Proceedings of IYSW, (2020), vol. **9**, pp 339-365.

Journal homepage: http://journals.sdu.edu.kz/index.php/iysw

International Young Scholars' Workshop

Applying machine learning models for predicting forest fires in Australia and the influence of weather on the spread of fires based on satellite and weather forecast data

Bakhtiyor Meraliyev, Kurmangazy Kongratbayev

Department of Computer Science, Suleyman Demirel University, Kaskelen

Abstract

What shall we expect from the year 2020? The coronavirus pandemic is not the worst thing that humanity can face in the near future. According to the observations of the scientists, in March, 2020, the planet temperature warmed up to the record-high level. Also, the temperature of the world's oceans exceeded its average temperatures by 80%, and prognosis of the meteorological observations is not good. The warming seas had already led to catastrophic disaster. The average temperature increase can also lead to hurricanes, drought, invasion of locusts and, the worst, to forest fires. Natural disasters lead to loss of life, destruction of properties and infrastructure, loss of animal natural habitats, displacement of humans. And the results of these all lead to humanitarian catastrophes, including social and economic.

The situations related to the nature are always very serious, as the whole world is involved. This is like butterfly effect, i.e., the natural disaster in Australia affect the economic and ecologic situation in USA and England. Taking the Australia, they faced problem that cannot be avoided. Nevertheless, the world can be prepared and prevent from the huge disasters. The forecasting of forest fires can really be helpful, as well as the inquiry of the weather impact on fires. The current paper is focused on the study of fire forecasting and weather influence on fire. The relevance of the study is important, as the global warming and human caused fires are increasing and there is a trend that Australia's fires became more dangerous and longer lasting. The artificial intelligence, particularly machine learning algorithms, can help to make appropriate calculations and predictions to safe the ecosystem and human lives.

According to the preliminary research results we acquire; in-depth multidimensional analysis confirms almost 100 percent dependence of bushfires on the weather conditions. Using the machine learning algorithms, it would be possible to predict the time and positioning of inflammation source.

Keywords: Machine learning, algorithms, data mining, wildfire prediction, artificial intelligence, analyzing.

Applying machine learning models for predicting forest fires in Australia and the influence of weather on the spread of fires based on satellite and weather forecast data.

Fire and its nature

Almost 50-degree heat and drought in December 2019 led to fires that are visible even from space: 8 million hectares of land were burned, and the flame reaches 70 m in height.

In the end of year 2019 similar texts and headlines dazzled in all newspapers and media. All mass media channels were broadcasting the consequences of bushfires in Australia. That was really unprecedented situation ever. The area of conservation land burnt in bushfires reached 4/3 million hectares. Just imagine the fire effects affected many areas. And here comes the question: What shall we expect from the year 2020? You know the coronavirus pandemic is not the worst thing that humanity can face in the near future.

According to the observations of the scientists, in March, 2020, the planet temperature warmed up to the record-high level. Also, the temperature of the world's oceans exceeded its average temperatures by 80 %, and prognosis of the meteorological observations is not good. The warming seas had already led to catastrophic disaster. The average temperature increase can also lead to hurricanes, drought, invasion of locusts and, the worst – to forest fires. Natural disasters lead to loss of life, destruction of properties and infrastructure, loss of animal natural habitats, displacement of humans. And the results of these all lead to humanitarian catastrophes, including both social and economic.

The current paper describes the phenomenon of natural disaster that happened in Australia in the beginning of the year 2020. The effect of global warming observed in Australia brought the following situations: heat and drought, frequent lightning strikes during thunderstorms, positive dipole of the Indian Ocean, unintentional and deliberate arson, as well as other artificial causes. The disaster faced by Australia is not the problem of a particular region. This is the situation when the whole world shall focus their attention on. It would take a long time for the nature to return to its initial state, as well as expensive for economic resources of the region. It is well known that the Australia's fauna is unique. Among the animal species that can be found only in Australia is marsupial mammals. Unfortunately, in South Australia state, the authorities decided to shoot ten thousand camels as they were dangerous for the citizens. The reason was that due to the fires, camels strayed into herds and wandered into villages, broke down the fences and tried to break into homes while were looking for the water. And these are the superficial problems.

The situations related to the nature are always very serious, as the whole world is involved. This is like butterfly effect, i.e., the natural disaster in Australia affect the economic and ecologic situation in USA and England. For Australia this is a problem that cannot be avoided, but we can be prepared. Within our research work, we found that the humanity can be prepared and can be warned beforehand. The current research work focused on forecasting of forest fires in Australia, as well as the impact of weather on them. We believe that our studies should help humanity be warned and be prepared. And in case if our efforts save at least one life; life of a person or an animal; it will be a great success for us.

