A picture containing drawing, game

Description automatically generated

Hand-over documentation of bushfire data

EMERGENCY MANAGEMENT SIT782 T1 2020 DEAKIN UNIVERSITY BURWOOD

Table of contents

Team Member details………………………………………………………………………………………………………………………..2

Project Overview……………………………………………………………………………………………………………………………….2

Document purpose……………………………………………………………………………………………………………………………2

Project Description……………………………………………………………………………………..…………………………………….3

Initial Client Requirements…………………………………………………..………………………………………………………….3

Milestone Objectives On Object Initiation…………………..…………………………………………………………………..4

What was agreed with the Customer?.................................................................................................4

Completed Deliverables…………………………………………………………………………………………………………………….5

Open issues……………………………………………………………………………………………………………………………………….6

Lessons learned…………………………………………………………………………………………………………………………………6

Technology and Tools……………..………………………………………………………………………………………………………..7

Development process………………………………………………………………………………………………………………………..8

Architecture Of The Project……………………………………………………………………………………………………………..8

Data Sourcing………………………………………………………………………………………………………………………………….8

Data Cleaning……………………..…………………………………………………………………………………………………………..9

Data Aggregation…………………………………………………………………………………………………………………………….9

Predictive Analytics………………………………………………………………………………………………………………………..10

Logistic Regression……………………………………………………………..……………………………………………………….10

Data Visualisation…………………………………………………………………………………………………………………………….11

Website Creation…………………………………………………………………………………..………………………………………..14

Source code…………………………………………………………………………………………………………………………………….14

Other Relevant Document……………………………………………………………………………………………………………….15

Appendices………………………………………………………………………………………………………………………………………15

# Team lead:-

Nikita Dighe

# Academic Mentor:-

Chetan Arora

Technical supervisor**:**-Matt cohan

# Team Member Details

|  |  |  |  |
| --- | --- | --- | --- |
| **Name** | **Role** | **Phone Number** | **Email ID** |
| Nikita |  | 0414612037 | ndighe@deakin.edu.au |
| Pavan Kumar Vadla |  | 0451225828 | pvadla@deakin.edu.au |
| Vijay Bathini |  | 0433884945 | vbathini@deakin.edu.au |
| Varsha Sankar |  | 0433953196 | vsankar@deakin.edu.au |
| Piyush Gandhi |  | 0402041560 | gandhipi@deakin.edu.au |
| Bala Tharun Reddy Shyamala |  | 0490357774 | bshyamala@deakin.edu.au |
| Nikhil Reddy Viraati |  | 0452426575 | nviraati@deakin.edu.au |
| Aiswarya  Subrahmanian |  | 0432247154 | Asubrahmanian@deakin.edu.au |
| Simran Rashpal Dhillon |  | 0410425500 | Srdhillon@deakin.edu.au |
| Nithin reddy gundre |  | 0469818857 | ngundre@deakin.edu.au |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

# Project Overview:-

The primary goal of the project is to predict areas which have a considerable risk for bushfires across bushfires and developing a website with information based on bushfire regions which have a high risk of facing fires

# Document Purpose:-

This document has all the information and artefacts developed during Trimester 2020, the following handover document includes:-

1. Initial purpose**:-**To predict areas which are prone to bushfires across Australia and create a website which gives information about bushfires across Australia, survival guidelines and emergency hotline numbers

2.Plan:- To collect data which has all the required features, improve the quality of the data by performing data cleaning, trying out models which give high accuracy, visualising interesting insights using Tableau Software and to make a website with bushfire information

3.Status report**:-** We have applied machine learning models to predict areas which are prone to bushfires across Australia, visualised some exciting insight with maximum temperature and fire radiation power and created a website which is used by the general audience

# 4.Future work:-

* We are going to develop a web application which uses a satellite image of active and inactive bushfires occurring across Australia
* We are going to include emergency hotline numbers of nearest fire brigades and survival guides to make sure you’re safe during a bushfire situation

# Project Description:-

**Emergency Management**:-

Emergency management is the managerial function charged with creating the framework within which communities reduce vulnerability to hazards and cope with disasters. The vision of emergency management seeks to promote safer, less vulnerable communities with the capacity to cope with hazards and disasters.

