Final Report

BNCS411

Forest Fire Prediction

Anton SCHMIDT
ID: 2020952069

Dong Jun KIM
ID: 2015170302

Jun Suk LEE
ID: 2016250208

TA: Jee Seok YOON

Seoul, 2020/12/13

Abstract

With the high media coverage of forest fires in the Amazon rainforest or Australia recently, this paper introduces a research on a machine learning based warning system that takes weather forecast data and predicts a scale indicating the fire danger on that specific day. The data used in this research is for Brazil, but in theory any country would work. The research considers KNN, Decision-Tree, Random-Forest, and Deep-Neural-Network(DNN) for the prediction of fire risk fed with various parameters like Temperature boundaries, Humidity and Wind speeds for the next day. The risk is categorized as an output of a value from 1 to 4 with 1 being a very low chance for forest fires and 4 being a very high chance for forest fires to occur. The research regards the data from 2001 to 2018 as a train data. The year 2019 was considered for the test of the trained model, and DNN returns about 58% accuracy, which showed to be superior to other classifiers. In analysis, the research discusses the correlation between weather and forest fires in regards of Brazilian Amazon. In addition, other external factors were analyzed as a possible source of unpredictability.

Table of content

1 Introduction

In the first week of September of 2020, the Amazon has burned more than twice as much as it did at that time last year.[1] While covid19 is occupying our news feeds, the biggest rainforest reservoir of the earth is disappearing at a steady rate. The first ten months of 2020 showed an increase of 25% compared to 2019.[2]
With a total forest loss of 718,927m² since 1970 in the Brazilian rainforest alone, we now stand at a loss of 17% of the natural forest cover that was once present.[3]

Considering the last years media coverage of forest fires like in Brazil and Australia, we asked ourselves if there was a way to address the problem using our machine learning skills. If the results showed that there was any indicative correlation between weather conditions and occurring forest fires, wanted to create a prediction system that could daily weather parameters such as temperature and wind speeds, and output a value on a scale that would be indicative of the fire-danger on that specific day.
Our decision fell on Brazil as the country is inhabiting more than 60% of the Amazon rainforest and therefore plays a prominent role when it comes to forest fires. This also meant that there was already a big audience gathering valuable data for us to use.
In addition to that, Brazil has probably the worst outlook on the future as the political situation does not particularly favour a radical change, but instead is heavily critiqued for their seemingly lacklustre countermeasures.[4] The political situation may also affect one of the major causes in Brazilian forest fires. Despite being illegal, deforestation is on a surge since 2019[Fig.1] with observed record numbers for 2020.[5]
Research is also indicating that climate change is affecting the regularity of droughts. Brazil would be no exception to this phenomenon and therefore could cause an increase in severity of the yearly Amazon fires.[6]

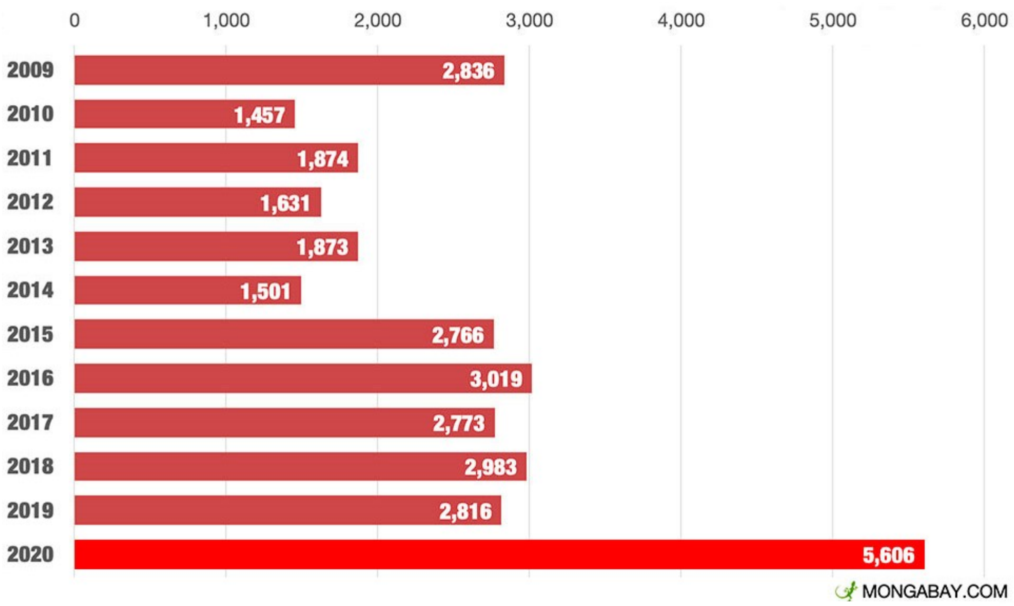


Figure 1 INPE: Accumated Amazon deforestation Aug1 thru Jan 31 (sq km)

2 Pre-processing

2.1 Weather Dataset

As the basis of our classification model we decided to use a collection of meteorological records collected in conventional weather stations[Fig. 2] of the Brazilian National Institute of Meteorology.[7] The data consists of more than 12 million rows, capturing a timespan of nearly 60 years between 1961 and 2019.

The original dataset holds 19 columns of which we decided on four as being the features we went on to use to train our model on.The challenge in selecting the features presented itself in multiple factors. The broad goal of the project was for the application to be applicable to other countries apart from Brazil, therefore a driving factor was choosing features that had a global availability and were not specific to our dataset. With this decision, features that might have helped improve the quality of our final model such as atmospheric pressure at sea level or average compensated temperature had to be excluded. Further we decided against the use of spatial data such as wind direction as this would have driven up the cost of computation as well as the complexity of the project. Our scope was to maintain the accessibility of the application as low as possible while still providing a useful prediction. The third and most important factor that influenced the final choice of features was the overall usefulness of the data. We decided that features like Piche evaporation, insolation and cloudiness were redundant to the final features we settled on.

We deemed following features to be the best choice:

- Maximum temperature (C°)
- Minimum temperature (C°)
- Average wind speed (m/s)
- Relative humidity (%)

The reasoning behind including both MinTemp as well as Max was to eliminate extreme values and to get a better daily average. Average wind speed and relative humidity were included as they are easily accessible and likely to be found in comparable datasets as well as them being red-flag indicators of fire danger according to the National Weather Service of the National Oceanic and Atmospheric Administration.[8]

Figure 2 Positions of INMET conventional weather stations

2.2 Fire Dataset

As the main source of historical fire data, we chose to use the collected data of NASA's MODIS (Moderate Resolution Imaging Spectroradiometer) instrument. MODIS is collecting daily data in 36 spectral bands, capturing natural processes happening on the surface.[9] The pixels identified as fires can be as small as 50m² (under perfect conditions).[10]

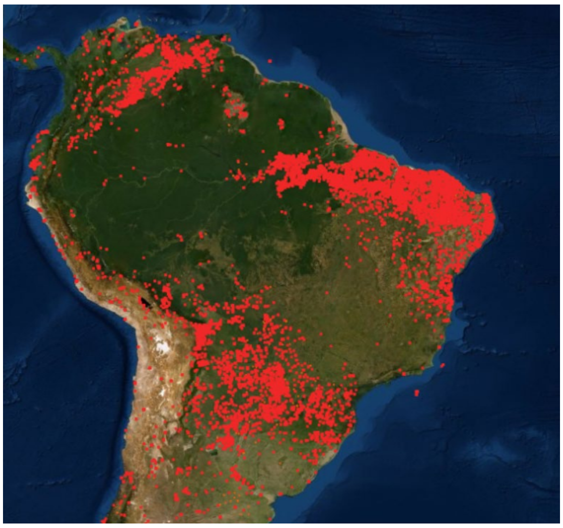


Figure 3 Fires collected my MODIS over the last 24h

To avoid false positives, we took advantage of the confidence value give for each datapoint and discarded any fires below 30%, to filter out ‘low’ confidence values.

Range	Confidence Class
$0\% \leq C < 30\%$	low
$30\% \leq C < 80\%$	nominal
$80\% \leq C \leq 100\%$	high

Figure 4 Fire-pixel confidence classes associated with the confidence level C computed for each fire pixel.

