Leveraging Analytics to Predict Geomagnetic Storms

Impact to Global Telecommunications

Taylor K. Larkin[1] and Denise J. McManus[2]
Information Systems, Statistics, and Management Science
Culverhouse College of Commerce
The University of Alabama
Tuscaloosa, AL 35487-0226
Email: tklarkin@crimson.ua.edu[1], dmcmanus@cba.ua.edu[2]

***Abstract*—Coronal mass ejections are colossal bursts of magnetic field and plasma from the Sun. These eruptions can have disastrous effects on Earth's telecommunication systems and power grid infrastructures costing millions of dollars in damages. Hence, it is imperative to construct intelligent predictive processes to determine whether an incoming coronal mass ejection will produce devastating impacts on Earth. One such process, called "stacked generalization," is an ensemble strategy that incorporates the predictions from a diverse set of models (base-learners) by using them as inputs for another model (a meta-learner). The goal of this meta-learner is to deduce information about the biases from the base-learners and improve generalization to make more accurate predictions. In this work, 30 models are chosen from the R package caret to serve as base-learners in order to predict a geomagnetic storm index value associated for 2,811 coronal mass ejection events that occurred between 1996 and 2014. Two meta-learners are explored: 1) standard linear regression 2) non-negative elastic net regression. Results show that for this dataset, stacked generalization with the latter meta-learner produces the lowest error and performs significantly better than any of the base-learners executed individually. Not only does non-negative elastic net regression have predictive advantages, but it provides sparser solutions and more reliable inferences at the meta-level compared to linear regression. This, in turn, encourages the idea of parsimony and consequently, improves the overall generalization behavior of this technique.**

***Keywords*–geomagnetic storms; stacked generalization; regularization; predictive modeling.**

I. INTRODUCTION

Coronal Mass Ejections (CMEs) are massive explosions of magnetic field and plasma components from the Sun (shown in Fig. 1). Typically, a CME travels at speeds between 400 and 1,000 kilometers per second [1] resulting in an arrival time of approximately one to four days [2]; however, they can move as slowly as 100 kilometers per second or as quickly as 3,000 kilometers per second (or around 6.7 million miles per hour) [3]. These phenomena can contain a mass of solar material exceeding 10^{13} kilograms (or approximately 22 trillion pounds) [4] and can explode with the force of a billion hydrogen bombs [5]. Naturally, CME events are often associated with solar activity such as sunspots [3]. During the solar minimum of the 11 year solar cycle (the period of time where the Sun has fewer sunspots and hence, weaker magnetic fields), CME events occur about once a day. During a solar maximum, this daily estimate increases to four or five. One plausible theory for these incidents taking place involves the Sun needing to release energy. As more sunspots develop, more coronal magnetic field structures become entangled; therefore, more energy is required to control the volatility and convulsion. Once the energy surpasses a certain level, it becomes beneficial for the Sun to release these complex magnetic structures [1]. When this force approaches Earth, it

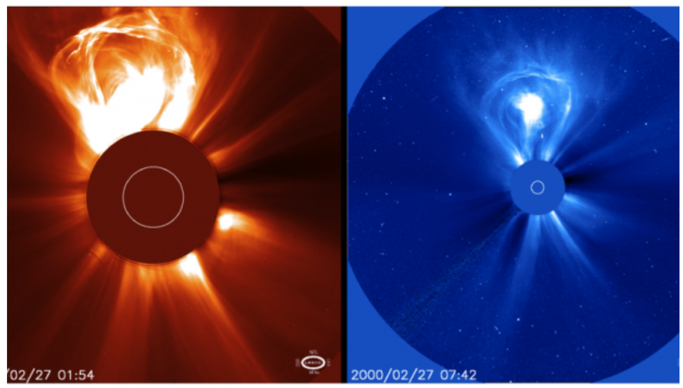

Figure 1. LASCO coronagraph images [3], courtesy of the NASA/ESA SOHO mission.

collides with the magnetosphere. The magnetosphere is the area encompassing Earth's magnetic field and serves as the line of defense against solar winds. The National Oceanic and Atmospheric Administration (NOAA) describes this event as "the appearance of water flowing around a rock in a stream" [6] as shown in Fig. 2. After the solar winds compress Earth's

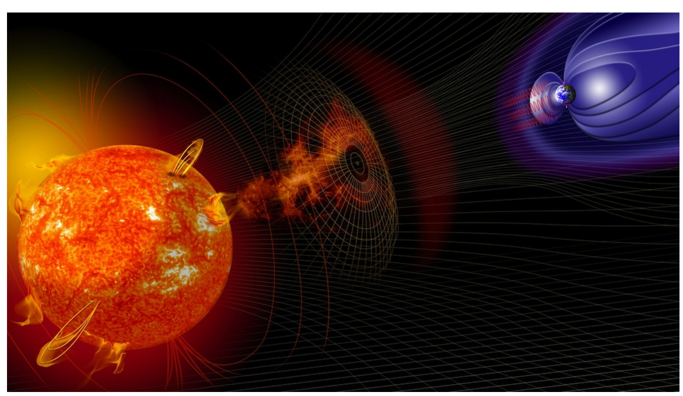

Figure 2. Rendering of Earth's magnetosphere interacting with the solar wind from the Sun [7], courtesy of the NASA's Goddard Space Flight Center.

