Predicting The Burned Area Of Forest Fires

Konrad Handke 210451506

1 Introduction

With millions of hectares of forests burning yearly, forest fires have become a significant global challenge that needs to be addressed. Forest fires have various negative impacts, including daily life disruption, ecosystem destruction, and global climate change contributions. For example, the 2024/25 California forest fires have had a massive impact on daily life and the environment. At the time of writing, nearly 40 thousand acres of land have been burned resulting in mass property damage, ecosystem damage, and mass evacuation [8]. Over 20 years, on average an increase of 5.4% of area burned by forest fires is seen yearly. In 2001, approximately 2.52 million hectares were burned while in 2023, this amount nearly quintupled to approximately 11.91 million hectares. These fires produce a large amount of pollutants that have a direct impact on global warming. As temperatures rise, forest fires become more frequent resulting in more pollutants, resulting in a self-sustaining loop where global temperatures increase and then fires occur [7]. Pollutants also affect air quality. Prolonged exposure to poor air quality will result in respiratory and cardiovascular issues [11].

With the problem introduced, better methods of preventing and understanding forest fires must be established. By analysing a large span of variables, this report will look at a diverse range of models that will be used to predict the total burned area in forest fires. This prediction should make it possible to provide an idea of the resource allocation necessary to tackle the fire and highlight the features that have the highest impact on the total burned area.

2 Dataset Analysis

The dataset [Fires prediction: Forest fires dataset] used in this report consists of 517 samples with 12 features and 1 target. The location features comprise spatial coordinates within the Monetesinho Natural Park in Portugal. The park spans 75 thousand hectares of land varying from around 400 meters to 1500 meters in altitude. It is a biodiverse location with many different types of vegetation [4]. In Figure 1, position (8, 8) has the largest mean area burned at 186 hectares. In comparison, positions (2, 3), (3, 6), and (5, 5) have the smallest mean area burned at 0 hectares. It can be observed that most high burn areas are around (6-9, 3-8), with some at (1, 2-5).

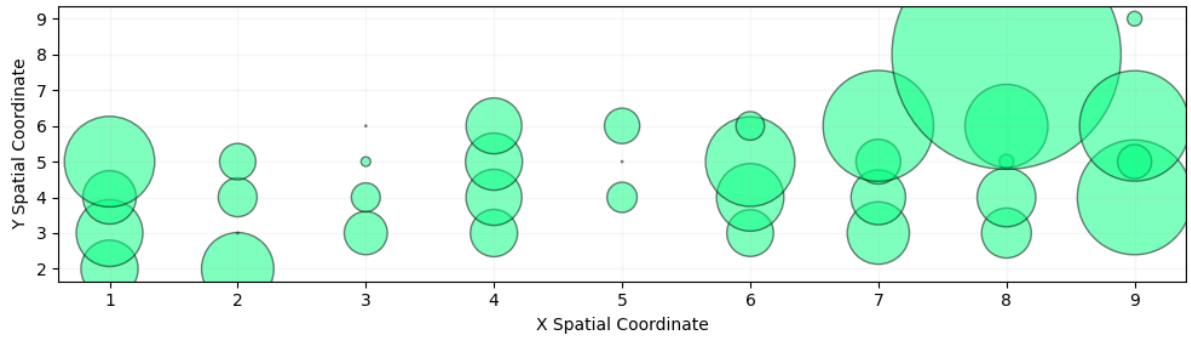

Figure 1: Spatial coordinates within the Montesinho Natural Park showcasing mean area burned

The temporal features consist of months of the year and days of the week. In Figure 2, it is observed that the months are concentrated around June to August, the summer months, during which the expected temperature should be the highest. This results in trees becoming drier and therefore easier to ignite and burn. An unexpected distribution is seen in the days of the week. It is known that weekends provide the most visitors to amenities, however, this should not have such an effect on fire occurrences. There

should not be an environmental factor that increases the likelihood of fires occurring during weekends, indicating an anomaly.

The index/code features contain parts of the Canadian Forest Fire Weather Index system. The Fine Fuel Moisture Code (FFMC), is a numerical value ranging from 0 to 101 representing the ease of ignition and flammability. The dataset contains an extremely negatively skewed set of values, mainly ranging from 80 to 96.2. The Duff Moisture Code (DMC), is a numerical value between 0 to 1001 that represents the availability of fuel in duff layers. As seen in Figure 2, there is a wide distribution of values, with a high density in the 80 to 150 range. The Drought Code (DC) is a numerical value ranging from 0 to 1000, it represents the average moisture content of deep compact organic layers. In the data distribution, there is an interesting spread, with most values hovering around 600 to 800, with some values around 0 to 130. This indicates either very moist or dry conditions. The Initial Spread Index (ISI) is a numeric value ranging from 0 to 50, representing the expected rate of fire spread. The data is positively skewed, mainly from 0 to 20, indicating a low spread rate [9][6].