Next, we applied the already in the dataset provided filter to only focus on presumed vegetation fire. The remaining relevant features after filtering were therefore:

- Latitude at the centre of fire pixel
- Longitude at the centre of fire pixel
- UTC year, month, day

With MODIS’ collection of fire data starting in November 2001 we were forced to use this timestamp as starting point for the final timespan used to train the model. This still left us with around 7 million detected fire pixels from 2001 to 2019.

2.2.1 Reverse geocoding

Early on we decided that the separation between the 26 Brazilian states was crucial for the danger scale to be accurate. As seen in Figure 3, the detected fire pixels tend to cluster around certain areas more than others, resulting in a very uneven distribution between states. [Fig. 5] The underlying problem was the missing state feature on our fire data. As we were only provided with Latitude/Longitude values for each fire pixel we went on use the google reverse geocoding API to gain Address data. The scope of our application focuses only on a research field and has no commercial purpose. Therefore, the usage of the API did not scale well regarding the number of requests that would have been necessary to receive the states of all fire occurrences in our dataset. Our solution was to use the limited 50.000 API

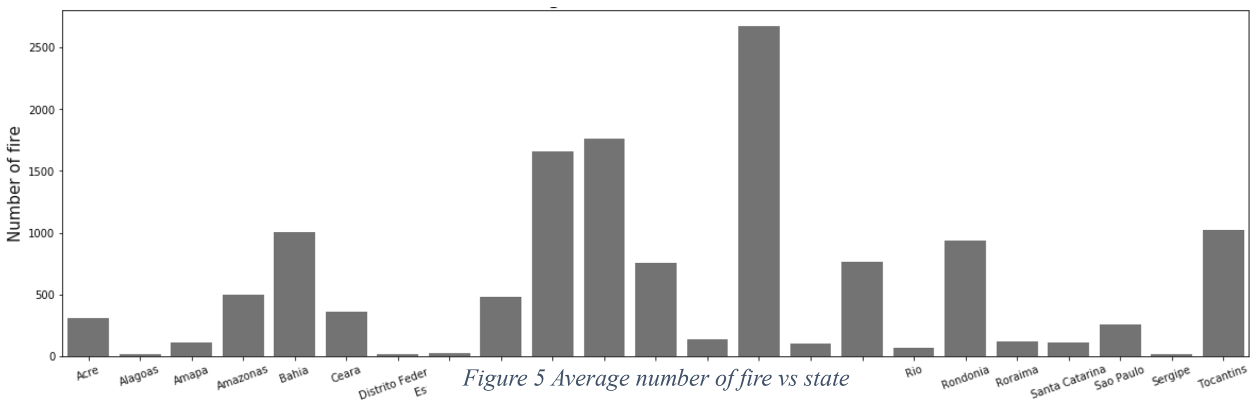


Figure 5 Average number of fire vs state

requests we had and using those as training set to predict the remainder of the 7 million fire-locations.

After converting the degree values to radians [Fig. 6], we achieved an 85% accuracy using the k-nearest-Neighbour algorithm with the haversine metric and a depth of 5. The main part of the errors consisted of zones with little fire occurrences [Fig. 5] and around borders between states.

$$\text{angle in radians} = \text{angle in degrees} \cdot \frac{\pi}{180^\circ}$$

Figure 6 Conversion between degrees and radians

3 Training

3.1 Data Modification

3.1.1 Applying Scale

As pre-processing is done, the dataset is almost fit for training. As mentioned in introduction, the target value is set as scale numbers from 1 to 4. The scale numbers are applied with quantile values of 25%, 75%, and 85% of the fire occurrence for each month and state. For example, to set the scale for November of state 'Amazonas', we gathered all the fire occurrence data in November at Amazonas and calculated the scale boundaries. Same process can be done for the whole dataset and gives boundaries as Table 1.

Month	25%	75%	85%
Jan	0.	3.	6.
Feb	0.	2.	4.
Mar	0.	1.	2.
Apr	0.	1.	1.
May	0.	1.	1.
Jun	0.	3.75	5.65
Jul	3.	29.	50.
Aug	43.	216.	284.6
Sep	39.	177.	264.6
Oct	9.	77.	121.8
Nov	1.	25.	43.3
Dec	0.	7.	12.

Table 1 Scale Boundary for Amazonas

After grading the whole dataset with scale numbers, the dataset was divided into input values and target values. Other 6 columns; Maximum/Minimum Temperature, Relative Wind Speed, Humidity, State Code, and Month were set as features for training. Scale values were applied with one-hot-encoding to make loss calculation capable.

	Date	State	Fires	MaxTemp	MinTemp	RelHum	WindVel	Scale	Month
	2000-11-01	BA	21.0	28.670370	20.216000	78.875000	2.615385	2	11

Figure 7 Feature Selection

3.1.2 Data Split

Data is split for cross-validation; year 2001~2018 for training and validation, and 2019 for testing.

3.2 Model Selection

3.2.1 Model Preperation

Several classifier models were tested on hand such as SVC Classifier, K-Nearest-Neighbors Classifier, Decision Tree Classifier, Gaussian-Naïve-Bayes Classifier, Random-Forest Classifier, and lastly a fully connected Deep Neural Network. For preparation, we tested them briefly on the dataset to see what problems occur. Two models were discarded from our selection; SVC was shown that it took to much time on calculation, and GaussianNB shown problems with loss calculations since it requires single output from training, while our dataset's target value has 4. Thus, 4 Classifiers. K-Nearest-Neighbors Classifier, Decision Tree Classifier, Random-Forest Classifier, and a Fully-Connected-Deep-Neural-Network (DNN) was executed for the project.

3.2.2 Dealing With Overfitting

Several methods were imported to prevent overfitting. For all the classifier models, KFold and shuffling were applied. This prohibits the model to overfit to a certain proportion of the data. Furthermore, for DNN,

we additionally applied batch normalization layers and dropout layers to prevent overfitting.

The DNN model is consisted as the following picture.

Model: "sequential"

Layer (type)	Output Shape	Param #
dense (Dense)	(None, 128)	896
batch_normalization (BatchNo	(None, 128)	512
activation (Activation)	(None, 128)	0
dense_1 (Dense)	(None, 256)	33024
dropout (Dropout)	(None, 256)	0
batch_normalization_1 (Batch	(None, 256)	1024
activation_1 (Activation)	(None, 256)	0
dense_2 (Dense)	(None, 512)	131584
dropout_1 (Dropout)	(None, 512)	0
batch_normalization_2 (Batch	(None, 512)	2048
activation_2 (Activation)	(None, 512)	0
dense_3 (Dense)	(None, 128)	65664
dense_4 (Dense)	(None, 4)	516

Total params: 235,268
Trainable params: 233,476
Non-trainable params: 1,792

Figure 8 DNN Model Structure

4 Evaluation

4.1 Train, Test Results

Model training and validation is executed for all classifiers, and with the weight saved from them, they are tested on the test data, year 2019. The result is as the table.

Model Type	Training Accuracy	Test Accuracy
K-Neighbors	44.9%	43.4%
Decision Tree	46.5%	45.0%
Random Forest	38.7%	37.2%
DNN	**58.0%**	**57.9%**

Table 2 Train, Test Results

4.2 DNN

Since DNN shows the highest accuracy above others, it was chosen as the final model of the project.

During Training, we executed 150 epochs total; 30 for each 5 KFold data sets and could come up with the learning curve below.

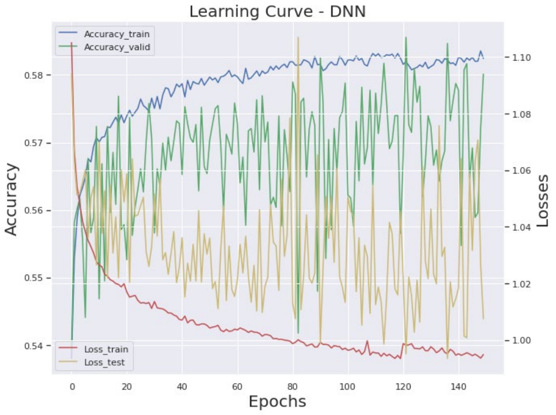


Figure 9 DNN Learning Curve

The learning curve shows that the accuracy is increasing gradually by epochs but seems to be stuck near 58~59% accuracy.

Accuracy can be plotted by state and month as the following plot. It shows that in certain states, our model fits well through the year, while some states fit only in few months. Some states have low accuracy throughout the whole year.

