**Master of AI**

Faculty of Engineering & IT

42172 AT1: Forest Fire Spread Prediction

UTS CRICOS 00099F

**Link to solution notebook:**

[Google Colab](https://drive.google.com/file/d/1geJoGXNZ3lavZtkXW4lxlCKxqTTORlMt/view?usp=drive_link)

[GitHub](https://github.com/Cyrus-Kwan/Forest-Fire-Prediction)

**Part: 2**

**Author**: Cyrus KWAN

**Student ID**: 25466929

**Date:** 30 SEPTEMBER

**Semester:** Spring

**Year:** 2024

**University of Technology Sydney**

Table of Contents

[Table of Contents 2](#_Toc178615334)

[Summary or Executive Summary 3](#_Toc178615335)

[Purpose 3](#_Toc178615336)

[Techniques 3](#_Toc178615337)

[Findings 3](#_Toc178615338)

[Conclusion 3](#_Toc178615339)

[Recommendation 3](#_Toc178615340)

[Introduction 3](#_Toc178615341)

[Report Body 3](#_Toc178615342)

[Methods/AI techniques used 3](#_Toc178615343)

[Implementation of AI techniques 6](#_Toc178615344)

[Comparisons of different AI techniques 12](#_Toc178615345)

[Conclusions and Recommendations 13](#_Toc178615346)

[Conclusion 13](#_Toc178615347)

[Recommendation 13](#_Toc178615348)

[List of References 14](#_Toc178615349)

[Individual reflection 15](#_Toc178615350)

[Planning 15](#_Toc178615351)

[Learnings 15](#_Toc178615352)

[Challenges 15](#_Toc178615353)

[Limitations 15](#_Toc178615354)

[Future 15](#_Toc178615355)

## Summary or Executive Summary

|  |
| --- |
| Purpose |
| Reevaluates the recommendation of previous literature by comparing data mining model solutions to predict the spread of forest fires in Montesinho park in Norther Portugal. |

|  |
| --- |
| Techniques |
| Analyses the regression task using support vector machine (SVM) and deep neural networks and compares the effect of both z-score and min-max normalization on each model. |

|  |
| --- |
| Findings |
| This report found that generally, a z-score normalized deep neural network performs best when predicting the area spread of forest fires. Neither normalization technique significantly affects the results of the SVM regressor. Although the min-max normalized neural network did not make the most accurate prediction, it still outperformed both SVM configurations. |

|  |
| --- |
| Conclusion |
| Forest fire prediction is directly related to the preparedness of first responders and thus requires the highest accuracy. The results of this paper contradict the findings in previous works that recommended a z-score SVM regressor. |

|  |
| --- |
| Recommendation |
| Authorities should generally use a z-score normalized neural network as they are more scalable and the absence of input bounds allows it to be implemented into real-time systems with high accuracy. |

## Introduction

|  |
| --- |
| What was the problem and its context? (including data sources) |
| Cortez and Morais (2007) produced a work that determined that a support vector machine (SVM) regressor is the most accurate data mining model in predicting the scale of forest fires in Montesinho park located in Northern Portugal by using meteorological data. The study compared SVM, random forests, decision trees, and a shallow neural network. |
| Why was it a problem? |
| The evaluation of only a shallow neural network in the previous literature is such the final recommendation could be argued as unsatisfactory. As the topic of natural disaster prevention bears much consequence. It is imperative that the most accurate model in these systems are chosen for the task. |
| Why was the project necessary and how was the problem solved? |
| Forest fire prediction is inherently a mission critical system that directly influences the preparedness of first responders. Hence, this report extends on the work of (Cortez & Morais, 2007) to determine if deep neural networks are able to more accurately predict the spread of forest fires. The same dataset is used as in previous literature and both z-score and min-max normalization are used on all input features to train 2 separate configurations of SVM regressor and deep neural network respectively. All 4 models are evaluated against a mean square error (MSE) and a final recommendation is made. |