So by using data science techniques, we are going to figure out the areas which are prone to bushfires and also predict which areas have a high risk of catching fire. To do this, we are going to collect data of bushfire from the last 6 months and weather data. To help people to know the places which have chances of bushfires we are creating a website which gives out information about bushfires across Australia which includes intensity levels of fires, visualisation of frequency of fires over the years, increases of heat over the years and also provides with the emergency hotline

The project includes complete data science life cycle – Data Collection, Data Cleansing, Data Aggregation, Data Visualisation, Model Building, Forecasting based on Data Analysis and creating a website

# Initial Client Requirements:-

A. Identify patterns of

* FRP:- Fire Radiation Power
* Suburbs
* State
* Region
* Temperature
* Humidity
* Windspeed
* Pressure

B.finding correlation between the features mentioned above

C. Creating a machine learning algorithms for bushfire data

D. Predicting the areas which are highly prone to bushfires

E. Creating a web application which includes information about bushfires across Australia

# Milestone objectives on object initiation:-

**3rd  March 2020 to 31st  March 2020**:-Decided to complete the visualisation of ambulance data, extracting required datasets for bushfires and weather and planning out a development plan for the project

**1st  April 2020 to 28th April 2020**:-cleaning the data to increase the accuracy of the data, testing out models which give out the highest accuracy and done with creating the main page for the website

**28th  April 2020 to 18th May 2020:-**Done with prediction analysis, visualisation of exciting patterns by using Tableau and connected the database to the website

**18th May 2020 to 30th May 2020**:- Making the website live and handover documentation of the bushfires

# What was agreed with the Customer?

the Client has provided with data in the following link:

<https://nrt4.modaps.eosdis.nasa.gov/>

at the beginning of the iteration, the data was not sufficient to satisfy the Client, so we started searching for the information which has all the requirements

so for bushfire, there was the need for finding weather data and bushfire data as the weather is the leading cause of the fire and by correlating both the datasets, we produce some interesting patterns and visualise them

Link for the weather data:-

<https://www.worldweatheronline.com/>

we agreed to create a website for everyone and agreed to include locations of bushfire prone areas, survival guide and emergency hotline numbers

Below Timelines were agreed upon discussion with the Client:-

**3rd  March 2020 to 31st  March 2020:-** Done with finding scope for the project and researching about the existing information about the bushfires, and development pan

**1st  April 2020 to 28th April 2020:-**Extracted the required data, improving the quality of the data by performing data cleaning and joining both the data sets, the main page for the website, testing out models for high accuracy

**28th  April 2020 to 18th May 2020:-** predicting the areas which are highly prone to fires, visualising exciting patterns by using Tableau and connecting the database to the website

**18th May 2020 to 30th May 2020:-**Integrating the visualisation part into the website, making the website live and preparing handover documents

# Completed Deliverables:-

|  |  |  |  |
| --- | --- | --- | --- |
| Deliverable | Description | Source | Contributor |
| Data sourcing | Collected data which required to predict areas which are prone to bushfires | For bushfire data:-<https://nrt4.modaps.eosdis.nasa.gov/>  For weather data:-  <https://www.worldweatheronline.com/> | 1.Nithin  2.Mohan  3.shikhar  4.Varsha  5.Vijay  6.Bala Tharun Shyamala |
| Data Cleaning | We have improved data by removing null values, unwanted attributes and duplicates | We have used Python for cleaning then data | 1Bala Tharun Shyamala  2.Nithin  3.Shikhar |
| Prediction | We have used machine learning models to get the highest accuracy  We choose to go with Logistic Regression as it has the accuracy of |  | 1.Aiswarya Subramanian  2.Nitin  3.Pavan Vadla  4.Nikita Dighe  5.Piyush |
| Data Visualisation | We have used Tableau to find exciting features like maximum temperature and fire radiation power |  | 1.Nikhil Viraati  2.Nithin  3.Pavan  4.Bala Tharun shymala  5.Nikita Dighe  6.Piyush |
| Database Creation | We have to connect the database to a website to start working |  | 1.Nikita Dighe  2.Varsha  3.Simran |
| Website Creation | Created a website with info on bushfires across Australia by using UI, ATOM and Microsoft Azure |  | 1.Simran  2.Nikita Dighe |
| Handover Documentation | Document about the project and links to files and website |  | Vijay Bathini |
| Git Repository |  |  | Pavan Vadla |

# Planned Work:-

In future, we want to create a web application which has a satellite image of active /inactive bushfires across Australia and includes a survival guide which includes how to protect yourself from fires and Emergency hotlines to the nearest fire brigades