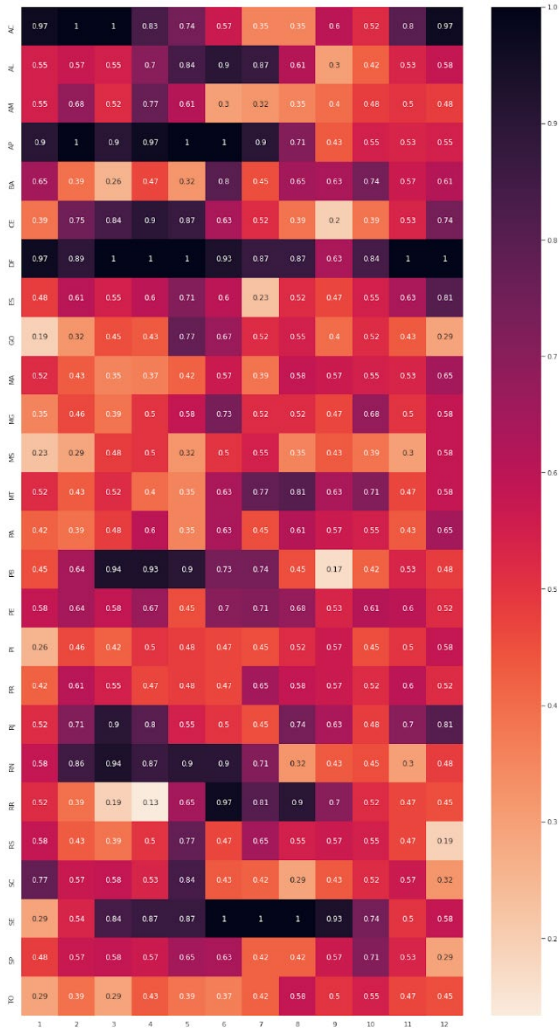


Figure 10 Accuracy by Month, State

5 Analysis

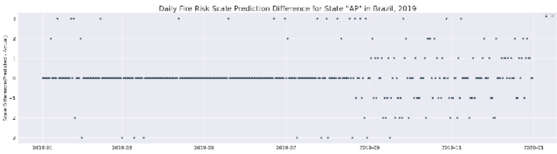


Figure 11 Daily Fire Risk Scale Prediction Difference for State "AP" in Brazil, 2019

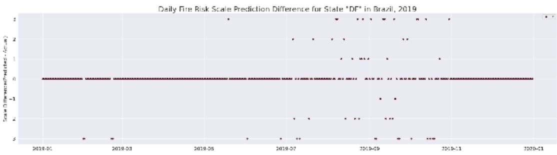


Figure 12 Daily Fire Risk Scale Prediction Difference for State "DF" in Brazil, 2019

This distribution of dots presented in figures 1 and 2 describes the accuracy of 4-classes of fire scale predictions. The middle lines in the figures represents the number of correct class prediction. Lines of dots above and below the middle line mean how much the accuracy is off-by from the correct class. The first line above the middle line would mean that the class was off by 1 step. For example, the original dot would be in class 1, but the predicted class was class 2 instead. The second line under the middle line would mean that the accuracy is off by 2 steps. For instance, it would be the case where the dot in class 1 should be in class 3. As seen in the graphs, it is visible that majority of class predictions were correct. Additionally, most of the errors were distributed in the first lines above and below the central line. This denotes that our model predicted to the point where it nearly matches the proper class prediction, however the final accuracy calculation would just consider it wrong. If we could increase tolerance of what we considered as error, the accuracy would increase by a lot. This could be facilitated by considering a minor mis-prediction by a step from the actual scale as a correct prediction. As seen in Table 3, this procedure would increase our overall accuracy to around 80%.

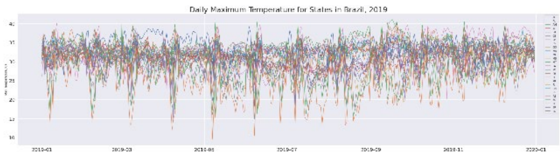


Figure 13 Daily Maximum Temperature State in Brazil, 2019

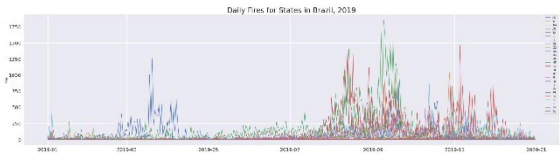


Figure 14 Daily fires for States in Brazil, 2019

We assumed that the tropical climate in Brazil was the biggest factor why overall accuracy converges around 60%. In Brazil, there is dry season (May to August) and rainy

	AC	AL	AM	AP	BA	CE	DF	ES	GO	MA	MG	MS	MT	PA	PB	PE	PI	PR	RJ	RN	RR	RS	SC	SE	SP	TO
0	220	190	143	255	143	142	311	152	122	116	146	126	146	158	184	162	114	135	167	177	172	144	158	231	138	122
-1	42	57	58	28	64	74	2	47	85	79	73	75	86	65	37	49	80	69	54	33	50	71	54	28	61	94
1	28	45	51	25	65	52	8	63	66	70	71	55	51	66	39	54	68	73	59	48	46	55	57	37	97	46
3	27	27	24	18	10	23	13	20	11	11	7	10	10	13	18	21	18	16	16	22	23	13	20	24	15	8
2	12	21	32	15	38	25	6	16	27	43	32	25	33	20	24	35	30	45	31	32	25	33	32	19	40	44
-3	19	11	23	10	15	15	17	31	21	12	3	20	10	13	31	18	26	8	13	12	24	17	19	11	4	16
-2	17	14	34	14	30	34	8	36	33	34	33	54	29	30	32	26	29	19	25	41	25	32	25	15	10	35
Dependency	**79.45%**	**85.88%**	**81.82%**	**90.32%**	**85.00%**	**84.81%**	**94.41%**	**87.92%**	**87.78%**	**83.07%**	**88.15%**	**87.97%**	**86.81%**	**89.75%**	**86.09%**	**82.55%**	**84.52%**	**81.95%**	**85.63%**	**82.69%**	**84.81%**	**85.44%**	**83.80%**	**87.32%**	**84.33%**	**83.44%**

Table 3 Dependency with higher tolerance

season (December to February) stays relatively same and there is no meaningful temperature change throughout the year [Fig. 13]. Yet, the fires accumulate around 3~4 months of the year [Fig. 14]. Each state would have its own average temperatures that are relatively similar, but fires differ dramatically. This makes features like temperature irrelevant in our report, which would originally be very indicative in other countries.

If we increased the number of parameters of the data, the accuracy could be improved. In the real world, there are a lot more to consider like terrain information, wind direction, and solar radiation when it comes to fires. However, we only had a limited number of parameters, which could have played a part in why the accuracy is not so high. In addition, man-made fires would often spread out and cause forest fires, which is an unpredictable factor. Furthermore, the errors in satellite recognition were also taken into account of low accuracy. In our report, we discarded fire data below 30% confidence.

Even though Brazil has a wet and dry season, most of the forest fires originate from a man-made error. Either legal or illegal deforestation, which is practiced in a so called 'Slash and Burn' method can often lead to the controlled deforestation fire escalating and becoming impossible to contain. It is common practice to burn unusable wood during the deforestation process. This and other fire hazards in the agricultural industry are, according to forestry experts the main source of fires in the Amazon. Even though factors like humidity and wind speed surely have an influence on the spread of fires, the source is mostly unpredictable and focuses around the times where deforestation is at its most active point during the year than around meteorological factors. [11]

6 Conclusion

In this paper, we tested various classifiers and a deep neural network to predict the forest fires in Brazil. As we progressed, we realized that Brazil was not an optimal choice of country to predict fires with due to its tropical climate. Through research, we discovered that similar experiments applied in non-tropical countries displayed much higher accuracies. [12] Despite all these considerations, we were still able to come up with relatively high accuracy of predicting the forest fires in Brazil. In conclusion, this research was not only a great practice on classifying machine learning, but also a good opportunity to try our hand at applying neural network in real world situation. In the end, we were satisfied with our results considering the limitations in our data.