magnetic field on the day side (the side facing the sun), they travel along the elongated magnetosphere into Earth's dark side (the side opposite of the Sun). The electrons are accelerated and energized in the tails of the magnetosphere. They filter down to the Polar Regions and clash with atmospheric gases causing geomagnetic storms. This energy transfer emits the brilliance known as the *Aurora Borealis*, or Northern Lights, and the *Aurora Australis*, or Southern Lights, which can be seen near the poles.

While mainly responsible for the illustrious Northern Lights, geomagnetic storms have the potential to cause cataclysmic damage to Earth. Normally, the magnetic field is able to deflect most of the incoming plasma particles from the Sun. However, when a CME contains a strong southward-directed magnetic field component (B_z), energy is transferred from the CME's magnetic field to Earth's through a process called magnetic reconnection [8][9][10] (as cited in [11]). Magnetic reconnection leads to an injection of plasma particles in Earth's geomagnetic field and a reduction of the magnetosphere towards the equator [1]. Consequently, more energy is amassed in the upper atmosphere, particularly at the poles. Moreover, this energy is impressed upon power transformers causing an acute over-saturation and inducing black-outs via geomagnetically induced currents (GICs) [12]. Some other residuals of this over-accumulation of energy include the corrosion of pipelines, deteriorations of radio and GPS communications, radiation hazards in higher latitudes, damages to spacecrafts, and deficiencies in solar arrays [13]. These ramifications pose a significant threat to global telecommunications and electrical power infrastructures as CMEs continue to be launched towards Earth [14]. From a business perspective, risk factor mitigation is an absolute necessity within the global business environment [15]. This can be accomplished using advanced analytical techniques on data collected about these phenomena.

The subsequent sections of this work read as follows. Section 2 briefly introduces some previous studies on predicting geomagnetic storms. Section 3 provides introductory detail about the basics of the methodology used, the dataset studied, and the experimental strategy. Section 4 displays and discusses the results. Section 5 concludes with a summary and postulates areas for future work.

II. BACKGROUND INFORMATION

A. Predicting Dangerous CMEs

CMEs present an ever-increasing threat to Earth as society becomes more dependent on technology, such as satellites and telecommunication operations. Nevertheless, because of this increase in technology, more data has been collected about these acts and solar wind in general. This, in turn, has allowed for empirical models to be developed. Burton, McPherron, and Russell [16] presented an algorithm to predict the disturbance storm index (DST) value [17] based on solar wind and interplanetary magnetic field parameters. The DST value is a popular metric to assess geomagnetic activity. Expressed in nanoteslas (nT) and recorded every hour from observatories around the world, it measures the depression of the equatorial geomagnetic field, or horizontal component of the magnetic field; thus, the smaller the value of the DST, the more significant the disturbance of the magnetic field [1]. Many researchers have used this information for building forecasting models to predict geomagnetic storms [18][19].

However, many of these systems only use *in-situ* data, or data that can only be measured close to Earth. To improve prediction, studies have included data from both initial CME observations and the near-Earth solar wind condition [20][21][22], especially considering CMEs remain the source of major geomagnetic disturbances [23][24][25] (as cited in [26]). These have ranged from using logistic regression [21] to neural networks [27][28] to make predictions based on this combination of data. Aside from the work by Dryer et al. [20], which used an ensemble of four physics-based models to predict shock arrival times, the idea of using ensembles of models has not been very prevalent in the literature. Stacked generalization is a type of ensemble that uses the individual predictions from a set of base models as inputs for another model to make a final prediction. This strategy has been the backbone of successful schemes in areas such as predicting financial fraud [29], bankruptcy [30], and user ratings in the famous Netflix Prize competition [31]. Therefore, leveraging more advanced ensemble frameworks for predictive modeling has the opportunity to increase accuracy in this field.

B. Stacked Generalization

The idea of stacked generalization was originally proposed by David Wolpert [32]. It can be simplified in the following way:

- Construct a dataset consisting of predictions from a set of level 0 (or base) learners using a training and a test set. Refer to this as the metadata, *MD*.
- Generate a level 1 (or meta) learner that utilizes the predictions made at the previous level as inputs. That is, train the meta-learner on *MD* as opposed to the original training data.

