Prediction models were built using Spark in python. The dependent variable was selected as state for the first model. Logistic regression was used after encoding the categorical data. The accuracy of the model was 0.86. The prediction variable has features such as probability. Each row in the data has a value for probability. This feature can provide the probability of bushfire for different places under various weather conditions.

A close up of a map

Description automatically generated

Figure -ROC CURVE OF LOGISTIC REGRESSION

Another model was built using Multinomial Logistic Regression. This model has an accuracy of 0.86. From the predictions it was seen that Suburbs such as Bairnsdale with temperatures higher than 38 degree Celsius had the highest probability of bushfire of 0.879. Also, it can be seen from the predictions that areas which had temperatures lower than 12 degrees Celsius and humidity greater than 30% had probabilities of bushfire of only 0.39 or lower. Hence, we can say that weather conditions do have an impact on bushfires.