The weather features consist of temperature, relative humidity, wind speed and rain. In Figure 2, the temperature is shown to have a slightly skewed normal distribution, with most values at around 20°C. Relative humidity reveals how much water vapour is in the air, compared to how much there could be at a given temperature. Lower values represent drier air which means a higher fire risk [10]. It is positively skewed, with most values around 25% to 45%. The wind speed is generally scattered appearing at particular intervals, mainly from 2 to 6 KM/H. Wind speed affects the spread of fire rather than the generation of a fire. The rain values are extremely skewed with most values appearing as 0 MM/M^2.

Lastly, the area is extremely positively skewed, with most values at 0 hectares. This suggests either the fire was contained quickly or there was no fire. As there is a large range here, in the following methodology section, the area will be logarithmically transformed to reduce the effect of outliers and to lower the variability.

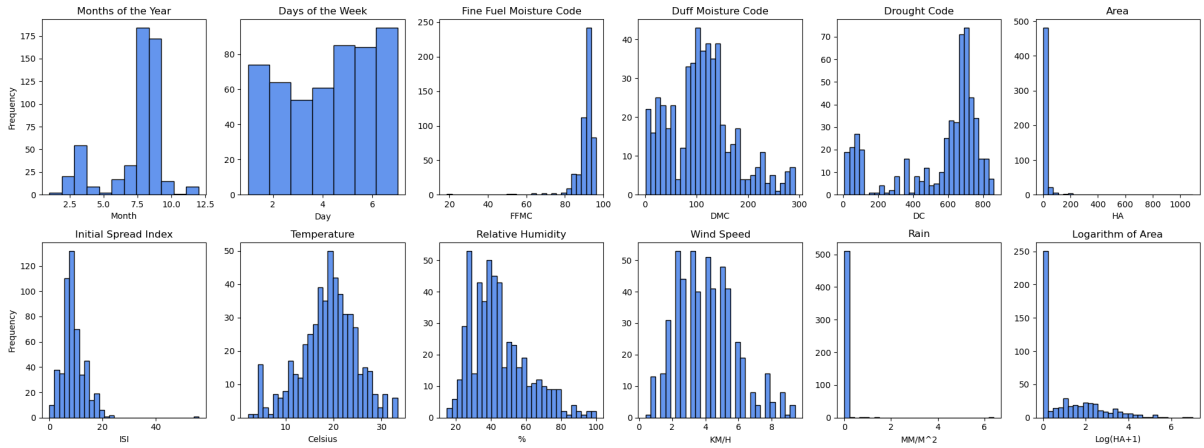

Figure 2: Distributions of features

In Figure 3, the temperature has the highest correlation with the area at 0.1, indicating a very weak positive relationship with area. The smallest negative correlation is with humidity, at -0.08. All values are between these points, suggesting very weak correlations across the board. This conveys that the area is not reliant on a single variable, but rather a combination of multiple variables.

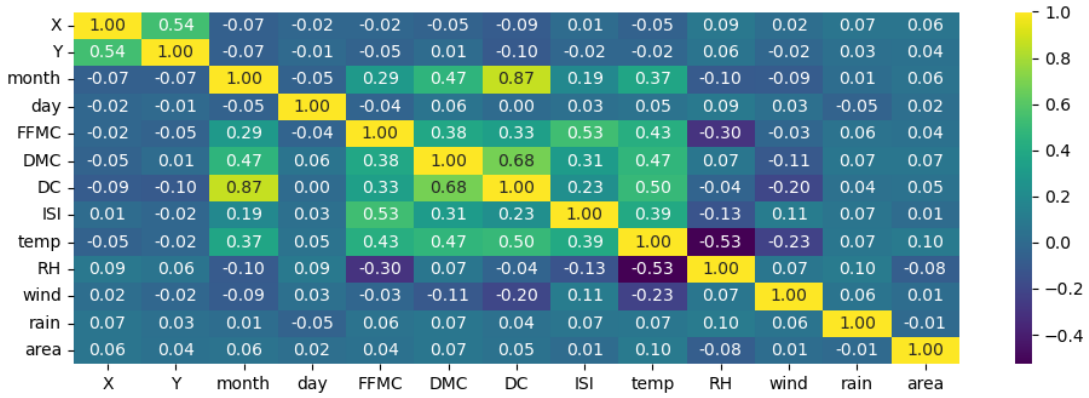

Figure 3: Correlation heatmap of variables

3 Methodology

3.1 Datasets

For each dataset, all columns were set to a standard normal distribution $\mathcal{N}(0,1)$ through $z = \frac{x-\mu}{\sigma}$.