Often times, the predictions from the base-learners are determined via k-fold cross-validation [33]. Define the dataset $S = \{(y_i, \boldsymbol{x}_i), i = 1, ..., n\}$ where $\boldsymbol{x}_i$ is a vector of predictor variables and y_i is the corresponding response value for the i^{th} observation. Specifically, split the dataset S into k near equal and disjoint sets such that $S_1, S_2, ..., S_k$. Let $S^{-k} = S - S_k$ and S_k be the training and test sets, respectively. Execute the base-learner on the first S^{-k} parts and produce a prediction for the held-out part S_k. Repeat this procedure until each subset of S has been used as a test set. Extract all the hold-out predictions to create *MD*. Because generating the metadata is an independent process across each base-learner, it can be parallelized for faster computation. That is, each base-learner can be trained at the same time. This is key as time plays a pivotal role in geomagnetic storm prediction [22].

The meta-learner's purpose is to gain information about the generalization behavior of each learner trained at the base-level. Popular choices for meta-learners have been linear models [34], especially those with a non-negativity constraint on the estimated coefficients in regression type problems [35][33]. While this ensemble strategy leverages the strengths and weaknesses of the base-learners, it can be prone to overfitting [36]. Therefore, in order to combat this issue, employing regularized linear methods can perform better than their non-regularized counterparts. Reid and Grudic [37] experimented with three regularization penalties: ridge [38], lasso [39]], and elastic net [40]. The authors showed that using stacked generalization with ridge regression as the meta-learner performs well on multi-class datasets. Their findings make sense given the advantages of ridge regression for highly correlated data [38], a natural consequence of well-tuned base-learner predictions. In addition, they commented that using the lasso and elastic net penalties can promote sparse solutions that can reduce the size of the ensemble at the meta-level. Pruning the size of an ensemble model has been explored in other works [41][42][43]. It can lead to better generalization and promote the necessary diversity in the base-learner predictions, or the ensemble members [32].

C. Review of Ridge, Lasso, and Elastic Net

Recall the ordinary least squares (OLS) solution for the coefficients in linear regression [44]:

$$\hat{\beta}^{ols} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y} \tag{1}$$

where $\boldsymbol{X}$ is the predictor matrix of dimension $n \times (p+1)$ and $\boldsymbol{Y}$ is the vector of outcomes of dimension $n \times 1$ for n observations and p predictor variables. In particular,

$$\boldsymbol{X} = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ 1 & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix} \quad \boldsymbol{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

Alternatively, equation (1) can be written as

$$\hat{\beta}^{ols} = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^{n} (y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2 \tag{2}$$

When multicollinearity exists in the predictor matrix $\boldsymbol{X}$, estimates for β can become erratic and demonstrate a large amount of variability. Large positive values of β cancel out with equally large negative values which cause issues for meaningfully interpreting the coefficients [38] [44]. This stems from the predictor matrix $\boldsymbol{X}'\boldsymbol{X}$ being nearly singular. Hoerl and Kennard [38] showed that a smaller mean square error can be achieved through the use of adding a positive constant λ to the diagonal of the predictor matrix

$$\hat{\beta}^{ridge} = (\boldsymbol{X}'\boldsymbol{X} + \lambda I)^{-1}\boldsymbol{X}'\boldsymbol{Y} \tag{3}$$

such that I is a $p \times p$ identity matrix. This makes the solution invertible even if the predictor matrix is not full rank. Hastie, Tibshirani, and Friedman [44] defined ridge regression as a shrinkage method which can be expressed as an optimization problem

$$\begin{aligned} &\underset{\beta}{\text{minimize}} \left\{ \sum_{i=1}^{n} (y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2 \right\} \\ &\text{subject to} \quad \sum_{j=1}^{p} \beta_j^2 \leq t, \quad t \geq 0 \end{aligned} \tag{4}$$

such that t controls the amount of shrinkage, or penalty, to introduce.

The ability of ridge regression to effectively penalize the coefficients in a continuous fashion mitigates the unstable behavior of the coefficients in the presence of multicollinearity. However, this penalty only shrinks the coefficients towards zero; hence, no variable selection is taking place. Tibshirani [39] posited lasso which imposes the following formulation:

$$\begin{aligned} &\underset{\beta}{\text{minimize}} \left\{ \sum_{i=1}^{n} (y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2 \right\} \\ &\text{subject to} \quad \sum_{j=1}^{p} |\beta_j| \leq t, \quad t \geq 0 \end{aligned} \tag{5}$$

This constraint allows for some of the coefficients to shrink completely to zero for a sufficiently small t. While this has become a wildly popular technique due to its sparse nature, it can have some drawbacks. For instance, it may only select one predictor variable from a highly correlated group. Hence, Zou and Hastie [40] offered the elastic net penalty

$$\begin{aligned} &\underset{\beta}{\text{minimize}} \left\{ \sum_{i=1}^{n} (y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2 \right\} \\ &\text{subject to} \quad \sum_{j=1}^{p} ((1-\alpha)\beta_j^2 + \alpha|\beta_j|) \leq t, \quad t \geq 0 \end{aligned} \tag{6}$$

which is a convex combination of both ridge and lasso penalties. Note that when $\alpha = 0$ this reduces to the ridge penalty while $\alpha = 1$ is equivalent to lasso. Thus, the elastic net can shrink coefficients to exactly zero while also handling groups of correlated predictor variables.