# Open issues:-

* During the data sourcing, we faced a problem with a limit to extract data from the website, so we had to collect keys from other teammates to extract more data
* In the process of aggregating two datasets, we faced problem in combing the timestamps, so we had to take the nearest time when the fire occurred
* For prediction, we had trouble choosing machine learning models as we have to have high accuracy so at last, we used Logistic Regression as has the highest accuracy among other models
* We had trouble choosing exciting insights to visualise and share with the public as we had data for only 6 months
* The process of making a website was the biggest challenge as most of the team do not know about User Interface so had to arrange much time in up-skilling in User Interface

# Lessons learned:-

* Most of the people in the group don’t have experience working in a team so, we learned 1.how to divide and conquer

2 how to respect other opinions regarding the project,

3.helping each other in difficulties,

4. time management

5.decision making

* Have good knowledge about using Python and learned different techniques like Predictive analysis, Data Visualisation by Tableau and website creation by using UI

# Technologies and Tools

* Tableau and Python for visualisation
* R – “**tidyverse**” – package for Data cleansing
* Python – used for building machine learning models and for visualisation
* UI(user interface), ATOM and Microsoft- This tools are used in creating the website

Why are we using Python?

* Compatibility:- you can run python code in any operating system by using Python interpreters, and it also allows us to run the Code on various platforms
* Python is a standard library from we can choose a wide range of modules according to the requirements
* Python is an open-source framework which decreases cost and time for the project and offers many features
* In Python, the syntax rules allow us to explain the concept without any

additional Code and features code readability and allow us to use English keyword

* Python can be used for plotting graphs for representing interesting patterns

Why are we using Tableau?

* For representing interesting patterns between two variable we use Tableau software which has a simple drag and drop option which helps us to analyse exciting insights
* Tableau is easy to integrate with different data sources as we have to link different data types in our project

Project Involves all the different stages in Data Science Lifecycle that is data collection, data cleansing, data aggregation, data visualisation, model building and prediction based on the analysis. As it was agreed with the Client, the team completed predicting the areas which are highly prone to bushfires, website creation, visualisation and pattern analysis based on each feature

Various stages in the development lifecycle of this project are as follows:

# 

# Development process

Analysing the Project Requirements:-

From the Initial Client Requirement, we found three crucial parts they are

1. Predicting the areas which are highly prone to bushfires
2. Visualisation of interesting patterns
3. Creating a website which has all the necessary information on bushfires across Australia

# ARCHITECTURE OF THE PROJECT

# Data Sourcing:-

A data source may be the initial location where data is born or where physical information is first digitised. However, even the most refined data may serve as a source, as long as another process accesses and utilises it.

We have extracted weather data and bushfire from the following links

For the weather data:-

<https://www.worldweatheronline.com/>

For the bushfire data:-

<https://nrt4.modaps.eosdis.nasa.gov/>

which has all the required features

* FRP:- Fire Radiation Power
* Suburbs
* State
* Region
* Temperature
* Humidity
* Windspeed
* Pressure

The challenge faced in the extracting the data is that there was a limit for a certain amount of data, so we had collect keys from some of the team members. By using the keys, we have extracted data from Dec 2019 to April 2020.

# Data cleaning:-

Data cleaning is a process of removing null values, duplicate values, removing unnecessary rows and columns which are not useful for the project and also to check if there are any outliers.

Data cleaning includes more actions than removing data, such as fixing spelling and syntax errors, standardising data sets, and correcting mistakes such as empty fields, missing codes, and identifying duplicate data points. Data cleaning is considered a foundational element of the data science basics, as it plays an essential role in the analytical process and uncovering reliable answers.

We used primary tools by importing packages of pandas and NumPy to the working directory to clean the data. our team has removed null values, duplicate values, and unnecessary rows and columns

# 

# Data aggregation:

Data aggregation is a process of combining data to get more information about a particular group based on specific variables such as age, profession and income

The problem faced during aggregating bushfire data and weather is using date timestamp as the date timestamp for bushfire and weather were different we have considered closest timestamp for both the data

# 

# Predictive analytics:-

As the name refers to it predicts the unknown feature events .predivtive analytics uses many techniques like Data mining, Statistics, Modeling, Machine learning and AI (artificial intelligence)

There are different models for predictive analytics, such as Logistic Regression, Multinomial Regression, Decision tree, Random Forest .we applied every model by taking forty per cent as the test data. To find the model which has the highest accuracy we used Multiclass classifier evaluator which splits the data into training data and test data by using random split function

To make sure that every data point has the same scale before applying Regression we normalised the data, so the primary way to normalise the data is to use Min-Max normalisation which turns minimum value into 1, and maximum value gets transformed into 1, and every other value gets transformed into a decimal between 0 and 1

# Logistic Regression:-

Logistic Regression is a statistical approach to predicting binary classes. The outcome variable is dual, which means that there are two possible outcomes.