3.1.1 Dataset 1

The first iteration of the dataset involved using the entire dataset with three conversions. The columns, months and days, were transformed to their counterpart numerical values. Months were assigned to digits 1-12 while days were assigned to digits 1-7. As seen previously, the area output is extremely skewed towards 0 values so $\log(1 + area)$ was used to reduce the total spread.

3.1.2 Dataset 2

The second iteration of the data involved 12 variables. In this version, the month column used a cyclical transformation based on temperature. This was achieved using $\sin(x)$, where $x \in \mathbb{R}$ and is divided into 12 points in the interval $[0, \pi]$. The day column was excluded as no meaningful cyclical transformation could be made. The columns: FFMC, DC, ISI, rain, and area were all logarithmically transformed to reduce variability.

3.1.3 Dataset 3

The third iteration of the dataset included only four variables. The Canadian Forest Fire Weather Index system is separated into three distinct categories: fire weather observations, fuel moisture codes, and fire behaviour indices. Using the last category, the Buildup Index (BUI) and Fire Weather Index (FWI) can be computed. The BUI represents the total amount of fuel available for the fire using a combination of the DMC and the DC. and can be derived by:

$$BUI = \begin{cases} 0.8 \cdot \frac{DMC \cdot DC}{DMC + 0.4 \cdot DC}, & \text{if } DMC \leq 0.4 \cdot DC. \\ DMC - (1 - \frac{0.8 \cdot DC}{DMC + 0.4 \cdot DC}) \cdot (0.92 + (0.0114 \cdot DMC)^{1.7}), & \text{Otherwise.} \end{cases} \tag{1}$$

The FWI is the measure of fire intensity combining both the ISI and the BUI. The FWI that will be used, is the S-scale, which is considered optimal. To compute the S-scale, the BUI has to be converted into the $f(D)$ where:

$$f(D) = \begin{cases} 0.626 \cdot BUI^{0.809} + 2, & \text{if } BUI \leq 80, \\ \frac{1000}{25 + 108.64 \cdot \exp(-0.023 \cdot BUI)}, & \text{Otherwise.} \end{cases} \tag{2}$$

then the B-scale can be obtained as $B = 0.1 \cdot ISI \cdot f(D)$. After obtaining this, the S-scale can be found from:

$$S = \begin{cases} \exp(2.72 \cdot (0.434 \cdot \log B)^{0.647}), & \text{if } B > 1, \\ B, & \text{Otherwise.} \end{cases} \tag{3}$$

This information was sourced from [2][1][3], where the original publication is Van Wagner (1987). The FWI obtained was used with the X coordinate and month data, where the month was changed using the cyclical temperature-based transformation. The area was logarithmically transformed as before.

3.2 K-Fold Cross-Validation

To train and test each model, a K-Fold Cross-Validation system will be used to separate the data into K parts. Once the data is split into K parts, the data will be trained on $K-1$ parts and tested on the last part. To ensure consistency, this system will be applied once to each dataset with $K = 5$. For each dataset, each model will use the same testing and training data to ensure consistency.

3.3 Modelling

3.3.1 Linear regression

The linear regression [5][p.13] model aims to create a function $f_w(x)$, denoted as:

$$f_w(x) = \langle x, w \rangle = \sum_{j=0}^{d} x_j w_j \tag{4}$$

where $f_w(x)$ is the predicted target variable, $\mathbf{X}$ is the data matrix with $\mathbf{x}_0 = 1$ to support the intercept $\mathbf{w}_0$, and $\mathbf{w}$ are the weights. To find the optimal weights the derivative of the Mean-Squared Error (MSE) is taken with respect to w and set to 0. This is done using the training fold from the K-Fold. This is written as:

$$MSE(\mathbf{w}) = \frac{1}{2s}\|\mathbf{X}\mathbf{w} - \mathbf{y}\|^2 \Rightarrow \nabla MSE(\mathbf{w}) = 0 = \frac{1}{s}\mathbf{X}^\top(\mathbf{X}\mathbf{w} - \mathbf{y}) \Rightarrow \mathbf{X}^\top\mathbf{X}\mathbf{w} = \mathbf{X}^\top\mathbf{y} \tag{5}$$

where:

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1d} \\ 1 & x_{21} & x_{22} & \cdots & x_{2d} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{s1} & x_{s2} & \cdots & x_{sd} \end{bmatrix} \mathbf{w} = \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_d \end{bmatrix} \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_s \end{bmatrix}. \tag{6}$$

Once the optimal weights are found, the testing set of the corresponding K-Fold is used in the Mean-Squared Error with the optimal weights. After the optimal weights and MSE are calculated for each fold, the mean and standard deviation will be computed. The mean weights and MSE will be computed for each other model as well.