## Report Body

### Methods/AI techniques used

|  |  |  |  |
| --- | --- | --- | --- |
| Regression Models | | | |
|  | What are regression models? | | |
|  | Regression models are stochastic methods of predicting a continuous target variable by learning the correlations and trends in a training set. | | |
|  | How are regression models evaluated? | | |
|  | Comparatively to classification problems, regression models are not limited to making predictions between 0 and 1. Hence, metrics such as accuracy, precision, recall, etc are not useful in evaluating the accuracy of regression models. Rather, regression models are evaluated based on the error from the predicted value to the observed value. Typical examples of regression metrics are mean square error (MSE), root mean square error (RMSE), and mean absolute error (MEA). | | |
|  | How do they relate to the problem context? | | |
|  | For this work, the selected models attempt to predict the spread of forest fires in the Montesinho park located in northern Portugal. | | |
|  |  | | |
|  | *Figure 1: Regression* | | |
| Neural Network | | | |
|  | Description of the technique | | |
|  | **Forward pass** | | |
|  | In the forward pass stage, the dot product of each input and their respective weights are parsed into each neuron in the hidden layers. An activation function is used on the aggregated dot product to form some non-linear function. This process continues for each hidden layer until a final output has been produced. | | |
|  |  | | |
|  | *Figure 2: Forward Pass* | | |
|  | **Back propagation** | | |
|  | In the back propagation stage, the weights (w) of each neuron that is composed of an activation function and a dot product is increased or decreased by a learning rate such that the result of the loss function (error) approaches the minima. This process starts from the end of the neural network and propagates backwards towards the input layer. | | |
|  | Loss function: | | Used to calculate how accurate the predicted output is compared to the true value. This work uses mean square error as the primary loss function for the configured neural network. |
|  |  |  | |
|  | Gradient Descent: | | To find the minima of the loss function, the neural network iteratively increases or decreases the weights in each neuron by the learning rate such that the gradient of the loss function approaches zero. |
|  |  | | |
|  | *Figure 3: Back propagation* | | |
|  | Why was this algorithm chosen? | | |
|  | In the work of Cortez and Morais (2007), a single hidden layer neural network was used to predict the area spread of forest fires in the Montesinho park. In contrast, this report evaluates a neural network configuration using multiple hidden layers. Additionally, neural networks are a method of parametric modelling that performs well in determining non-linear patterns in data. | | |
| Support Vector Regressor | | | |
|  | Description of the technique | | |
|  | Unlike the support vector machine (SVM) classifier, the support vector regressor produces a prediction function that simultaneously minimizes the error rate (ξ) from the decision boundary and maximizes the number of data points within the decision boundary. The resulting hyperplane function is used to make predictions for subsequent data points. | | |
|  |  | | |
|  | *Figure 4: SVM regressor* | | |
|  | Why was this algorithm chosen? | | |
|  | SVM was the recommended model in the work of Cortez and Morais (2007). Hence, it was selected to be evaluated against neural networks with multiple hidden layers. Although SVM is typically associated with classification modelling, it is still able to learn and form a non-linear prediction function. | | |

### Implementation of AI techniques

|  |  |  |
| --- | --- | --- |
| Programming Environment | | |
|  | **Variable** | **Description** |
|  | CPU | Intel(R) Core(TM) i7-4600M CPU @ 2.90GHz 2.89 GHz |
|  | RAM | 8.00 GB |
|  | Python | Version 3.12.4 |
| Programming language for model implementation. |
|  | Anaconda | Version 24.5.0 |
| Jupyter notebooks interface for local testing environment. |
|  | Google Colab | Uniform external testing environment for jupyter notebooks. |
| Libraries | | |
|  | **Package** | **Usage** |
|  | Pandas | Data handling interface for slicing and transforming forest fire data. |
|  | NumPy | Array transformations and mathematical operations. |
|  | MatplotLib | Data visualizations. |
|  | Scikit-learn | Data normalization/standardization, SVM regression implementation, and mean square error metrics evaluation. |
|  | Tensorflow | Neural network regression implementation. |
| Data Source | | |
|  | The data for this report was acquired from the work of Cortez and Morais (2007) that includes 517 rows and 13 columns of meteorological data and features from the fire weather index (FWI). | |