We have categorical data so to convert it into numerical data we use one-hot encoder which is the representation of categorical variables as binary vectors this requires the categorical values mapped to integer values then each integer value is represented as a binary vector that is all zero values except the index of the integer, which is marked with a 1

After encoding, we apply Logistic Regression to the data to know the probability and also apply multiclass classifier evaluator to find the accuracy which turned to be 0.90. After applying Logistic Regression, the probability of finding places which are prone to fires is easy

A close up of a map

Description automatically generated

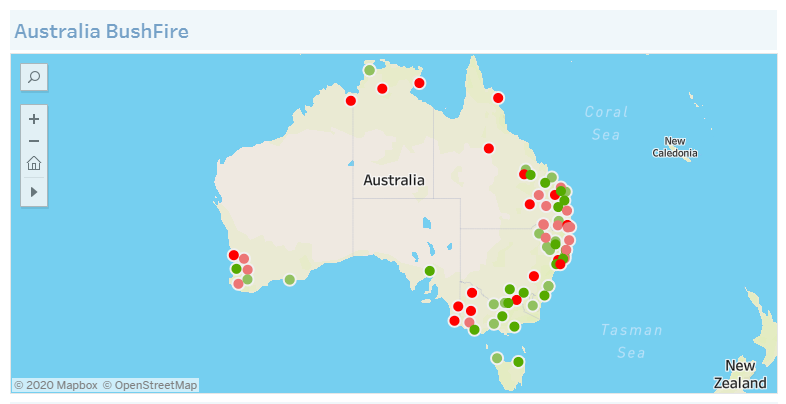
During the prediction analysis, the challenge faced is to choose a model that would give high accuracy

# Data Visualisation

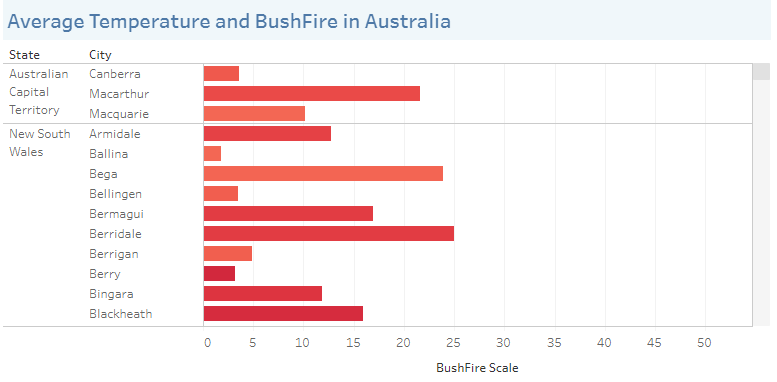
To visualise exciting patterns, we are going to use Tableau software and plotty software in R programming for interactive visualisation .in order to start the process the data should be cleaned and formatted.

Tableau Software has an option to upload the dashboards to Tableau public or private servers and then the HTML code is produced

Here are some of the fascinating insights that we have done using Tableau software