3.3.2 Polynomial Regression

In the polynomial regression [5][p.14] model, the data inputs X are transformed to $\mathbf{\Phi}(\mathbf{X})$. This is a polynomial transformation that can be set to some degree d. In this report, each fold was set to a degree between 1 and 50, where degree 1 is the linear regression model. The model is set to:

$$f_w(x) = \langle \phi(x), w \rangle = \sum_{j=0}^{d} \phi(x)_j w_j \tag{7}$$

and the optimal weights are found through the same method:

$$MSE(\mathbf{w}) = \frac{1}{2s}\|\mathbf{\Phi}(\mathbf{X})\mathbf{w} - \mathbf{y}\|^2 \Rightarrow \nabla MSE(\mathbf{w}) = 0 = \frac{1}{s}\mathbf{\Phi}(\mathbf{X})^\top(\mathbf{\Phi}(\mathbf{X})\mathbf{w} - \mathbf{y}) \Rightarrow \mathbf{\Phi}(\mathbf{X})^\top\mathbf{\Phi}(\mathbf{X})\mathbf{w} = \mathbf{\Phi}(\mathbf{X})^\top\mathbf{y} \tag{8}$$

where:

$$\mathbf{\Phi}(\mathbf{X}) = \begin{bmatrix} 1 & \phi(x_{11}) & \phi(x_{12}) & \cdots & \phi(x_{1d}) \\ 1 & \phi(x_{21}) & \phi(x_{22}) & \cdots & \phi(x_{2d}) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \phi(x_{s1}) & \phi(x_{s2}) & \cdots & \phi(x_{sd}) \end{bmatrix} \mathbf{w} = \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_d \end{bmatrix} \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_s \end{bmatrix}. \tag{9}$$

Once the optimal weights are found for each training fold, the model is then tested using the testing fold on the MSE, written as (8). It is important to note that the higher the degree, the higher the chance of overfitting.

3.3.3 Ridge Regression

The ridge regression [5][p.22,23] model builds on the foundation of the MSE model by including the extra term $\frac{\alpha}{2}\|w\|^2$ to the existing $\frac{1}{2}\|\mathbf{\Phi}(\mathbf{X})\mathbf{w}-\mathbf{y}\|^2$. The term α is a regularisation parameter that balances out the weights and helps reduce overfitting. This model will use a polynomial base ranging from 1 to 25, written as (9). The final model is written as (4) or (7). With the new term, the optimal weights are found by setting the derivative of the optimisation problem to 0 as shown below.

$$E(\mathbf{w}) = \frac{1}{2}\|\mathbf{\Phi}(\mathbf{X})\mathbf{w}-\mathbf{y}\|^2 + \frac{\alpha}{2}\|\mathbf{w}\|^2 \Rightarrow \nabla f(\mathbf{w}) = 0 = \mathbf{\Phi}(\mathbf{X})^\top(\mathbf{\Phi}(\mathbf{X})\mathbf{w}-\mathbf{y}) + \alpha\mathbf{w} \tag{10}$$

$$\Rightarrow (\mathbf{\Phi}(\mathbf{X})^\top\mathbf{\Phi}(\mathbf{X}) + \alpha I)\mathbf{w} = \mathbf{\Phi}(\mathbf{X})^\top\mathbf{y} \tag{11}$$

Once the optimal weights are found, the MSE, written as (8) is computed. The difficulty here lies in the optimisation of α. If α is large then the weights will be small highlighting underfitting, whilst small α does the opposite. A variety of alpha will be tested to find the optimal degree with the optimal alpha and lowest MSE.