|  |  |
| --- | --- |
| Data Dictionary | |
| |  |  |  | | --- | --- | --- | | **Feature** | **Data Type** | **Description** | | X | Interval | x-axis spatial coordinate within the Montesinho park map: 1 to 9 | | Y | Interval | y-axis spatial coordinate within the Montesinho park map: 2 to 9 | | month | Ordinal | month of the year: "jan" to "dec" | | day | Ordinal | day of the week: "mon" to "sun" | | FFMC | Ratio | FFMC index from the FWI system: 18.7 to 96.20 | | DMC | Ratio | DMC index from the FWI system: 1.1 to 291.3 | | DC | Ratio | DC index from the FWI system: 7.9 to 860.6 | | ISI | Ratio | ISI index from the FWI system: 0.0 to 56.10 | | temp | Interval | temperature in Celsius degrees: 2.2 to 33.30 | | RH | Ratio | relative humidity in %: 15.0 to 100 | | wind | Ratio | wind speed in km/h: 0.40 to 9.40 | | rain | Ratio | outside rain in mm/m2 : 0.0 to 6.4 | | area | Ratio | the burned area of the forest (in ha): 0.00 to 1090.84 | | |
| Data Exploration | |
| Time | |
| |  |  | | --- | --- | | A graph of blue bars  Description automatically generated | A graph of a number of blue bars  Description automatically generated | | |
| *Figure 5: Forest fire outbreaks appear to be most frequent from Friday to Sunday and during the months August to September.* | |
| Location | |
| |  |  | | --- | --- | |  | *A graph of a forest fire y frequency  Description automatically generated* | |  | | | |
| *Figure 6: The frequency of forest fires is shown to appear on the central right and central left areas in Montesinho park.* | |
| FWI System | |
| |  |  | | --- | --- | |  | A graph of different colored squares  Description automatically generated | | A graph of a number of different colored boxes  Description automatically generated | A graph of different colored boxes  Description automatically generated | | |
| *Figure 7: The FWI feature at each month appears somewhat correlated to forest fire frequency.* | |
| Weather | |
| |  |  | | --- | --- | | A graph of different colored boxes  Description automatically generated with medium confidence | A graph of different colored and red squares  Description automatically generated with medium confidence | | A graph of different colored squares  Description automatically generated with medium confidence | A graph with numbers and a line  Description automatically generated with medium confidence | | |
| *Figure 8: Temperature during the summer-spring months is shown to be higher and thusly have a higher frequency of forest fires. Interestingly, wind speeds do not appear to be closely correlated to time of year nor the frequency of forest fires. Additionally, rain is infrequent in Montesinho park.* | |
| Area | |
| |  | | --- | | A graph of a forest fire area  Description automatically generated | | |
| *Figure 9: Burned area is composed of a wide range (0.00 to 1090.84 ha) with most instances distributed at zero for all months.* | |
| Data Preprocessing | |
| Feature Scaling | |
|  | |  |  |  | | --- | --- | --- | | Feature | Transformation | | | Area**:** | As seen in the [data exploration](#_Data_Exploration) section, the “Area” feature shows a wide range where many data points are distributed around zero. As this study employs the use of neural networks, wide target variable ranges may slow down the learning rate of the model. Hence a transform was applied to the target variable to reduce its range. | | |  | | A graph of a log and log  Description automatically generated | | *Figure 10: Comparison of pre and post-scaled forest fire area data.* | | | |
| Normalizations | |
|  | For this report, normalization/standardizations were performed on all input features such that the ranges between each feature are not disproportionate to one another. This is significant as for SVM, variable feature ranges could influence the determining hyperplane and decision boundaries. While for neural networks, inconsistent feature ranges may cause oscillations or increase the number of iterations required to learn from the data. |
|  | |  |  | | --- | --- | | **Technique** | **Description/Justification** | | Z-Score: | Transforms the feature such that standard deviation is equal to 1 and the median is set to 0. This method is useful for models that may receive new instances that fall outside the maximum and minimum ranges of the existing dataset. | |  | | Min-Max: | This technique distributes the feature between 0 and 1. It is useful in ensuring that all features have the same weight during model training. However, it is not designed to be used for instances where inputs fall outside the existing minimum-maximum feature range. | |  | |
| Correlations | |
|  | The correlations matrix shows that excluding the month feature, no other feature appears particularly indicative of predicting the target variable (area) for both pearson and spearman correlations. Hence for this report, no features were removed nor created for model training. |
|  | |  |  | | --- | --- | | zscore\_ff.corr(method="pearson")["area"]  X 0.061995  Y 0.038838  month 0.114280  day 0.000208  FFMC 0.046799  DMC 0.067153  DC 0.066360  ISI -0.010347  temp 0.053487  RH -0.053662  wind 0.066973  rain 0.023311  area 1.000000  Name: area, dtype: float64 | zscore\_ff.corr(method="spearman")["area"]  X 0.060499  Y 0.046018  month 0.117681  day -0.016799  FFMC 0.025300  DMC 0.071920  DC 0.061633  ISI 0.012496  temp 0.078696  RH -0.024221  wind 0.053196  rain -0.064073  area 1.000000  Name: area, dtype: float64 | | minmax\_ff.corr(method="pearson")["area"]  X 0.061995  Y 0.038838  month 0.114280  day 0.000208  FFMC 0.046799  DMC 0.067153  DC 0.066360  ISI -0.010347  temp 0.053487  RH -0.053662  wind 0.066973  rain 0.023311  area 1.000000  Name: area, dtype: float64 | minmax\_ff.corr(method="spearman")["area"]  X 0.060499  Y 0.046018  month 0.117681  day -0.016799  FFMC 0.025300  DMC 0.071920  DC 0.061633  ISI 0.012496  temp 0.078696  RH -0.024221  wind 0.053196  rain -0.064073  area 1.000000  Name: area, dtype: float64 | |

