

Firetrace: Modelling Bushfires At Continental Scale

[Details hidden for privacy]

Abstract

We document the development of Firetrace; an AI model and web interface that uses projected weather data to predict the severity of bushfire events at a continental scale. It uses a deep neural network trained on temperature and Southern Oscillation Index (SOI - a measure of La Nina and El Nino events) information from BOM, fire information from NASA's MODIS satellite as well as date information to represent seasonality and climate change trends. Severe bushfire events cause sudden habitat loss, endanger many animal species and drastically impact the economy. Accurate modelling of bushfires has the potential to inform future policy decisions and prepare for the challenges of a bushfire event.

While not ready for production use in scientific applications, Firetrace offers some insight with reasonable accuracy into the relationship between the analysed factors and the size of bushfire events. With further work, Firetrace could help bring awareness to the effects of climate change, and other environmental factors, on the size of bushfires.

1 Introduction

We aim to create a publicly and freely accessible, reasonably accurate method of predicting the scale of future bushfires across the land area of Australia, as existing solutions are difficult to access outside of scientific institutions. An accessible solution could increase awareness of the relationship between climate and fire events and inform decisions around resource management in Australia.

1.1 Inspiration

Bushfires are a pressing environmental issue with large ecological and economic consequences. Part of the inspiration for this project came from the devastating aftermath of the 2019-2020 Black Summer bushfires. An estimated 3.2 billion dollars of assets were lost during the fires and projections for the total economic impact of the event are significantly higher. (Wittwer et al., 2021). Additionally, an estimated 3 billion vertebrates were displaced, harmed or killed in the fires (World Wide Fund for Nature, 2020).

Further investigation reveals that this is an actively researched field. Solutions currently in development such as CSIRO's Spark specialise in accurate, small-scale predictions with details such as predicted spread patterns and speed. (CSIRO, n.d.). We wanted to explore different approaches that prioritise continental-scale prediction as well as accessibility to people without extensive fire management training.

2 Approach

2.1 Selecting and Finding Data

A non-exhaustive list of potential factors that affect the size of bushfires was formed.

- Maximum daily temperature
- Density/land area of vegetation
- Humidity
- Rainfall
- Solar irradiance (the energy from the sun that makes it to Earth's surface)
- Wind speed

To reduce the difficulty of finding suitable datasets, potential approximations for the above factors were used.

No approximations found

- Maximum daily temperature
- Vegetation - assumed to be constant across bushfire seasons.
- Wind speed - unable to be accounted for.

Time of year

The following factors may be seasonal and have a strong correlation with time of year.

- Humidity
- Rainfall
- Solar irradiance

Year since measurements began

The following factors may be linked to long-term effects such as climate change and are correlated to time.

- Humidity
- Rainfall

Southern Oscillation Index

The following factors may be linked to the Southern Oscillation Index (typically used as the definition for El Nino and La Nina events).

- Temperature
- Humidity
- Rainfall
- Wind speed

Pre-published datasets on Kaggle were used, sourced from NASA's MODIS satellite, a BOM weather observatory in Sydney, a BOM weather observatory in Brisbane and BOM's monthly measurements of the Southern Oscillation Index. Refer to Section 6.2.

2.2 Cleaning Data

The following describes the method used to clean data irrespective of where the implementations are written.

Satellite Data on Australian Fires

1. Removed any potential sources of inaccuracy by dropping rows with confidence less than 95.
2. Daily totals for fire size were generated.
3. Unused columns were dropped.

Australian Weather Observation Data

1. A 'date' column was created, replacing the 'day', 'month' and 'year' columns.
2. Columns were renamed to more sensible names (e.g. "Maximum temperature (Degree C)" to "max_t")
3. Unused columns were dropped.

Southern Oscillation Index

1. Column names were added.

2.3 Feature Engineering

Time of Year

Encoding time-of-year information as a combination of sine and cosine waves is a relatively common method. (NVIDIA, 2022). Encoding data in such a way means that for every day of the year, there is a unique pair of values (from the sine and cosine) functions. Furthermore, this method represents the cyclical nature of a year, as days that appear to be on different ends (e.g. December 31st and January 1st) are, in reality, close together.

2.4 Designing and Validating Models

Initial trials in Orange revealed that the best model suited for the problem may be a Deep Neural Network (DNN), rather than simpler methods such as linear regression.