3.3.4 Gradient Descent

Gradient Descent [5][p.27,28] is a different procedure where the optimal weights are found through an iterative process that minimises a differentiable function. Each gradient descent will follow this iterative process. This is written as:

$$w^{k+1} = w^k - \tau\nabla E(w^k) \tag{12}$$

where w are the weights, E is a differentiable energy function and τ is the step-size parameter. It is important to note that the step-size parameter has to be > 0 and optimised. Optimisation involves the same procedure as α from the ridge regression and has to be done to prevent slow convergence or divergence. Once the weights are computed, the MSE, written as (5), is calculated. As K-Fold is used, the procedure is slightly altered, as there will be K versions of $\nabla E(w^k)$. This method will choose to take the mean of $\nabla E(w^k)$ rather than take the mean of the weights at the end. The starting weights will written as zeros the length of the data inputs base. After the weights converge, the MSE will be calculated. There will be two versions of this, one with linear energy function and the other with ridge energy function. The linear gradient descent will set E to (5). The ridge gradient descent will set the energy function as (10). In the ridge version, the optimal degree and alpha will be taken and substituted into the gradient.

3.3.5 Gradient Descent with Huber Loss

Gradient Descent with Huber Loss [5][p.31,32] includes the function: $H_\tau(w) = \sum_{j=0}^{d}|w_j|_\tau$. It is also multiplied by some α that needs to be optimised: $\alpha H_\tau(w)$. In this case, τ represents the smoothing parameter. This function is added to the energy function written in (12). This means that the derivative of the Huber Loss function needs to be computed, in this case, it is written as:

$$|w_j|_\tau = \begin{cases} \frac{w_i}{\tau}, & \text{if } |w_i| \le \tau, \\ 1, & \text{if } w_i > \tau, \\ -1, & \text{if } w_i < -\tau. \end{cases} \tag{13}$$

In this method, a linear base has been chosen, due to previous regression results. The Huber Loss gradient will be added to the MSE gradient written in (5). Once this is done, the method is the same as the normal gradient descent, (12).

3.3.6 Proximal Gradient Descent

Proximal Gradient Descent [5][p.27,32,33] is an optimisation method that consists of normal gradient descent and an extra feature. The original energy function is written as $E(w)+R(w)$ where the function E is convex and continuously differentiable, but the function R is proper, convex and lower semi-continuous. R is a regularisation function and can be written as $\frac{\alpha}{2}\|\mathbf{w}\|^2$ from ridge regression. In this method, the addition of R is seen once (12) has been calculated, through a proximal map. This method will use the

function $R(w) = a\|w\|$, the LASSO regularisation term, which results in the proximal map:

$$|w_j| = \begin{cases} 0, & \text{if } |w_i| \leq \tau\alpha, \\ w_i - \tau\alpha, & \text{if } w_i > \tau\alpha, \\ w_i + \tau\alpha, & \text{if } w_i < -\tau\alpha. \end{cases} \tag{14}$$

where α is a regularisation parameter and τ is the smoothing parameter. Once the weights are calculated, the MSE, (5), is computed.

4 Results

Due to the K-Folds, there will not be a constant MSE for each model as it will change depending on the configuration of the data. For each model, the MSE will remain very similar but will have some deviation. It is important to know that for each gradient descent graph if the step size was greater than shown, the MSE would increase rapidly to extreme values.

4.1 Dataset 1 Results

The linear regression resulted in an MSE of $0.5294 +- 0.02$ with a standard deviation of 0.09. Consistently, the month column seems to have the strongest relationship with a weight around 0.2479 while the Y coordinate column seems to be the least relevant at -0.0019. The polynomial regression most commonly reveals the best degree is degree 1, suggesting a linear relationship captures this the best. The ridge regression revealed a similar pattern. The MSE here is $0.4997 +- 0.015$ with a 0.0239 standard deviation. The lower MSE could suggest a better model, but this is not the clear case. In this data, the degree directly matches the polynomial regression degree. If it is degree 1, then the ridge regression is degree 1. A degree 1 ridge shows an extremely large alpha, that seems to depict a pattern of $\alpha \to \infty$, weights $\to 0$. On rare occasions, the ridge regression will show a degree 2 with a low alpha. In Figure 4 the Linear Gradient Descent has the lowest MSE of 0.4813 starting from around step size 0.05. Interestingly, this is also the case for Huber Loss and proximal. What is seen is that as $\alpha \to 0$ with different step sizes, the MSE for proximal starts to converge. What is interesting here is that the τ seems to have little effect on the outcome. In Huber Loss, the τ and α are both important. As with Proximal, if α is 0 then the MSE will be the lowest, however, what is different here is the τ. With a large τ, no matter the α, the MSE will also converge to the same value. The gradient descent with ridge has a comparable MSE to the basic ridge regression, but to achieve this the step size has to be minuscule. On rare occasions, when the ridge regression optimal degree is 2, the MSE for the ridge gradient descent will be lower than all the other models. This seems to be a rare case so will not be looked at as it cannot be standardised to all data. The best model here would be the linear gradient descent model. In Figure 5 the final model is presented. The MSE for this model is 0.4811, which is what was shown in the above figure. The weight vector for this model is:

$$\mathbf{w}^\top = \begin{bmatrix} 8.39 \cdot 10^{-6} & 7.03 \cdot 10^{-2} & -6.90 \cdot 10^{-4} & 2.48 \cdot 10^{-1} & 1.28 \cdot 10^{-2} & 2.00 \cdot 10^{-2} & \\ 1.08 \cdot 10^{-1} & -2.07 \cdot 10^{-1} & -8.12 \cdot 10^{-2} & 2.95 \cdot 10^{-2} & -4.94 \cdot 10^{-2} & 7.43 \cdot 10^{-2} & 1.80 \cdot 10^{-2} \end{bmatrix} \tag{15}$$

This has been graphed using the mean weights for a step size of 0.05 which was shown optimal. In the first graph, the prediction seems to mimic the lower values around 0 quite well however does not deliver on the higher values. In the second graph, the problem remains however even the lower values are not completely correct. The third graph mimics the behaviour of the first graph and nothing new is seen. The fourth and fifth graphs both show an interesting development. There are attempts to represent the larger actual area values. The values are far from the actual, but an attempt was made.

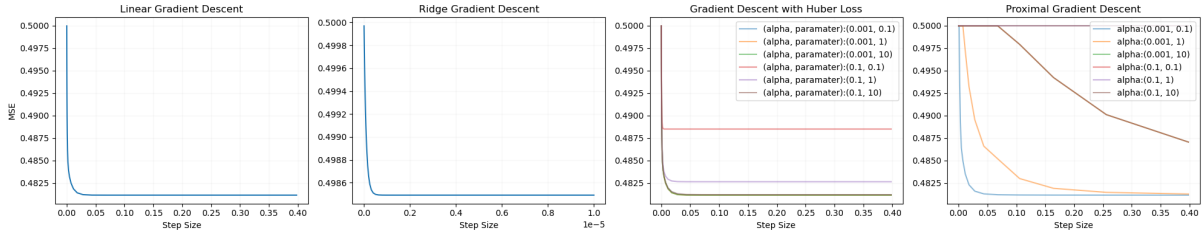

Figure 4: MSE variability in different regression models

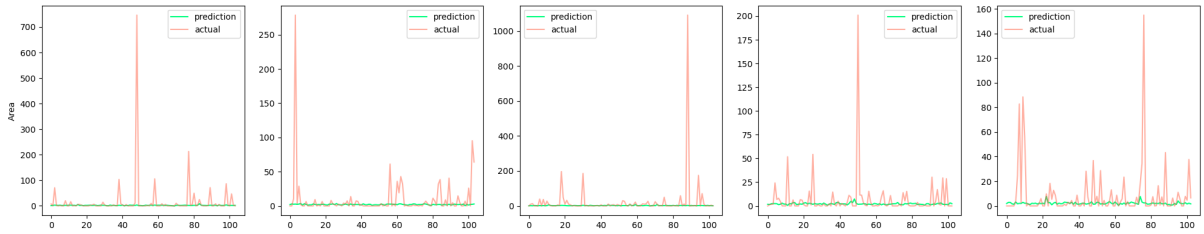

Figure 5: Final model for dataset 1

4.2 Dataset 2 Results

In the data, the linear regression revealed a $0.5087 + -0.01$ MSE with a standard deviation of 0.0926. This is very close to the previous dataset, although the features have been drastically changed. As with the first dataset, the Y coordinate is the least relevant, and although the month has been altered, it is still the most important variable. The polynomial analysis is a consistent degree 1, meaning that the MSE and weights are identical. The ridge regression follows the same pattern as the previous dataset, with the alpha being exceptionally large. The MSE here is $0.4993 + -0.005$ with a standard deviation of 0.0707. In Figure 6 the MSE is graphed according to step size and other parameters. As before, the ridge gradient descent presents a small step size with a relatively high MSE compared to the rest. On rare occasions, this ridge gradient descent can have a lower MSE than everything else even though it is still degree 1. The reason for this is unclear. The linear gradient descent, like previously has the same step sizes but with a lower MSE. The lowest MSE is around 0.479 which is the best current for this sample. As with the Huber Loss and proximal from the first dataset, the α and τ mimic the same behaviour. The best model for this dataset would be linear gradient descent. In Figure 7 the final model is presented. The MSE for this model is 0.4794, which is what was shown in the above figure. The weight vector for this model is:

$$\mathbf{w}^\top = \begin{bmatrix} -4.96 \cdot 10^{-6} & 7.47 \cdot 10^{-2} & 4.25 \cdot 10^{-4} & -2.07 \cdot 10^{-1} & 7.40 \cdot 10^{-2} & 8.82 \cdot 10^{-2} \\ 8.31 \cdot 10^{-1} & -6.53 \cdot 10^{-1} & 1.41 \cdot 10^{-2} & 1.81 \cdot 10^{-2} & 9.52 \cdot 10^{-2} & -2.55 \cdot 10^{-2} \end{bmatrix} \tag{16}$$