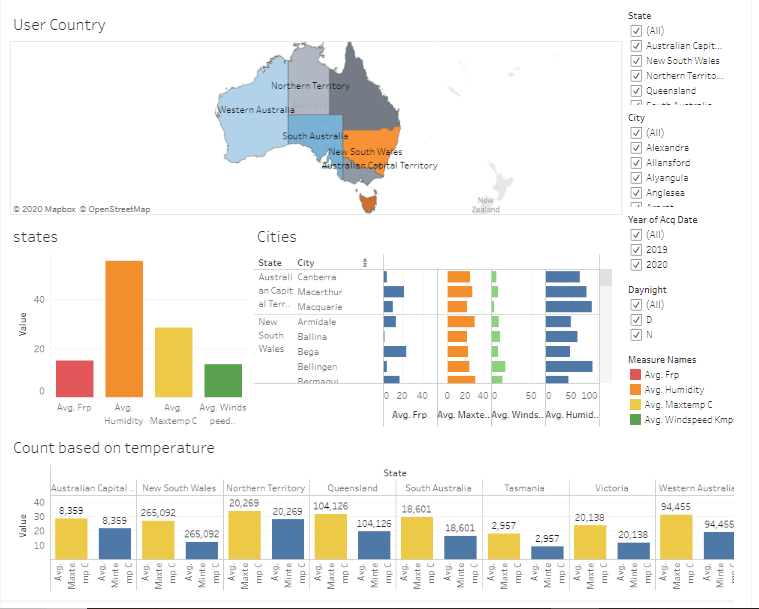
From the Australian Bushfire Dashboard, we can see the areas which are prone to bushfires throughout Australia, including the suburban areas. The intensity of the bushfire is scaled using a different colour. The green colour indicates the least affected areas, while red indicates the most affected areas. Bushfires from Dec 2019 to April 2020 is visualised using the scale



The leading cause of bushfires is a rise in temperature, so we made a visualisation based on maximum temperature and fire radiation power, which shows the correlation between both parameters across Australia

**Note:-**

: **Fire Radiation Power (FRP)** - FRP depicts the pixel-integrated fire radiative power in MW (megawatts). Given the unprecedented spatial and spectral resolution of the data, the VIIRS 375 m fire detection algorithm was customised and tuned to optimise its response over small fires while balancing the occurrence of false alarms. Frequent saturation of the mid-infrared I4 channel (3.55-3.93 µm) driving the detection of active fires requires additional tests and procedures to avoid pixel classification errors. As a result, sub-pixel fire characterisation (e.g., fire radiative power [FRP] retrieval) is only viable across small and low-intensity fires. Systematic FRP retrievals are based on a hybrid approach combining 375 and 750 m data. Starting in 2015 the algorithm incorporated additional VIIRS channel M13 (3.973-4.128 µm) 750 m data in both aggregated and unaggregated format.



This dashboard contains information of all the cities in different states where bushfires occurred. Variable of Interest Fire Radiation Power(FRP) is plotted against weather attributes like temperature, humidity and wind speed. The total number of cases in each state is compared with the minimum and maximum temperatures, respectively. Data can also be filtered using the Year of acquisition and Day/night filters present in the dashboard. There is an interactive dashboard created using Tableau, and it is uploaded in Tableau Public so that the users can have access to it. [Click Here](https://public.tableau.com/profile/nikhil.reddy8097#!/vizhome/BushfireDashboard_15904216328460/Dashboard1) to have a look at the Dashboard in Tableau Public.

# Website Creation:-

The aspect of creating a website is to locate areas across Australia which are prone to bushfires by including interactive features offered by webpages. In the process of making a website, we used various coding languages such as Python, HTML, and Tableau.

The initial objective of the Webpage

* Locate areas where there are ongoing fires
* Extract data from past fires onto the map
* Including the survival plan which assists you in bushfire situation and emergency hotline numbers to report a fire
* Integrating the visualisation part into the website
* Deploy the website for the general public

**Tools:-**

**UI:-**

To make interactive webpages, we need to use UI, which is a graphical layout of the website. It consists of the buttons users click on, the text they read, the images, sliders, text entry fields, and all the rest of the items the user interacts.

**ATOM:-**

To run the HTML code, we are going to use ATOM which is an open-source text and source editor developed by GitHub. ATOM enables users to install third-party packages and themes to customise the features and looks of the editor, so you can set it up according to your preferences and with ease

**Microsoft Azure:-**

to make the website live we are going to use Microsoft azure which is a public computing platform

To create an interactive webpage, we faced many challenges as it involves different coding languages and platforms of visualisation.

The most prominent hurdle face by the team is to integrate HTML and Python, so after discussion, we concluded to use “**FLASK**” which is a lightweight Web Server Gateway Interface WSGI web application framework. So we used flask to integrate Python with HTML and Tableau with HTML webpage

As there are only a few people in the group who knows how to create a website and coding it was tough to produce a website at the right time

# Source Code:-

The link for the Repository:-

Go through the above link to understand about the project .it contains all the necessary files to run the project

1.the data set we used for the project

2.data visualisation files

3.website

Link for the website:-

# Other relevant documents:-

All the required documents are uploaded in the repository and teams .

# Appendices:-