The final DNN was written in Python with Keras.

```
# Utility function for the model
def build_and_compile_model():
    model = tf.keras.Sequential()

    model.add(normalizer)

    for i in range(5):
        model.add(tf.keras.layers.Dense(40, activation="relu"))

    model.add(tf.keras.layers.Dense(1, activation="relu"))

    model.compile(loss=custom_loss, metrics=["mean_absolute_error"], optimizer=tf.keras.optimizers.Adam(learning_rate=0.001))
    return model
✓ 0.0s
```

Figure 2.4: *The 28th July version of the model.*

Model Shape

While there are numerous “rule-of-thumb” methods available to determine the shape of a neural network, it was found that the best results came from some experimentation with the number of

hidden layers and input nodes. A very deep model (relative to the complexity of the problem) was chosen as it had a better chance of picking up subtle patterns while keeping file size and inference latency low.

2.5 Training

Custom Loss

It was decided that overestimating fire sizes would be less harmful than underestimating fire sizes. To reflect this priority in the training of the model, a custom loss function was written that added additional penalties for underestimating fire sizes.

```
# Custom loss to punish undershooting. Based on mean_squared_error.
def custom_loss(y_true, y_pred):
    error = y_true - y_pred
    squared_error = tf.square(error)
    undershoot_error = tf.square(tf.math.maximum(0.0, error)) / 2 # undershooting is 1.5x as bad as overshooting.
    return tf.reduce_mean(squared_error + undershoot_error, axis=-1)
```

Figure 2.5.1: *The custom loss function*

The loss function is based on Mean Squared Error, which punishes large errors significantly more than small errors.

Early Stopping

To prevent overfitting (the model “remembering” training data and being unable to generalise to new data), the model was configured to automatically stop training after validation loss stops improving.

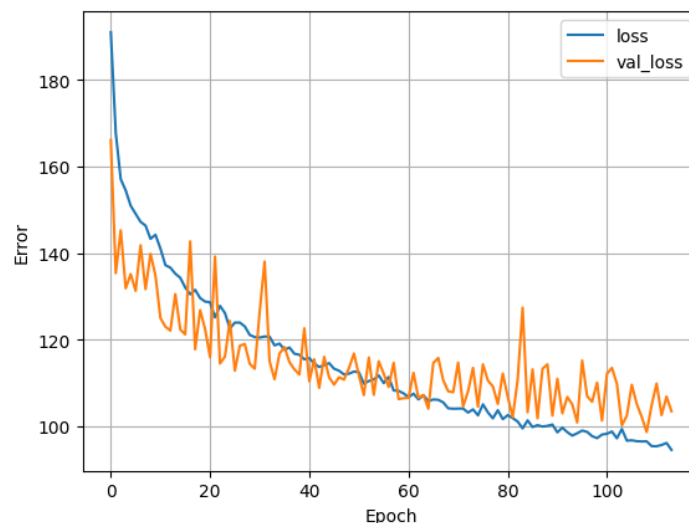


Figure 2.5.2: *Graph displaying the training curves during the training process of the model*

The training curves (values shown are in terms of Mean Absolute Error for interpretability) indicate no sign of under/overfitting. After training, the checkpoint that performs the best is loaded and saved.

2.6 Validation

To visualise the performance of the model, the model's predictions were plotted against the actual, ground-truth values of the test dataset.