In the graphs, the predictions seem to be more deviated with more attempts to predict large areas. Apart from this, the graphs seem to follow the same patterns as the first dataset. This would suggest that although a variable was changed and another left out, there is little change observed. This model is better than the previous one.

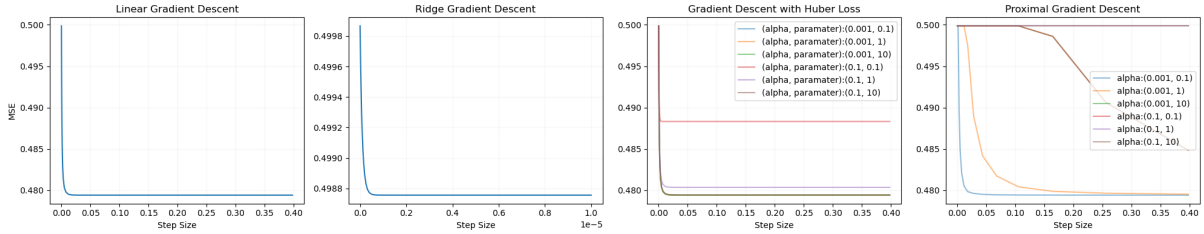

Figure 6: MSE variability in different regression models

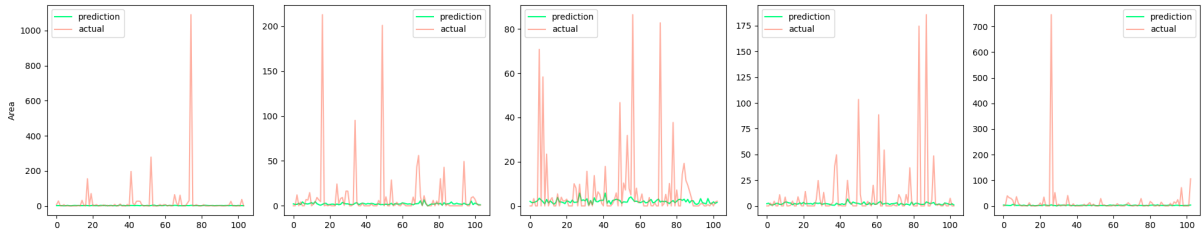

Figure 7: Final model for dataset 2

4.3 Dataset 3 Results

The linear regression here gives an MSE of $0.4984 + -0.01$ with a standard deviation of 0.0382. The lower variable count should make the modelling easier however might make the data too simple. The polynomial

regression reveals that the optimal degree is consistently degree 2, with an MSE of $0.4946 + -0.01$. The ridge regression shows that the optimal degree is degree 4, however, this changes frequently between 2 and 6. The alpha here is very large at 2685, however, this is still relatively low to some of the other results that can be achieved. This would suggest that the model is being simplified. The MSE here is $0.4921 + -0.005$ with a standard deviation of 0.0361. In Figure 8 the same pattern is present for the linear, Huber Loss, and proximal gradient descent. The lowest is around 0.4944. The most important graph here is the ridge gradient descent. This is the most interesting as it is different to all other previous results suggesting it could be the best model out of all of the models. The MSE here is around 0.4880 at a step size of $1. \cdot 10^{-6}$. The final model here will be the ridge gradient descent model. In Figure 9 the model is shown. The MSE here is 0.4881 with the final weights at:

$$\mathbf{w}^\top = \begin{bmatrix} -0.0024 & 0.0064 & -0.0004 & 0.0039 & 0.0021 & -0.0064 & \\ -0.0047 & 0.0086 & -0.0014 & 0.0145 & 0.0108 & 0.0041 & -0.0058 \end{bmatrix} \tag{17}$$

Looking at this model, there is not much to comment on as it is significantly worse than the previous two even though the MSE is small which would indicate this is better. Some notable points get the area spot on, but most of the values are incorrect.