D. A Suitable Meta-learner

As noted before, it has been shown that a non-negative constrained linear model performs well in stacked generalization for regression tasks. Given the popularity of the penalty functions mentioned in the previous section, a natural extension is to implement a meta-learner that combines these constraints into one model. Recently, Mandal and Ma [45] proposed an efficient multiplicative iterative path algorithm to estimate the entire regularization path for a variety of non-negative generalized linear models with ridge, lasso, and elastic net penalties. Therefore, to capitalize on the results of previous works, this work institutes a regularized linear model with a non-negativity constraint as the meta-learner. That is, with the elastic net Gaussian objective function [46], the following optimization problem is formed:

$$\begin{aligned} &\underset{(\beta_0,\beta)\in\mathbb{R}^{p+1}}{\text{minimize}} \left\{ \frac{1}{2n} \sum_{i=1}^{n} (y_i - \beta_0 - x_i'\beta)^2 \right. \\ &\qquad\qquad \left. + \lambda\Big[\frac{(1-\alpha)}{2}||\beta||_2^2 + \alpha||\beta||_1\Big] \right\} \\ &\text{subject to} \quad \beta, \lambda \geq 0, \quad 0 \leq \alpha \leq 1 \end{aligned} \tag{7}$$

In this case, λ has a one-to-one correspondence with t and is considered the shrinkage parameter. By regularizing the coefficients with both a non-negativity constraint and a penalty function, sparse solutions may be realized, even in the presence of high correlation. This meta-learner will be referred to as the non-negative elastic net (NNEN). For comparison, standard linear regression will also be implemented at the meta-level.

 ISBN: 978-1-61208-510-4

III. Methodology

A. Data

Four sources are considered to construct the experimental dataset: near-Earth CME information provided by Richardson and Cane [47] [48], OMNI 2 hourly averaged solar wind data at one AU from the Coordinated Data Analysis (Workshop) Web [49], CME measurements given by the Large Angle and Spectrometric Coronagraph (LASCO) located on the Solar and Heliospheric Observatory (SOHO) satellite [50], and some Sun characteristics recorded by NOAA [51]. These data are combined so that each CME has been assigned interplanetary variable values (such as B_z) prior to the DST minimum during a predicted area of effect on Earth. Establishing these values before the DST minimum gives a lead time prior to the climax of the geomagnetic storms and allows for a more realistic prediction scenario. Also included are the initial measurements about the speed and size of a CME at the time of ejection from the Sun and daily Sun characteristics on the day of ejection. After filtering out missing values and some unnecessary rows, a dataset composed of 2,811 CME events from 1996 to 2014 with 28 predictor variables is ready for analysis.

B. Implementation

1) Experimental Set-up: The analysis is performed in the R environment version 3.2.5 [52], mainly by using the **caret** (Classification And REgression Training) package [53]. This package allows for a streamlined user interface for applying various sets of predictive models from different packages. It has options to perform several resampling techniques to tune model parameters and create visualizations of model performance. The stacked generalization framework is constructed using models from this package. The amount of parameter tuning for all models is assigned at **caret**'s default value with the final parameter combinations determined by those which deliver the lowest root mean square error (RMSE).

The RMSE is calculated from an average of ten repeats of 10-fold (10×10) nested cross-validation to ensure a good estimation of error in the presence of parameter tuning [54]. Furthermore, significance tests between the meta-learner and the individual base-learners are conducted on the population of RMSEs (100 estimates from the 10×10 nested cross-validation) using the corrected repeated k-fold cross-validation test [55]. It is important to test for significant differences to investigate if the extra computation of stacked generalization is worth the effort compared to simply using the best performing model [56]. All base-learners and meta-learners are trained over the same folds with the only difference being that the meta-learners use *MD* as its inputs instead of the CME predictor variables. *MD* is generated using 10-fold cross-validation [34]. Note that this cross-validation is separate from the nested cross-validation used to estimate the error.