References

[1] Spring, J. (2020, September 9). *Brazil's Amazon fires worsen in September, threaten virgin forests*. Reuters. https://www.reuters.com/article/us-brazil-environment-fires/brazils-amazon-fires-worsen-in-september-threaten-virgin-forests-idUSKBN2601NP

[2] Spring, J. (2020, November 2). *Fires in Brazil's Amazon rainforest jump in October*. Reuters. https://www.reuters.com/article/brazil-environment-idUSKBN27H1MT

[3] Butler, R, A. (2018, April 24). *Calculating Deforestation Figures for the Amazon*. Mongabay. https://rainforests.mongabay.com/amazon/deforestation_calculations.html

[4] Landau, L. & Phillips, T. (2020, September 2). *Amazon tragedy repeats itself as Brazil rainforest goes up in smoke*. Guardians. https://www.theguardian.com/world/2020/sep/02/amazon-fires-brazil-rainforest-bolsonaro-destruction

[5] Shukman, D. (2020, November 30) *Brazil's Amazon: Deforestation 'surges to 12-year high'*, BBC. https://www.bbc.com/news/world-latin-america-55130304

[6] Cunha, A, P. & Cunningham, C. & Brito, S. & Marengo, J. & Coutinho, M. (2017, November). *Climate change and drought in Brazil*. https://www.researchgate.net/publication/321184207_Climate_change_and_drought_in_Brazil

[7] Saraiva, M. (2020, January 15). *Brazil Weather, Conventional Stations (1961-2019)*. Kaggle. https://www.kaggle.com/saraivaufc/conventional-weather-stations-brazil

[8] National Oceanic And Atmospheric Administration. *Fire Weather Criteria*. National Weather Service. https://www.weather.gov/gjt/firewxcriteria

[9] Frazier F. & Maccherone, B. (2020). *Moderate Resolution Imaging Spectroradiometer*. NASA. https://modis.gsfc.nasa.gov/

[10] Giglio, L. & Schroeder, W. & Hall, J, V. & Justice, C, O. (2018, December). *MODIS Collection 6 Active Fire Product User's Guide Revision B*. NASA. http://modis-fire.umd.edu/files/MODIS_C6_Fire_User_Guide_B.pdf

[11] Stewart, M. (2019, November 28). *An Analysis of Amazonian Forest Fires* . Towards Data Science https://towardsdatascience.com/an-analysis-of-amazonian-forest-fires-8facca63ba69

[12] Zhao, Jianhui & Zhong, Zhang & Han, Shizhong & Qu, Chengzhang & Yuan, Zhi-Yong & Zhang, Dengyi. (2011). *SVM based forest fire detection using static and dynamic features.* Comput. Sci. Inf. Syst.. 8. 821-841. 10.2298/CSIS101012030Z.

List of figures

[1] Butler, R, A. (2020, February 09). *Deforestation in Brazil continues torrid pace into 2020*. Mongabay.
https://news.mongabay.com/2020/02/deforestation-in-brazil-continues-torrid-pace-into-2020/

[2] Saraiva, M. (2020, January 15). *Brazil Weather, Conventional Stations (1961-2019)*. Kaggle.
https://www.kaggle.com/saraivaufc/conventional-weather-stations-brazil

[3] Wolfe, R. E. (2020, December 12). *Fire Information for Resource Management System.* NASA.
https://firms.modaps.eosdis.nasa.gov/map/

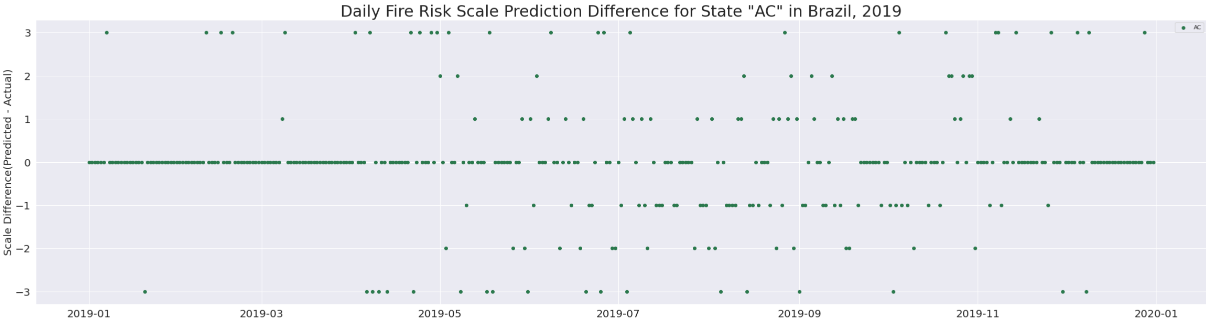
[4] Giglio, L. & Schroeder, W. & Hall, J, V. & Justice, C, O. (2018, December). *MODIS. Collection 6 Active Fire Product User's Guide Revision B*. NASA.
http://modis-fire.umd.edu/files/MODIS_C6_Fire_User_Guide_B.pdf

[5] Jhalani H. (2020, February 20). *Analysis of Forest fires in Brazil*. Analytics Vidhya.
https://medium.com/analytics-vidhya/analysis-of-forest-fires-in-brazil-dec001ddf718

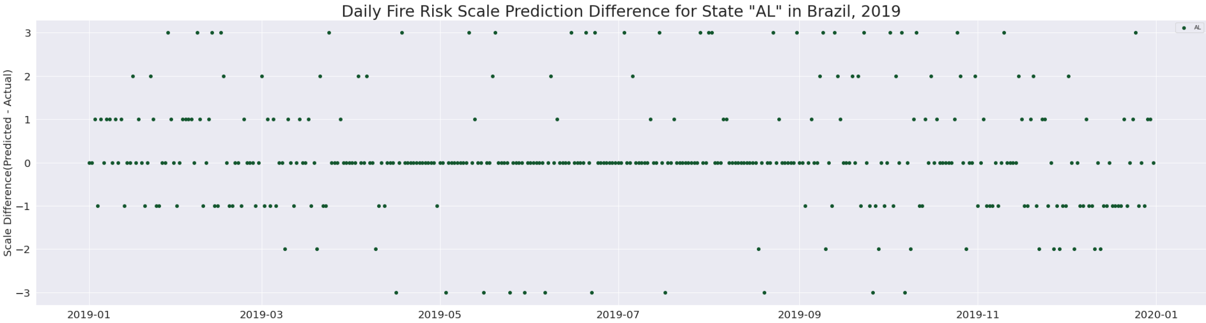
[6] Wikipedia: The free encyclopedia. (2020, December 9). *Radian* Wikipedia.
https://en.wikipedia.org/wiki/Radian

Appendix: Scale differences

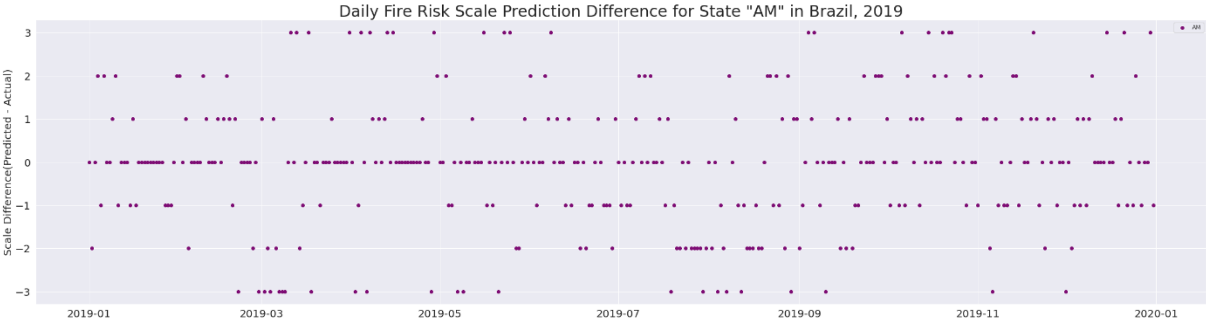
Daily Fire Risk Scale Prediction Difference for State "AC" in Brazil, 2019