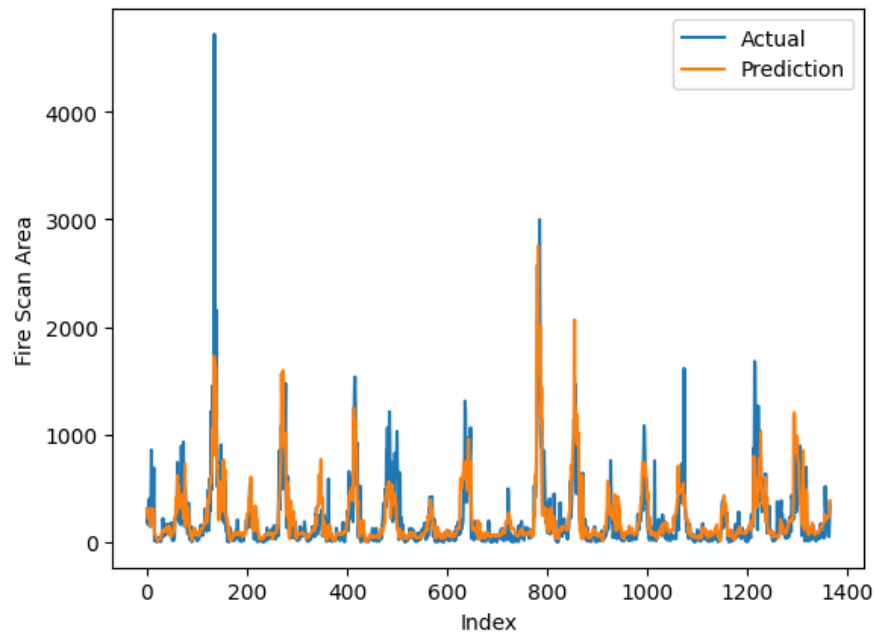


Figure 2.6: Graph displaying a comparison between the model's predicted data and actual data

Section 3.1 discusses the results in more detail.

2.7 Distribution

A user interface to interact with the model was written with Gradio and hosted on HuggingFace to ensure that the model was accessible to the general public. The user interface also offers a few useful insights to assist with interpreting the model's output, to put the size of these fires into perspective.

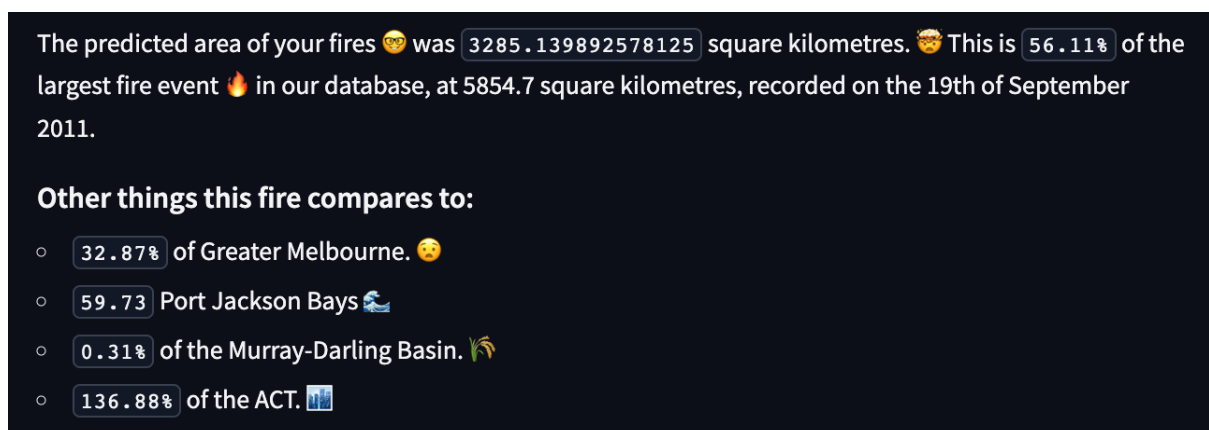


Figure 2.7: A screenshot of a part of the user interface displaying how the output is presented

An API accessible through HTTP requests is also available.

3 Limitations and Challenges

3.1 Results and Limitations

The model can accurately predict the timing of fire events without fire size information from previous days, significantly exceeding expectations.

However, it does not appear to be very effective at determining the scale of bushfire events. As shown by Figure 2.5.2, the model is off from the ground truth predictions by ~100 square kilometres on average. Accuracy is also expected to decline when trying to predict fires further into the future. Figure 2.6 reveals that the model is significantly less accurate during extreme events. The custom loss function addresses this to a limited extent, and the strength of punishments for underestimated values needs to be re-evaluated.

The model's reliance on projected weather information further reduces its accuracy. Projected weather data may not be as accurate as the historical data the model was trained on.

Additionally, the model was only trained on data leading up to (approximately) 2016 and validated on data between 2016 and 2019. The inclusion of up-to-date data is likely to increase the model's accuracy in predicting future events.

Because of these shortcomings, Firetrace in its current state should not be used as a forecasting tool for mission-critical applications.