2) Learners: Care is taken to make sure a diverse set of base-learners is utilized [41]. The complete list of the 30 models chosen can be found in Table I. As for implementing the NNEN, a custom model is created within the **caret** framework using the *nnlasso* function from the **nnlasso** R package developed by Mandal and Ma [57]. It is important to use a custom model so that this learner is trained on the same folds as the other learners and so that the variable selection takes place within the training folds [44]. As with the popular *glmnet* function in the **glmnet** package [58], two main parameters are tuned in *nnlasso*: the mixing weights of the ridge and lasso penalties α and the amount of shrinkage to be applied λ. For this work, three values of α, $\{0, 0.5, 1\}$, are tested for each iteration in the 10×10 nested cross-validation process. For determining the amount of shrinkage, the values of λ are identified the same way as in **caret**'s implementation of *glmnet*. For the comparison meta-learner, linear regression is executed at the meta-level by calling **caret**'s *lm* method.

TABLE I. List of Base-learners

Name	**caret** Method
Bayesian Lasso Regression	*blasso*
Bayesian Regularized Neural Network	*brnn*
Bayesian Ridge Regression	*bridge*
Boosted Linear Model	*BstLm*
Boosted Tree	*bstTree*
Classification and Regression Tree	*rpart*
Conditional Inference Random Forest	*cforest*
Conditional Inference Tree	*ctree*
Cubist	*cubist*
Extreme Gradient Boosting with Linear Booster	*xgbLinear*
Extreme Learning Machine	*elm*
Generalized Additive Model using Splines	*gamSpline*
k-Nearest Neighbors	*kknn*
Lasso and Elastic Net Regression	*glmnet*
Least Angle Regression	*lars*
Linear Regression	*lm*
Linear Regression with Stepwise Selection	*leapSeq*
Multi-Layer Perceptron	*mlp*
Multivariate Adaptive Regression Splines	*earth*
Neural Network with Feature Extraction	*pcaNNet*
Non-Convex Penalized Quantile Regression	*rqnc*
Partial Least Squares	*pls*
Quantile Random Forest	*qrf*
Random Forest	*ranger*
Self-Organizing Map	*bdk*
Spike and Slab Regression	*spikeslab*
Stacked AutoEncoder Deep Neural Network	*dnn*
Stochastic Gradient Boosting	*gbm*
Supervised Principal Component Analysis	*superpc*
Support Vector Machine with Radial Basis Function Kernel	*svmRadialSigma*

IV. Results and Discussion

Table II reflects the results of the analysis. The first column lists both meta-learners and the ten most accurate base-learners ranked in ascending order by the average RMSE displayed in the second column. The third column represents the average RMSE for those CME events which triggered a strong geomagnetic disturbance (DST value ≤ -100 nT) [59]. The asterisk denotes instances where a significant difference between NNEN and the other learners are *not* found at the conventional 0.05 significance level.

Not surprisingly, the top five base-learners are all bagging or boosting ensemble models. However, NNEN yields the lowest RMSE compared to these and all the other learners, even for the strong CME events. In addition, NNEN performs statistically better than all of the base-learners for all CME events and better than the majority for the strong ones. This provides evidence that the implementation of stacked generalization here has more predictive power than using just one model. Ting and Witten [34] indicated in their analysis that stacked generalization delivers substantial improvements in accuracy for larger datasets. This is likely due to a more accurate estimation from the cross-validation process when generating the metadata. Hence, it is probable that with more data, stacked generalization can continue to enhance geomagnetic storm prediction. While the predictive improvements may seem minor at a higher cost in computation, this work, as well as others, show that this ensemble technique can lead to statistically significant improvements even against sophisticated, well-tuned models. Because of the danger that these geomagnetic storms present, even marginal improvements can make a difference.

TABLE II. PREDICTIVE PERFORMANCE

Learner	All CMEs	Strong CMEs
NNEN	**17.64**	**45.35**
Linear Regression	17.75*	45.38*
Random Forest	18.17	49.50
Cubist	18.19	47.61*
Conditional Inference Random Forest	19.06	54.28
Boosted Tree	19.10	51.87
Stochastic Gradient Boosting	19.38	52.33
Extreme Gradient Boosting with Linear Booster	19.46	51.97*
Multivariate Adaptive Regression Splines	19.81	53.29
Generalized Additive Model using Splines	20.06	55.56
Bayesian Regularized Neural Network	20.11	54.24
k-Nearest Neighbors	22.12	67.11

Although it shows its superiority over the base-learners, NNEN does not perform statistically different than simply using linear regression as the meta-learner with only very minor increases in predictive performance. While these techniques may seem equal here, recall that linear regression does not inherently perform variable selection; thus, it utilizes all 30 base-learner predictions. In addition, any attempt at making any inference regarding the coefficients at the meta-level is frivolous due to the high amount of correlation and likely presence of negative coefficients. On the other hand, NNEN selected only 19.39 ensemble members on average during the resampling process. Furthermore, some interpretation regarding the contribution of each ensemble member to the final prediction can be made by analyzing the positive, non-zero coefficients. Hence, NNEN should be preferred over standard linear regression for this dataset since it can produce sparser and more interpretable solutions with statistically similar error. The quality of being able to dynamically select which base-learners are most useful for prediction at the meta-level may help improve on the fixed form bias issues of stacked generalization mentioned by Vilalta and Drissi [60].