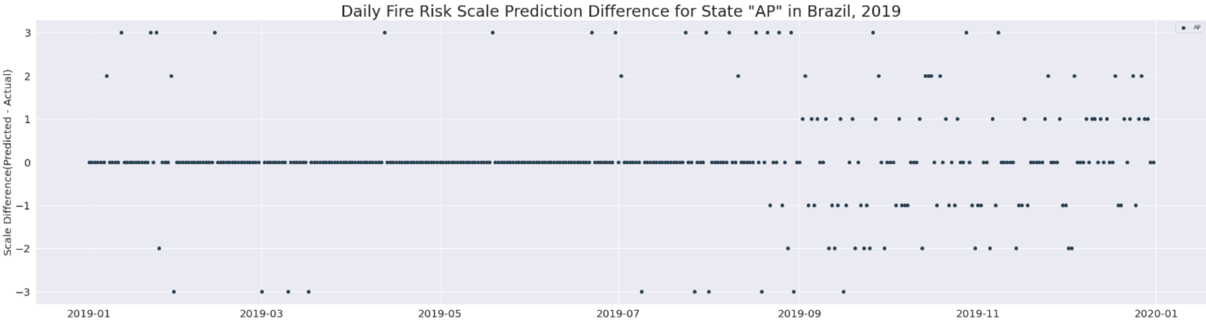
Daily Fire Risk Scale Prediction Difference for State "AL" in Brazil, 2019



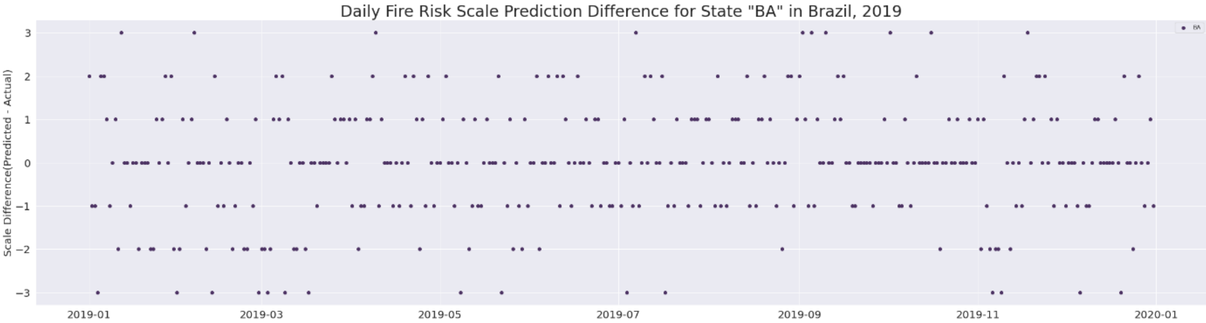
Daily Fire Risk Scale Prediction Difference for State "AM" in Brazil, 2019



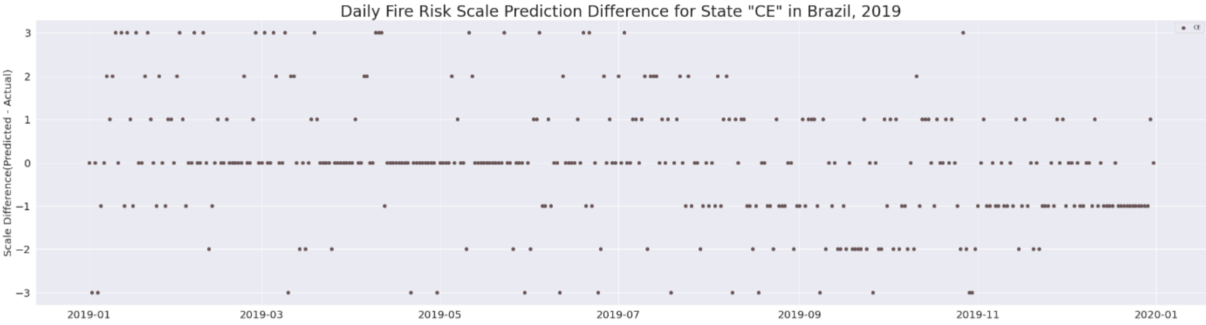
Daily Fire Risk Scale Prediction Difference for State "AP" in Brazil, 2019



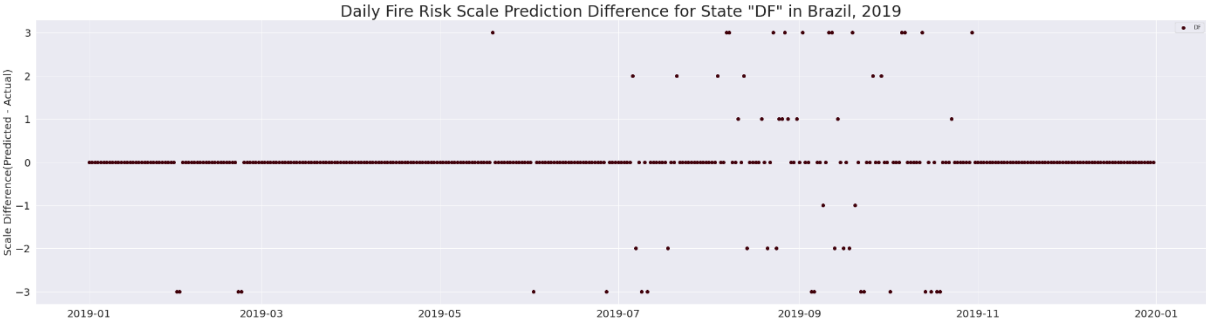
Daily Fire Risk Scale Prediction Difference for State "BA" in Brazil, 2019



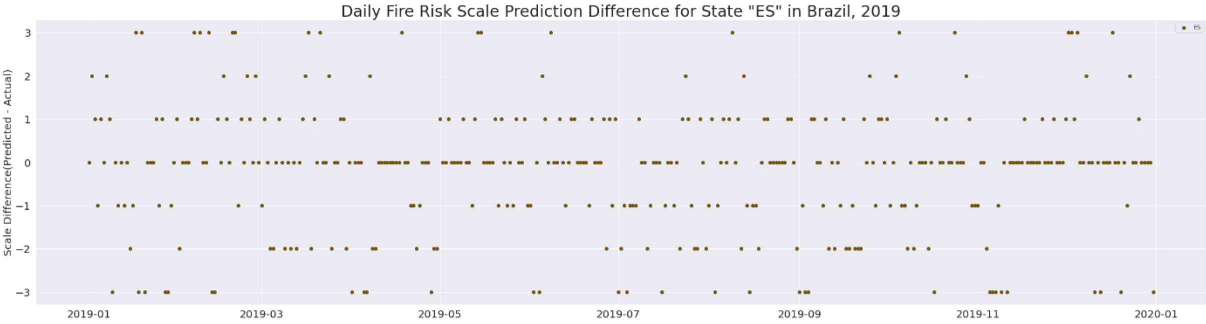
Daily Fire Risk Scale Prediction Difference for State "CE" in Brazil, 2019



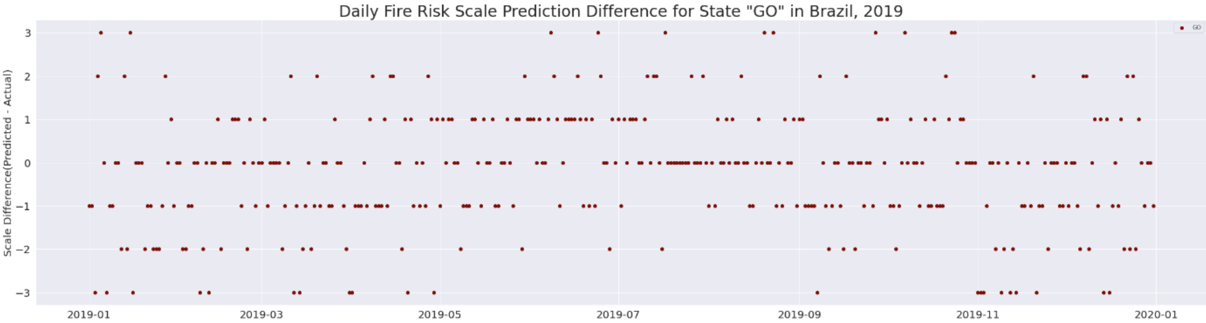
Daily Fire Risk Scale Prediction Difference for State "DF" in Brazil, 2019