3.2 Challenges

The initial focus of the project was to use time-series forecasting to predict the scale of future fires based on the size of previous fires. Time series forecasting depends on strong patterns in historical data. However, bushfires do not have strong means that they were difficult to predict from time series alone, regardless of which algorithm we used.

Methods considered:

- Linear regression
- Prophet (from Meta/Facebook)
- SARIMA/SARIMAX
- Deep neural network

Therefore, we adapted our method to use weather data to predict the size of fires instead. Meteorological information such as temperature has been heavily researched for centuries, and accurate methods exist to predict future values. There is also a stronger relationship between temperature and size of fire events making it easier to fit a model to this data.

Additionally, programming new solutions for every dataset and configuration proved to be an extremely arduous and inefficient process. The use of graphical, low-code tools such as Orange greatly reduced the time required for prototypes to be made. Once confident, a more polished model was created using Tensorflow/Keras, and model hyperparameters (number of nodes and layers, learning rate, etc) were further experimented with.

A lack of easily obtainable information surrounding wind speed, fuel loads and humidity also made it more difficult to construct an accurate bushfire prediction model. Unfortunately, this issue has no easy resolution and will require the assistance of domain experts such as machine learning and climate scientists.

4 Improvements

This project could have been completed more effectively if conversations with climate and fire experts were conducted to increase domain knowledge, allowing for a more tailored solution. Additionally, using more data points such as projected carbon dioxide and vegetation (fuel load) levels, as well as including more recent data, would have improved the accuracy of the model. The model could also have additional outputs, such as the predicted economic impact of fires and the number of animals that may be affected by the fires. Also, if sufficient time was provided, a visualisation of the prediction (ie. a map demonstrating the area covered) could be included for more convenient usage.

5 Conclusion

Firetrace is a bushfire modelling tool that uses machine learning to predict the severity of bushfire events in terms of the area covered. It is distributed for public use through a Gradio interface hosted on HuggingFace. The reasonably high level of accuracy allows the public to explore the relationship between fires and a range of factors, gaining an understanding on the behaviour of fire events. Though there is significant room for improvement, Firetrace represents a significant step towards highly useful natural disaster forecasting systems, presenting potential to inform future policy decisions and prepare for the challenges of a bushfire event. In the future, a range of other factors and methods may be explored to further increase accuracy.

6 References

6.1 Websites/Reports

CSIRO. (n.d.). *Spark: Predicting bushfire spread*. CSIRO. Retrieved July 16, 2023, from <https://www.csiro.au/en/research/technology-space/ai/Spark>

Three Approaches to Encoding Time Information as Features for ML Models. (2022, February 17). NVIDIA Technical Blog. <https://developer.nvidia.com/blog/three-approaches-to-encoding-time-information-as-features-for-ml-models/>

Wittwer, G., Li, K., & Yang, S. (2021, June). *The economic impacts of the 2019-20 bushfires on Victoria*. Department of Treasury and Finance Victoria. <https://www.dtf.vic.gov.au/victorias-economic-bulletin-volume-5/economic-impacts-2019-20-bushfires-victoria>

World Wide Fund for Nature. (2020). *Australia's 2019-2020 Bushfires: The Wildlife Toll*. World Wide Fund for Nature.
https://assets.wwf.org.au/image/upload/v1/website-media/resources/Animals_Impacted_Interim_Report_24072020_final?_a=ATO2Bfg0

6.2 Datasets

Australian Bureau of Meteorology. (2023). [Monthly data regarding the Southern Oscillation Index] [Data set] http://www.bom.gov.au/climate/enso/soi_monthly.txt

Cheng, A. (2020). Australian Weather Observation Data. [Data set] *Kaggle*.
<https://www.kaggle.com/datasets/alcheng10/bom-weather-observation-data-select-stations>

Gutierrez, G. B. (2020). Satellite Data on Australian Fires. [Data set] *Kaggle*. Retrieved July 14, 2023, from
<https://www.kaggle.com/datasets/gabrielbgutierrez/satellite-data-on-australia-fires>

7 Appendices

7A Source Code and Demo

<https://github.com/jtpotato/firetrace>

A GitHub repository containing archives of all Jupyter Notebooks, source code for the web UI, some data used for training as well as the final trained AI model.

<https://huggingface.co/spaces/jtpotato/firetrace>

HuggingFace demo of the model, open for anybody to use.