Further study of the performance of the base-learners offers interesting directions for future work, specifically with the use of quantile models. The Quantile Random Forest finishes as the 26^{th} least accurate base-learner with a value of 31.60; however, its RMSE on strong CME events has a value of 46.02, which is nearly as accurate as the NNEN. Given that these strong CME events do not occur very often (131 in this dataset) but pose the most risk to society, it makes sense to adapt the meta-learner to focus on a specific quantile as opposed to the conditional mean. In this way, by focusing more on the outliers, better predictions can be made on the more important observations. Naturally, a balance would need to be constructed so that the predictions for weaker storms are not rendered useless.

V. SUMMARY

In this work, stacked generalization is used to predict geomagnetic storms driven by CMEs. Using data from a variety of sources, this technique is executed on a realistic dataset to investigate its predictive performance against using only one statistical model or machine learning algorithm. Based on insights from previous research about regularization and pruning, a NNEN model is implemented. NNEN shows its advantages on this dataset in terms of predictive accuracy while using fewer base-learner predictions. Future work consists of implementing more data for geomagnetic storm prediction to see if further improvements can be made with this methodology. In addition, increasing the number of base-learners can give NNEN more opportunities to find an optimal combination of ensemble members. Moreover, exploring the usage of regularized quantile methods can provide a useful alternative than typical mean predictions. Given the potential cataclysmic damage that CMEs can wreak on telecommunications and power companies, advanced techniques for improving accuracy are an absolute necessity for saving these industries millions of dollars.

ACKNOWLEDGMENT

We would like to thank NASA for their images and the creation of the CME catalog. This CME catalog is generated and maintained at the CDAW Data Center by NASA and The Catholic University of America in cooperation with the Naval Research Laboratory. SOHO is a project of international cooperation between ESA and NASA. In addition, we would like to thank the Goddard Space Flight Center/Space Physics Data Facility (GSFC/SPDF), OMNIWeb, and NOAA for their public use databases.

REFERENCES

[1] T. Howard, Coronal mass ejections: An introduction. Springer Science & Business Media, 2011, vol. 376.

[2] N. Srivastava and P. Venkatakrishnan, "Solar and interplanetary sources of major geomagnetic storms during 1996–2002," Journal of Geophysical Research: Space Physics (1978–2012), vol. 109, no. A10, 2004, pp. 1–13.

[3] N. Oceanic and A. Administration, "Coronal mass ejections," Available: http://www.swpc.noaa.gov/phenomena/coronal-mass-ejections [accessed: 2016-07-27].

[4] R. MacQueen, "Coronal transients: A summary," Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, vol. 297, no. 1433, 1980, pp. 605–620.

[5] N. Aeronautics and S. Administration, "Coronal mass ejections," Available: http://helios.gsfc.nasa.gov/cme.html [accessed: 2016-07-27].

[6] N. Oceanic and A. Administration, "Earth's magnetosphere," Available: http://www.swpc.noaa.gov/phenomena/earths-magnetosphere [accessed: 2016-07-27].

[7] N. Aeronautics and S. A. G. S. F. Center, "Rattling earth's force field," Available: https://svs.gsfc.nasa.gov/10954 [accessed: 2016-07-27].

[8] J. W. Dungey, "Interplanetary magnetic field and the auroral zones," Physical Review Letters, vol. 6, no. 2, 1961, pp. 47–48.

[9] D. H. Fairfield and L. Cahill, "Transition region magnetic field and polar magnetic disturbances," Journal of Geophysical Research, vol. 71, no. 1, 1966, pp. 155–169.

[10] W. D. Gonzalez and B. T. Tsurutani, "Criteria of interplanetary parameters causing intense magnetic storms (dst <- 100 nt)," Planetary and Space Science, vol. 35, no. 9, 1987, pp. 1101–1109.

[11] Y. Wang, P. Ye, S. Wang, G. Zhou, and J. Wang, "A statistical study on the geoeffectiveness of earth-directed coronal mass ejections from march 1997 to december 2000," Journal of Geophysical Research: Space Physics (1978–2012), vol. 107, no. A11, 2002, pp. SSH 2–1–SSH 2–9.

[12] J. Kappernman and V. D. Albertson, "Bracing for the geomagnetic storms," Spectrum, IEEE, vol. 27, no. 3, 1990, pp. 27–33.

[13] S. S. Board et al., Severe Space Weather Events–Understanding Societal and Economic Impacts: A Workshop Report. National Academies Press, 2008.