|  |
| --- |
| Experiments |
| Normalization |
| This work compares the effect of normalization methods on both SVM and neural network models as forest fire prediction is a mission critical system, it is important to select the model that produces the most accurate results. As outlined in the [data preprocessing](#_Data_Preprocessing) section, z-score and min-max normalization methods were selected for this study. Additionally, variations in feature scales/ranges may affect the accuracy of each model due to distances between data points. |
| Support Vector Regressor |
| |  |  |  | | --- | --- | --- | | **Parameter** | **Configuration** | **Justification** | | Kernel | Radial Basis Function | RBF is a popular kernel for lower dimensional data and is the same method used by Cortez and Morais (2007) on the same dataset. | | Epsilon | 1 | Multiple epsilon values were tested between the ranges from the default 0.1 to 2. Ultimately, the values producing the best performance were between 0.8 and 1.5. This study settled on the value of 1. | |
| Neural Network |
| |  |  |  | | --- | --- | --- | | **Parameter** | **Configuration** | **Justification** | | Hidden layers | 3 | Configurations of hidden layers between 1 and 4 were evaluated. This work prioritised increasing the depth of the network before depth and found 3 hidden layers generally produced lower MSE scores compared to alternate configurations where the model would be over or underfitted. | | Neurons | |  |  |  | | --- | --- | --- | | Layer | Z-score | Min-Max | |  | 12 | 12 | |  | 256 | 32 | |  | 128 | 16 | |  | 32 | 8 | |  | 1 | 1 | | As mentioned previously, model depth was prioritised over breadth. Neuron selection began from 100 neurons at each hidden layer. Based on the resulting MSE score, the number of neurons at each hidden layer were manually configured until an acceptable error was produced. | | Activation Function | Rectified Linear Unit | The RELU activation function was adopted as opposed to sigmoid due to the nature of the regression problem. | | Epochs | 10 | Epoch selection began from 100 as outlined in previous literature (Cortez & Morais, 2007) and were manually adjusted until an acceptable MSE was produced. | | Batch Size | 8 | Manually selected in exponents of 2 to limit over/underfitting. | | Loss Function | Mean Square Error | Same as the regression metric that this study evaluates both models on. | | Optimizer | Adam | Efficient method for adjusting the learning rate of the neural network that is a combination of gradient descent and RMSProp optimizer. (prakharr0y, 2024) | |
| K-Fold Cross-Validation |
| The k-fold cross-validation method was selected for this work as opposed to a single train-test split validation for increased robustness in model evaluations. |
| |  |  |  | | --- | --- | --- | | **Parameter** | **Configuration** | **Justification** | | n splits | 5 | Each split iteration tests against 20% of the dataset. This configuration is common among test splitting and attempts to retain the representativeness of the dataset in each split. | | Shuffle | True | Randomizes the dataset before splitting. This is useful in the case that the dataset is initially ordered. | |
| Model Evaluations |
| Mean Square Error |
| MSE was chosen as the primary metric of evaluation for simple interpretability and the direct relation to the loss function in the neural network model. |
| Results |
| |  |  |  | | --- | --- | --- | |  |  |  | | Z-score NN | 1.788 | 2.809 | | Z-score SVM | 1.959 | 3.054 | | Minmax NN | 1.897 | 2.965 | | Minmax SVM | 1.944 | 3.032 | |
| Findings |
| In both normalization cases the neural network performed better than the SVM regressor. Additionally, the z-score neural network appears to have significantly increased the accuracy of predicted results comparatively to every other model. Notably, neither normalization technique appears to produce any significant difference in model accuracy for SVM models. |