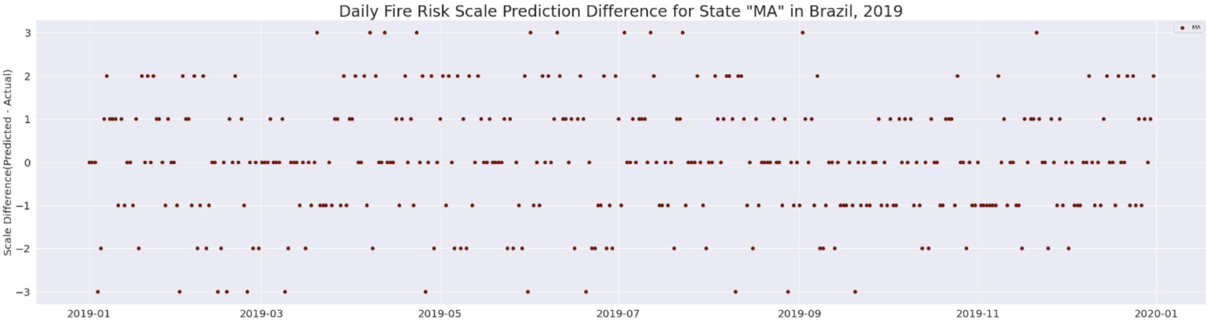
Daily Fire Risk Scale Prediction Difference for State "ES" in Brazil, 2019



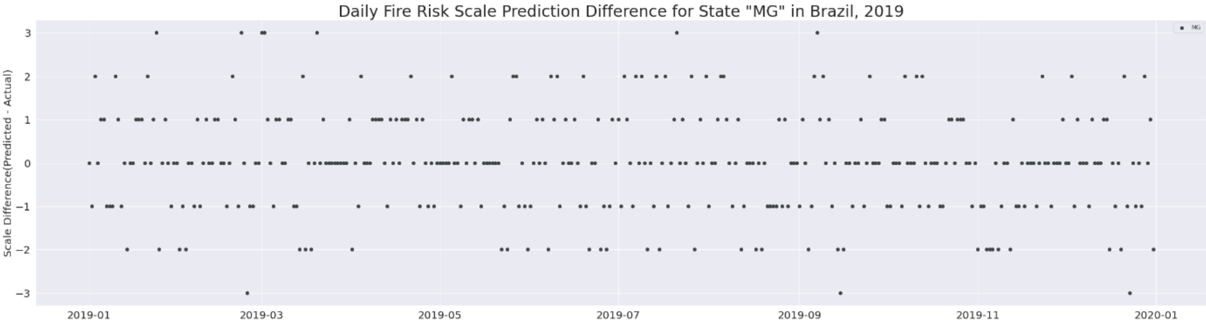
Daily Fire Risk Scale Prediction Difference for State "GO" in Brazil, 2019



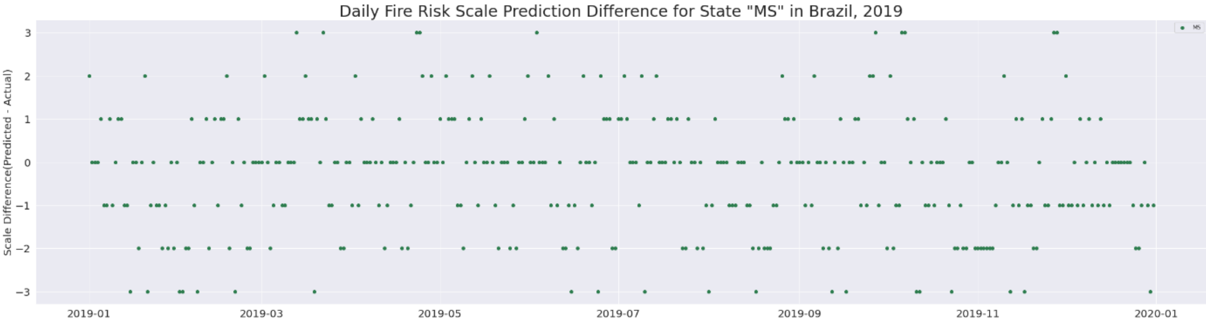
Daily Fire Risk Scale Prediction Difference for State "MA" in Brazil, 2019



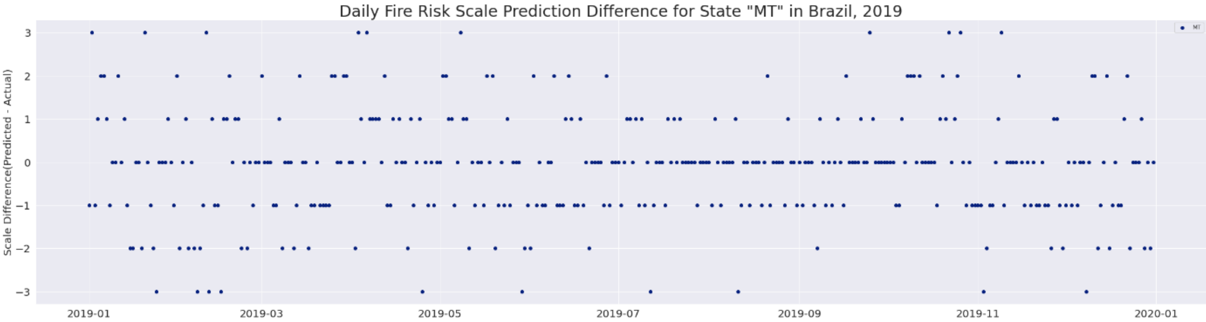
Daily Fire Risk Scale Prediction Difference for State "MG" in Brazil, 2019



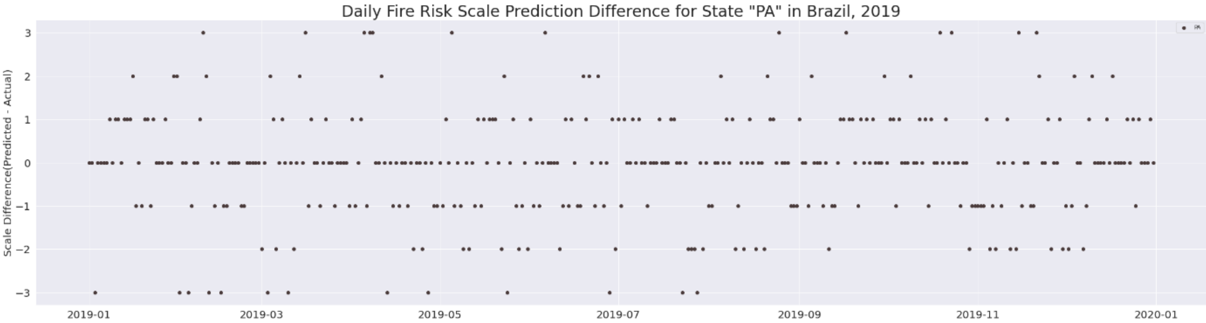
Daily Fire Risk Scale Prediction Difference for State "MS" in Brazil, 2019



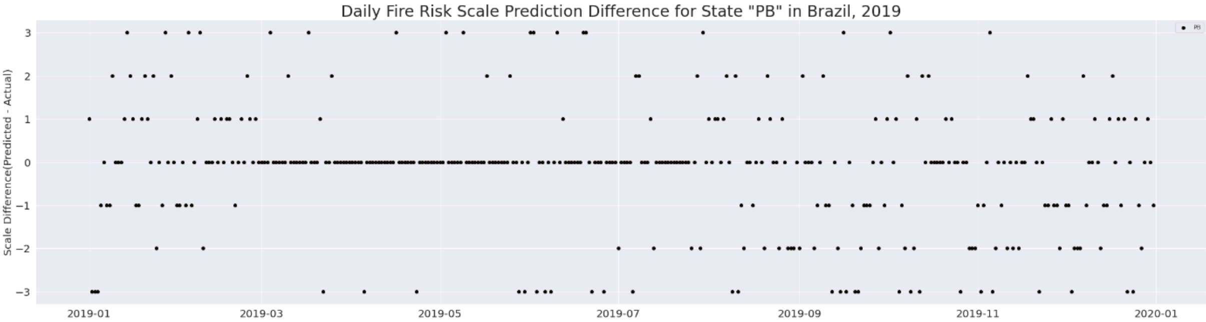
Daily Fire Risk Scale Prediction Difference for State "MT" in Brazil, 2019