[14] D. Baker, X. Li, A. Pulkkinen, C. Ngwira, M. Mays, A. Galvin, and K. Simunac, "A major solar eruptive event in july 2012: Defining extreme space weather scenarios," Space Weather, vol. 11, no. 10, 2013, pp. 585–591.

[15] D. McManus, H. Carr, and B. Adams, "Wireless on the precipice: The 14 th century revisited," Communications of the ACM, vol. 54, no. 6, 2011, pp. 138–143.

[16] R. K. Burton, R. McPherron, and C. Russell, "An empirical relationship between interplanetary conditions and dst," Journal of geophysical research, vol. 80, no. 31, 1975, pp. 4204–4214.

[17] M. Sugiura, "Hourly values of equatorial dst for the igy," Ann. int. geophys. Yr., vol. 35, 1964, pp. 1–44.

[18] E.-Y. Ji, Y.-J. Moon, N. Gopalswamy, and D.-H. Lee, "Comparison of dst forecast models for intense geomagnetic storms," Journal of Geophysical Research: Space Physics, vol. 117, no. A3, 2012, pp. 1–9.

[19] T. Andriyas and S. Andriyas, "Relevance vector machines as a tool for forecasting geomagnetic storms during years 1996–2007," Journal of Atmospheric and Solar-Terrestrial Physics, vol. 125, 2015, pp. 10–20.

[20] M. Dryer, Z. Smith, C. Fry, W. Sun, C. Deehr, and S.-I. Akasofu, "Real-time shock arrival predictions during the halloween 2003 epoch," Space Weather, vol. 2, no. 9, 2004, pp. 1–10.

[21] N. Srivastava, "A logistic regression model for predicting the occurrence of intense geomagnetic storms," in Annales Geophysicae, vol. 23, no. 9, 2005, pp. 2969–2974.

[22] R.-S. Kim, Y.-J. Moon, N. Gopalswamy, Y.-D. Park, and Y.-H. Kim, "Two-step forecast of geomagnetic storm using coronal mass ejection and solar wind condition," Space Weather, vol. 12, no. 4, 2014, pp. 246–256.

[23] J. Gosling, S. Bame, D. McComas, and J. Phillips, "Coronal mass ejections and large geomagnetic storms," Geophysical Research Letters, vol. 17, no. 7, 1990, pp. 901–904.

[24] V. Bothmer and R. Schwenn, "The interplanetary and solar causes of major geomagnetic storms." Journal of geomagnetism and geoelectricity, vol. 47, no. 11, 1995, pp. 1127–1132.

[25] B. T. Tsurutani and W. D. Gonzalez, "The interplanetary causes of magnetic storms: A review," Washington DC American Geophysical Union Geophysical Monograph Series, vol. 98, 1997, pp. 77–89.

[26] J. Zhang, K. Dere, R. Howard, and V. Bothmer, "Identification of solar sources of major geomagnetic storms between 1996 and 2000," The Astrophysical Journal, vol. 582, no. 1, 2003, pp. 520–533.

[27] J. Uwamahoro, L. McKinnell, and J. Habarulema, "Estimating the geo-effectiveness of halo cmes from associated solar and ip parameters using neural networks," Annales Geophysicae-Atmospheres Hydrospheresand Space Sciences, vol. 30, no. 6, 2012, pp. 963–972.

[28] A. Singh and P. Mishra, "Prediction of intense geomagnetic storms using artificial neural network," International Journal of Advances in Earth Sciences, vol. 4, no. 1, 2015, pp. 1–7.

[29] A. Abbasi, C. Albrecht, A. Vance, and J. Hansen, "Metafraud: a meta-learning framework for detecting financial fraud," Mis Quarterly, vol. 36, no. 4, 2012, pp. 1293–1327.

[30] C.-F. Tsai and Y.-F. Hsu, "A meta-learning framework for bankruptcy prediction," Journal of Forecasting, vol. 32, no. 2, 2013, pp. 167–179.

[31] J. Sill, G. Takács, L. Mackey, and D. Lin, "Feature-weighted linear stacking," arXiv preprint arXiv:0911.0460, 2009, pp. 1–17.

[32] D. H. Wolpert, "Stacked generalization," Neural networks, vol. 5, no. 2, 1992, pp. 241–259.

[33] L. Breiman, "Stacked regressions," Machine learning, vol. 24, no. 1, 1996, pp. 49–64.

[34] K. M. Ting and I. H. Witten, "Issues in stacked generalization," J. Artif. Intell. Res.(JAIR), vol. 10, 1999, pp. 271–289.

[35] M. LeBlanc and R. Tibshirani, "Combining estimates in regression and classification," Journal of the American Statistical Association, vol. 91, no. 436, 1996, pp. 1641–1650.