### Comparisons of different AI techniques

|  |  |
| --- | --- |
| Support Vector Regressor | |
| **Advantages** | **Disadvantages** |
| * Robust to small datasets * Optimal margin | * Scalability * Difficulty handling high-dimensional data |
| The robustness of the SVM regressor was such that MSE did not significantly change despite the differing normalization transforms on the dataset. Although the epsilon parameter was tested throughout multiple ranges, the model performed well in identifying a similar optimal margin with an epsilon range between 0.8 to 1.5.  The Montesinho park forest fire dataset is relatively small such that the model was able to make reasonably accurate predictions. However, increasing the number or rows to a greater time series range or adding columns/dimensions may introduce complexity that is difficult for this method to handle. | |
| Neural Network | |
| **Advantages** | **Disadvantages** |
| * High complexity handling * Scalability | * Data dependency * Overfitting |
| This report showed that deep neural networks were able to capture the complexity in forest fire data. In the earlier parameter configuration stages, it was found that incorrectly tuning these parameters would cause overfitting or underfitting.  Despite the limited size of the Montesinho park forest fire dataset, the neural network model was able to make more accurate area spread predictions compared to both minmax and z-score normalized configurations of SVM. It was found that the feature scaling of input variables has a significant impact on model accuracy where z-score normalization produced more accurate results. | |

## Conclusions and Recommendations

|  |
| --- |
| Conclusion |
| The results of this study show that for both normalization methods, deep neural networks can more accurately predict the spread of forest fires in Montesinho park compared to SVM regressors. In the case of mission critical systems, making accurate predictions of the potential harms are imperative for first responders. Previous work by Cortez and Morais (2007) found that an SVM regressor were able to predict the scale of forest fires more accurately than a neural network with a single hidden layer. This report expands on previous literature to identify a solution that produces more accurate solutions in a state of emergency. Further work into this topic would enable authorities to appropriately respond to outbreaks and could be scaled to different geographies and meteorological conditions. |

|  |
| --- |
| Recommendation |
| When comparing each configuration of neural networks and SVM regressors, it is recommended that authorities choose some configuration of deep neural network. As shown in the [model evaluations](#_Model_Evaluations), deep neural networks outperformed SVM regressors across a robust k-fold cross-validation method. This is justified as the accuracy of mission critical systems directly influence the preparedness of first responders. When selecting any given deep learning method, a z-score normalized deep neural network generally produces more accurate results. As z-score normalization is not bound by minimum and maximum ranges, this configuration can be applied to real-time predictions systems. |

## List of References

Cortez, P., & Morais, A. d. J. R. (2007). A data mining approach to predict forest fires using meteorological data.

prakharr0y. (2024, 20 Mar, 2024). *What is Adam Optimizer?* Geeks for Geeks. Retrieved 28 Sep, 2024 from <https://www.geeksforgeeks.org/adam-optimizer/>

## Individual reflection

|  |
| --- |
| Planning |
| Read through the reference material. |
| Identified the hyperparameters used in previous literature. |
| This report did not attempt to perfectly mirror the reference literature, but to determine the intuition behind the study and incorporate justifiable actions into this work. |
| Any intuition that did not appear completely sound was reevaluated or generalized. For example: normalization process in previous literature only considered z-score normalization where this work used both normalization processes in the experimentation. |

|  |
| --- |
| Learnings |
| Clarified the process of back propagation and gradient descent |
| Approached SVM from a new perspective where instead of classification, the model was used for a regression task. |

|  |
| --- |
| Challenges |
| Encountered difficulty tuning the hyperparameters for each configuration of neural network. |
| Perhaps performing some method of cross-validation to find the optimal configuration of neural network and SVM parameters would be more correct. |
| The reference material covered some level of depth that is beyond my current knowledge. Hence it was difficult to mirror the exact process that was used and to effectively justify the chosen parameters (e.g. Sequential Minimum Optimization algorithm). |

|  |
| --- |
| Limitations |
| The experimentation in previous work was done using the R programming language. |
| Differing programming environments may have variations in the default parameters and/or implementations of data mining models. Thus, it was difficult to directly mirror the prior case study. |

|  |
| --- |
| Future |
| Apply meteorological data for different regions to evaluate the model effectiveness in different geographies. |
| Vary the size of the dataset to evaluate accuracy for more complex data. |
| Implement selected models into real-time systems and determine the accuracy of predictions while the system is in production. |