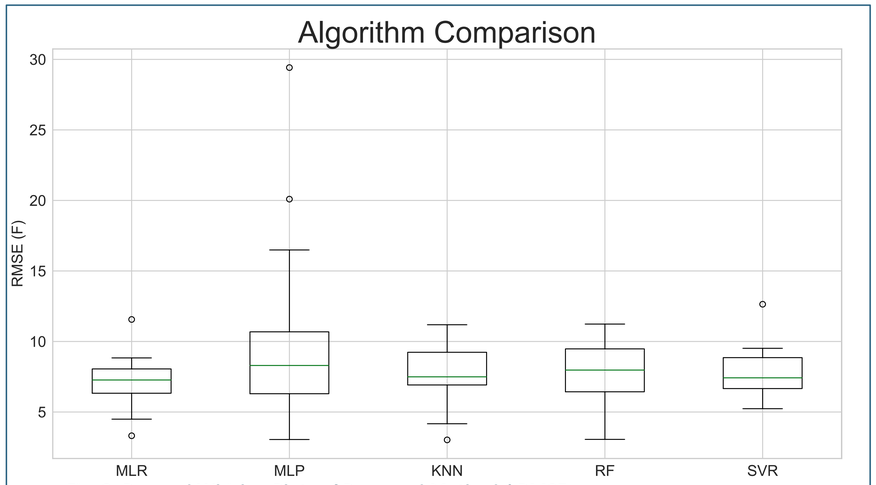

Fig. 2. Box and Whisker Plots of Proposed Methods' RMSE.

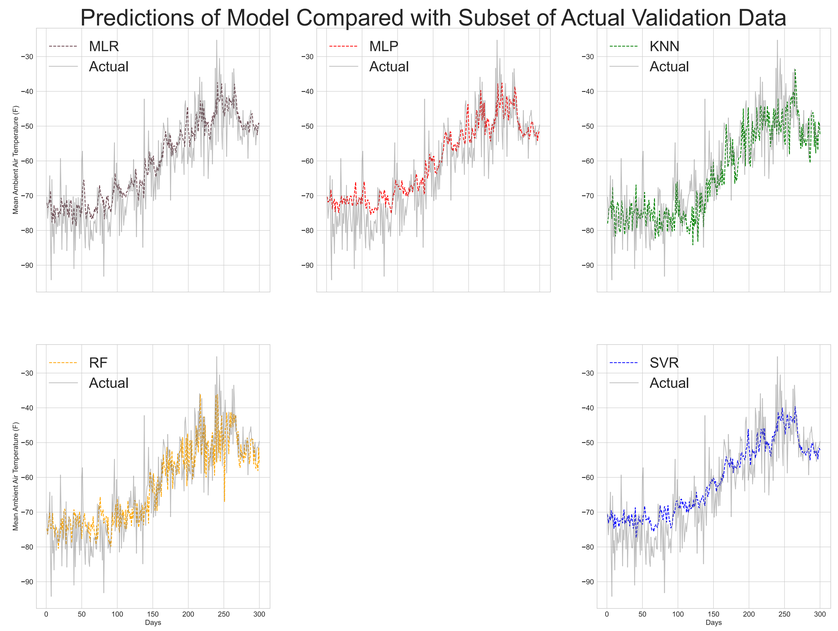

Fig. 3. Plots of Predictions Using Different Models.

Metric	Models				
	MLR	MLP	KNN	RF	SVR
RMSE in Fahrenheit (mean ± std)	7.05 ± 1.91	11.71 ± 6.34	7.75 ± 2.10	7.74 ± 2.17	7.78 ± 1.82

Table 1. RMSE for Different Models.

Fig. 4. Preliminary Results.

CONCLUSIONS AND FUTURE WORK

It was found that backward elimination followed by PCA eliminated multicollinearity and improved the MLR model. PCs of features were used as inputs for all other models. The RMSE for MLR was the lowest of all models tested. Other transformations for data may need to be considered due to violation of assumptions. Hyperparameter tuning for models other than MLR has not yet been conducted. Thus, the current algorithm comparison is not entirely indicative of optimal performance. In the future, multi-step time series forecasting will be implemented and models such as LSTM-RNN will be included.

REFERENCES

[1]. N. Anusha, M. S. Chaithanya and G. Reddy, "Weather Prediction Using Multi Linear Regression Algorithm," in IOP Conference Series: Materials Science and Engineering, Tamil Nadu, 2019.

[2]. S. Cramer, M. Kampouridis, A. A. Freitas and A. K. Alexandridis, "An extensive evaluation of seven machine learning methods for rainfall prediction in weather derivatives," Expert Systems with Applications, vol. 85, pp. 169-181, 2017.

ACKNOWLEDGEMENTS

This work was supported by MTSU Undergraduate Research Experience and Creative Activity funding. Dr. Sal Barbosa and Dr. Joshua Phillips also provided consultation on methods.