[36] R. Caruana, A. Niculescu-Mizil, G. Crew, and A. Ksikes, "Ensemble selection from libraries of models," in Proceedings of the twenty-first international conference on Machine learning. ACM, 2004, pp. 18–25.

[37] S. Reid and G. Grudic, "Regularized linear models in stacked generalization," in Multiple Classifier Systems. Springer, 2009, pp. 112–121.

[38] A. E. Hoerl and R. W. Kennard, "Ridge regression: Biased estimation for nonorthogonal problems," Technometrics, 1970, pp. 55–67.

[39] R. Tibshirani, "Regression shrinkage and selection via the lasso," Journal of the Royal Statistical Society. Series B (Methodological), 1996, pp. 267–288.

[40] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," Journal of the Royal Statistical Society: Series B (Statistical Methodology), vol. 67, no. 2, 2005, pp. 301–320.

[41] G. Zenobi and P. Cunningham, "Using diversity in preparing ensembles of classifiers based on different feature subsets to minimize generalization error," in Machine Learning: ECML 2001. Springer, 2001, pp. 576–587.

[42] Z.-H. Zhou, J. Wu, and W. Tang, "Ensembling neural networks: many could be better than all," Artificial intelligence, vol. 137, no. 1, 2002, pp. 239–263.

[43] N. Rooney, D. Patterson, and C. Nugent, "Pruning extensions to stacking," Intelligent Data Analysis, vol. 10, no. 1, 2006, pp. 47–66.

[44] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning. New York: Springer, 2009.

[45] B. Mandal and J. Ma, "l1 regularized multiplicative iterative path algorithm for non-negative generalized linear models," Computational Statistics & Data Analysis, vol. 101, 2016, pp. 289–299.

[46] T. Hastie and J. Qian, "Glmnet vignette," 2014.

[47] H. Cane and I. Richardson, "Interplanetary coronal mass ejections in the near-earth solar wind during 1996–2002," Journal of Geophysical Research: Space Physics (1978–2012), vol. 108, no. A4, 2003, pp. SSH 6–1–SSH 6–13.

[48] I. Richardson and H. Cane, "Near-earth interplanetary coronal mass ejections during solar cycle 23 (1996–2009): Catalog and summary of properties," Solar Physics, vol. 264, no. 1, 2010, pp. 189–237.

[49] J. King and N. Papitashvili, "Solar wind spatial scales in and comparisons of hourly wind and ace plasma and magnetic field data," Journal of Geophysical Research: Space Physics, vol. 110, no. A2, 2005, pp. 1–8.

[50] N. Gopalswamy, S. Yashiro, G. Michalek, G. Stenborg, A. Vourlidas, S. Freeland, and R. Howard, "The soho/lasco cme catalog," Earth, Moon, and Planets, vol. 104, no. 1-4, 2009, pp. 295–313.

[51] N. Oceanic and A. Administration. Index of /pub/warehouse. Available: ftp://ftp.swpc.noaa.gov/pub/warehouse [accessed: 2016-07-27].

[52] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2016, Available: https://www.R-project.org/ [accessed: 2016-07-27].

[53] M. K. C. from Jed Wing, S. Weston, A. Williams, C. Keefer, A. Engelhardt, T. Cooper, Z. Mayer, B. Kenkel, the R Core Team, M. Benesty, R. Lescarbeau, A. Ziem, L. Scrucca, Y. Tang, and C. Candan., caret: Classification and Regression Training, 2016, Available: https://CRAN.R-project.org/package=caret [accessed: 2016-07-27].

[54] S. Varma and R. Simon, "Bias in error estimation when using cross-validation for model selection," BMC bioinformatics, vol. 7, no. 1, 2006, p. 91.

[55] R. R. Bouckaert and E. Frank, "Evaluating the replicability of significance tests for comparing learning algorithms," in Advances in knowledge discovery and data mining. Springer, 2004, pp. 3–12.

[56] S. Džeroski and B. Ženko, "Is combining classifiers with stacking better than selecting the best one?" Machine learning, vol. 54, no. 3, 2004, pp. 255–273.

[57] B. N. Mandal and J. Ma, nnlasso: Non-Negative Lasso and Elastic Net Penalized Generalized Linear Models, 2016, Available: https://CRAN.R-project.org/package=nnlasso [accessed: 2016-07-27].

[58] J. Friedman, T. Hastie, and R. Tibshirani, "Regularization paths for generalized linear models via coordinate descent," Journal of statistical software, vol. 33, no. 1, 2010, pp. 1–22.

[59] C. Loewe and G. Prölss, "Classification and mean behavior of magnetic storms," Journal of Geophysical Research: Space Physics, vol. 102, no. A7, 1997, pp. 14 209–14 213.

[60] R. Vilalta and Y. Drissi, "A perspective view and survey of meta-learning," Artificial Intelligence Review, vol. 18, no. 2, 2002, pp. 77–95.