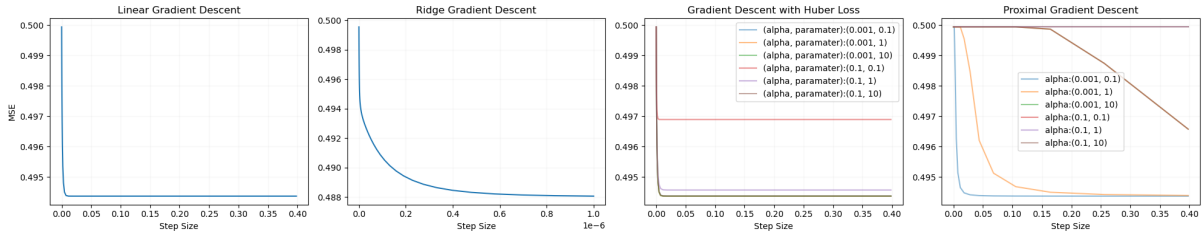

Figure 8: MSE variability in different regression models

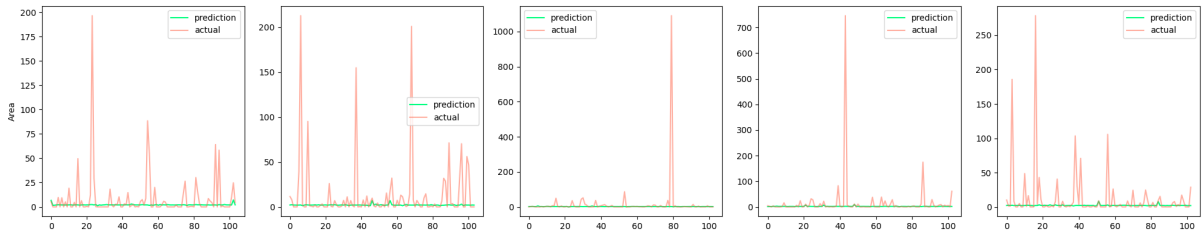

Figure 9: Final model for dataset 3

5 Conclusion

This report created multiple regression models to predict the burned area of forest fires. The final models created have significant limitations to them. The last model is the worst representation even though it achieved the lowest MSE. The two linear models are significantly better, being able to somewhat accurately predict the area. Low area values around zero can be predicted with a slight degree of inaccuracy. The main problem encountered in each attempt is the high area values. High values are unable to be predicted, and this is most probably due to the extreme skew of the area. As most values are zero, it is difficult to model which variables are more important than others. One way to address this is to remove the outlier values for the testing phase to check which features are more important. However, even when the model is tested on the outlier samples, the area would still most likely not be close. Another issue is the picking and choosing of features. Each model will have a feature that is less important than the others, and even if that feature is taken out, another feature will replace it. A potential way to combat this would be to design a couple of models that remove the feature with the lowest importance until only a few remain. To conclude, the second dataset final model is the best model however still should not be used as it is not accurate.

References

[1] Buildup index. https://wikifire.wsl.ch/tiki-index8720.html?page=Buildup+index, -.

[2] Canadian forest fire weather index system. https://wikifire.wsl.ch/tiki-indexbdbc.html?page=Canadian+forest+fire+weather+index+system, -.

[3] Fire weather index. https://wikifire.wsl.ch/tiki-index259b.html?page=Fire+weather+index, -.

[4] Montesinho natural park. https://amontesinho.pt/en/the-region/montesinho-natural-park/, 2025.

[5] Martin Benning. Machine learning: Lecture notes 2024–25, 2024. Adapted and modified by Nicola Perra.

[6] Francesca Di Giuseppe. User guide. https://confluence.ecmwf.int/display/CEMS/User+Guide, 2024.

[7] James MacCarthy; Jessica Richter; Sasha Tyukavina; Mikaela Weisse; Nancy Harris. The latest data confirms: Forest fires are getting worse. https://www.wri.org/insights/global-trends-forest-fires, 2024.

[8] Tim Stelloh; Marlene Lenthang; Rebecca Cohen; Phil Helsel. California wildfires: What we know about l.a.-area fires, what caused them, who is affected and more. https://www.nbcnews.com/news/us-news/california-wildfires-what-we-know-palisades-eaton-los-angeles-rcna188239, 2025.

[9] Government of Cananda. Canadian forest fire weather index (fwi) system. https://cwfis.cfs.nrcan.gc.ca/background/summary/fwi, -.

[10] Met Office. Humidity. https://www.metoffice.gov.uk/weather/learn-about/weather/types-of-weather/humidity, -.

[11] Shakshi. How wildfires affect air quality: Recent global case examples. https://www.aqi.in/blog/en-gb/wildfires-affect-air-quality-recent-case/, 2024.