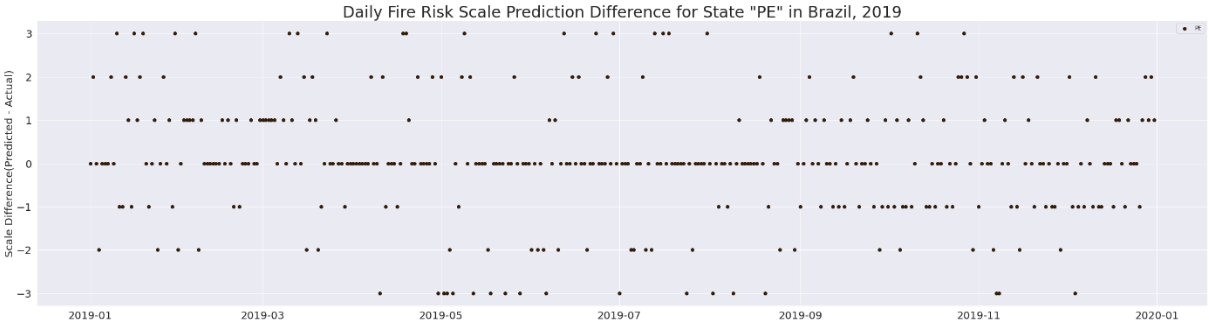


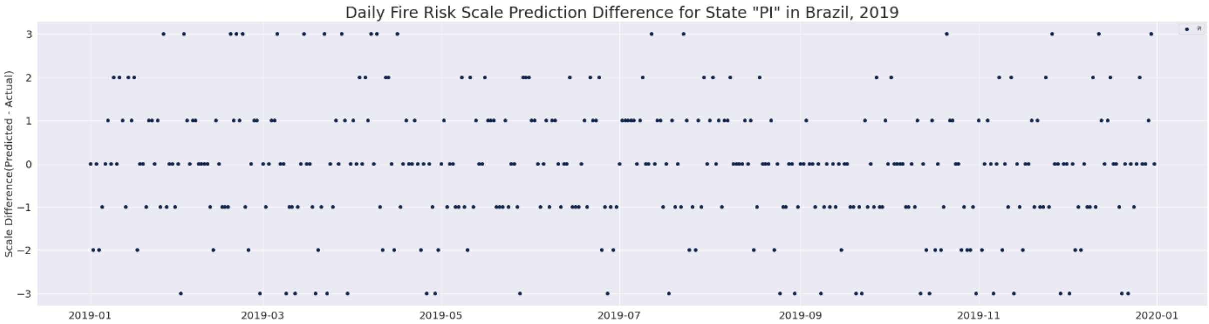
Daily Fire Risk Scale Prediction Difference for State "PA" in Brazil, 2019

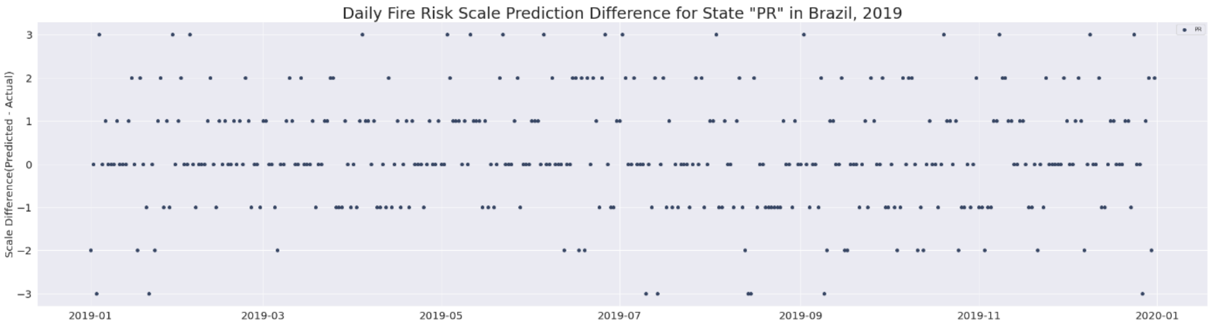


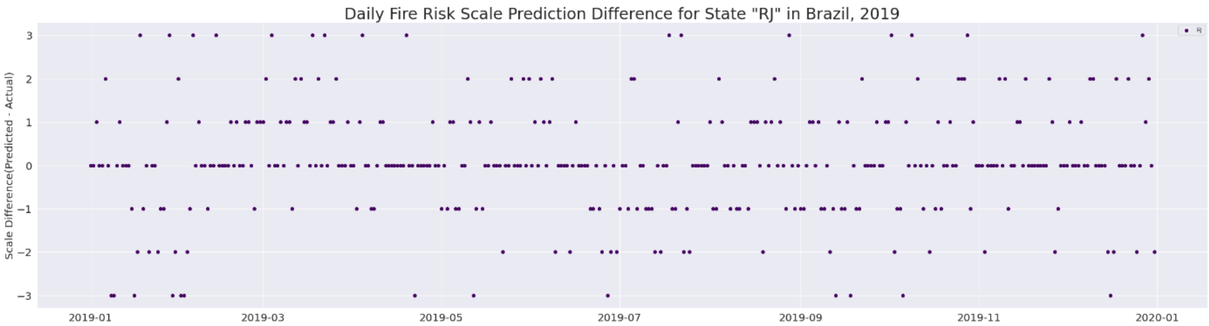
Daily Fire Risk Scale Prediction Difference for State "PB" in Brazil, 2019

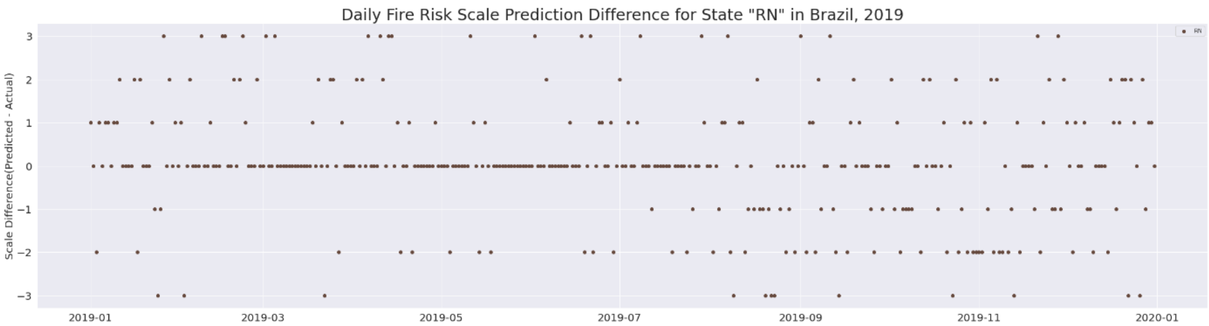



Daily Fire Risk Scale Prediction Difference for State "PE" in Brazil, 2019
Scale Difference(Predicted - Actual)
3
2
1
0
−1
−2
−3
2019-01
2019-03
2019-05
2019-07
2019-09
2019-11
2020-01

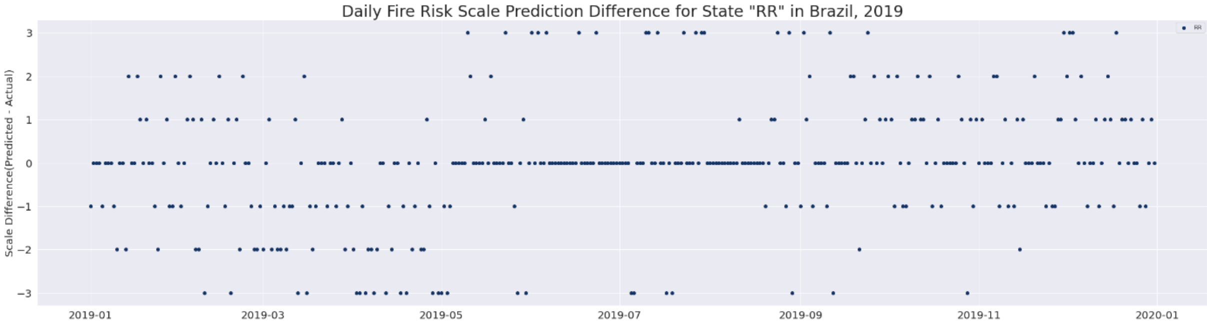

Daily Fire Risk Scale Prediction Difference for State "PI" in Brazil, 2019
Scale Difference(Predicted - Actual)
3
2
1
0
−1
−2
−3
2019-01
2019-03
2019-05
2019-07
2019-09
2019-11
2020-01


Daily Fire Risk Scale Prediction Difference for State "PR" in Brazil, 2019
Scale Difference(Predicted - Actual)
3
2
1
0
−1
−2
−3
2019-01
2019-03
2019-05
2019-07
2019-09
2019-11
2020-01

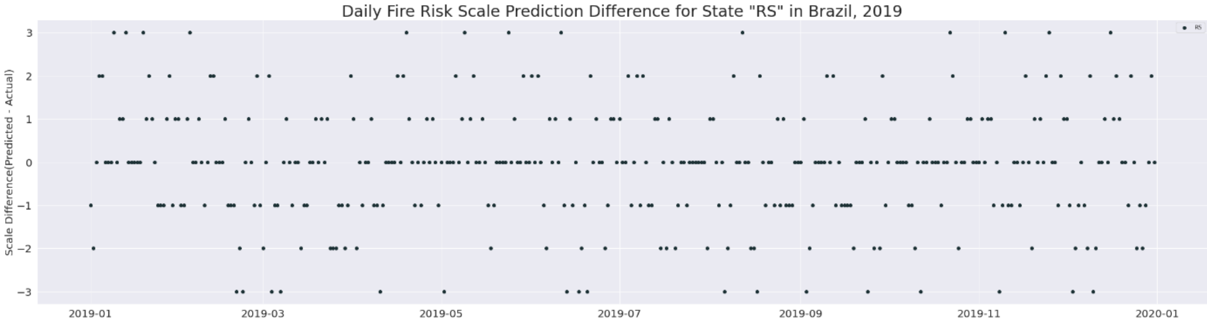

Daily Fire Risk Scale Prediction Difference for State "RJ" in Brazil, 2019
Scale Difference(Predicted - Actual)
3
2
1
0
−1
−2
−3
2019-01
2019-03
2019-05
2019-07
2019-09
2019-11
2020-01


Daily Fire Risk Scale Prediction Difference for State "RN" in Brazil, 2019
Scale Difference(Predicted - Actual)
3
2
1
0
−1
−2
−3
2019-01
2019-03
2019-05
2019-07
2019-09
2019-11
2020-01

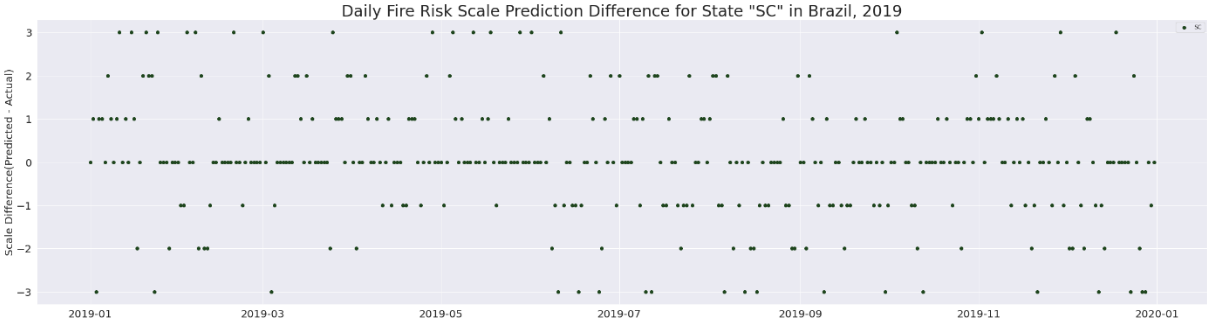
Daily Fire Risk Scale Prediction Difference for State "RR" in Brazil, 2019



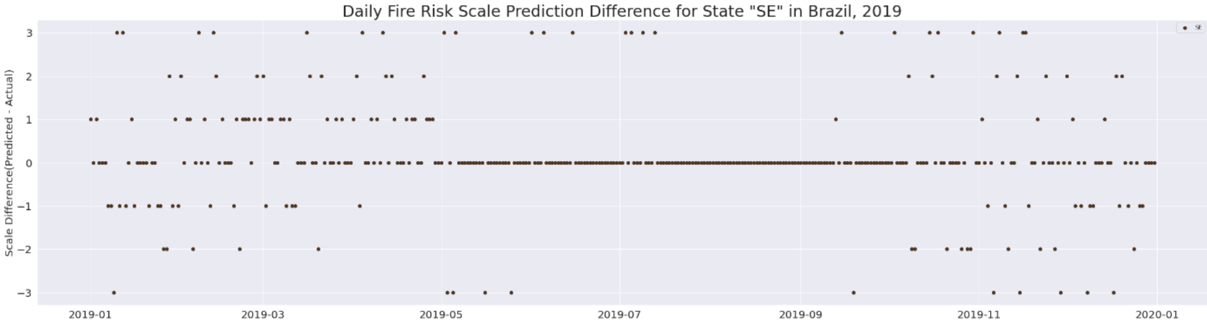
Daily Fire Risk Scale Prediction Difference for State "RS" in Brazil, 2019

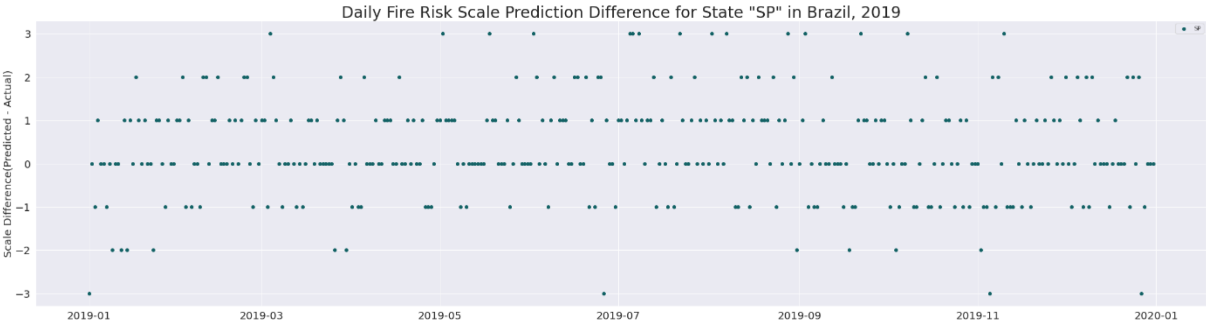


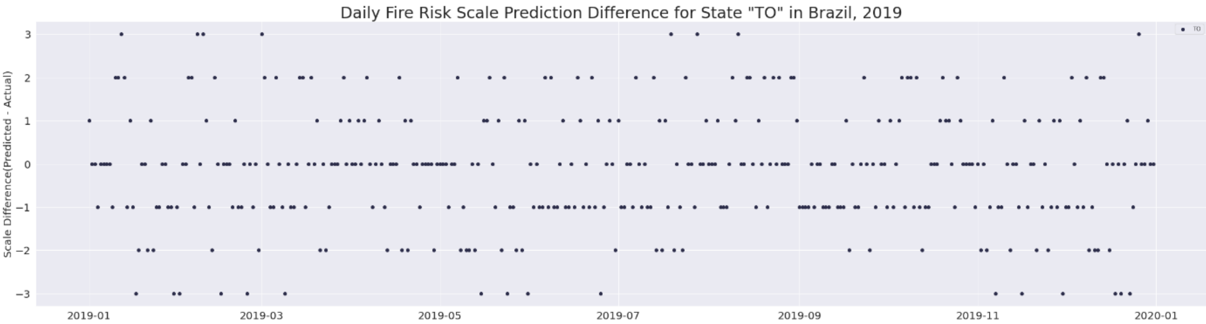
Daily Fire Risk Scale Prediction Difference for State "SC" in Brazil, 2019



Daily Fire Risk Scale Prediction Difference for State "SE" in Brazil, 2019




Daily Fire Risk Scale Prediction Difference for State "SP" in Brazil, 2019
Scale Difference(Predicted - Actual)
3
2
1
0
−1
−2
−3
2019-01
2019-03
2019-05
2019-07
2019-09
2019-11
2020-01


Daily Fire Risk Scale Prediction Difference for State "TO" in Brazil, 2019
Scale Difference(Predicted - Actual)
3
2
1
0
−1
−2
−3
2019-01
2019-03
2019-05
2019-07
2019-09
2019-11
2020-01