As the result of deep multidimensional analysis and preliminary research, it was confirmed almost 100 percent dependence of fires on the weather conditions. And with the help of forecasting these parameters it would be possible to predict the time and the most probable region of Australia where the forest fire can probably pass. At least such predictions can help the lives and economic situation.

Forest fires are accepted as the most uncontrolled and spontaneously spread for the huge areas which leads to partial or complete burning out of vegetation, forest litter, fertile soil layer. The most epic is that the forest fires causes the death of forest inhabitants who have not managed to escape from the fire, mainly these are newly born animals, including some species of rare mammals.

Absolutely all forest fires are extremely dangerous. The fire flashes up very quickly and for the considerable area. And despite the constant monitoring of fire hazard areas the natural disaster cannot be avoided. The most dangerous ones are fires that occur during drought, as they spread for the hundreds of thousands of hectares, destroying settlements and agricultural lands located near the forests.

The forest fires were mainly in the southeast of the country, and they have been ongoing since August 2019, the spring for Australia. Due to the geographical location of Australia, the weather and nature there differs from ours. The scale of fires is much higher during the annual drought season that is usually going from December to March, during the summer period in Australia. It is apparently that the forest fires of the years 2019 - 2020 became the most destructive in the entire history of Australia.

A forest fire is a dynamic phenomenon that changes its behavior in time. The behavior of fire is determined by the complex heat transfer and thermochemical processes. Understanding these can help making predictions of fire behavior in future. Believe that even the preliminary findings can save the nature and economics of the country. The current research work is aimed to use artificial intelligence to make the most reliable predictions of bushfires.

The fire must ripen

The research methods described in this paper are focused on predicting a random process as a fire. Fire depends on many factors like an accidental match, a lightning strike, a piece of a bottle that can work like a magnifying glass in hot weather. Fire is like an uncontrolled uncertainty.

The experts have the expression as ‘The fire must ripen’. Try to explain this expression… There must be appropriate environment for the fire to occur. And by analyzing these factors, we can predict the hotbed of ignition. For the research purposes, it is important to collect as much data as possible on the weather temperature, humidity level, wind direction, and other meteorological indicators. It was decided to create a mechanism that analyses above mentioned indicators. For the purposes of the research work, it was taken the changes in meteorological indicators from December 11, 2019 till January 11, 2020. As the result of the research we shall acquire the report on areas that are ‘ripe’ and where the fire can most probably occur.

There are more than 27 known mathematical models defined to describe the forest fire behavior. Each model was built according to the different experiences in different countries with forest fire. Every model differs from the others according to the input and the environmental parameters. For instance, the fires in Canada have some special causes, whether Russia’s and Australia’s fires’ occur due to other factors.

Earth-orbiting satellites and air-floating devices have been set for the observation and detection of forest fires. Satellite images mainly generated by two satellites that were launched specifically for forest fire detection purposes: the advanced very high-resolution radiometer

(AVHRR) launched in 1998 and the moderate resolution imaging spectroradiometer (MODIS) launched in 1999. These satellites can provide images of the Earth once every 2 days and being frankly, this is a long delay for fire scanning. In addition, the quality of satellite images directly depends on weather conditions.

Bushfire simulators and analysis in Australia

Researchers were tense with helplessness in situations related to bushfires. They tried to use all available and applicable methods to help the Earth avoiding the fires. Researchers and scientists are struggling with fires at the state level in Australia. The research results and experience of Australian scientists have helped in studying the root of the problem and causes of fire. The bushfire website of the Australian Bureau of Meteorology is the best result of researchers' work. During the current research work, we have studied the work of Timothy Neale and Daniel May 'Bushfire simulators and analysis in Australia: insights into an emerging sociotechnical practice' as it was somehow related to our current research work. Above mentioned researchers have prepared the ground for studies of the social dimensions of bushfire prediction by investigating how simulators and predictive practitioners have been mobilized. As the probability of bushfires increases, it is very important to find the answers on issues related to them. And T. Neale and D. May raised several questions that bushfire practitioners, policymakers and researchers will need in order to prevent disaster.

Machine Learning Models

Machine learning tasks are divided into 2 main categories: classification and regression. It was analyzed and implemented five machine learning models during the current research work. And we would like to explain these models in details.

Support Vector Machine (SVM) is a linear algorithm used in classification problems. SVM can also be used as a regression method, maintaining all the main features that characterize the algorithm (maximal margin). The Support Vector Regression (SVR) uses the same principles as the SVM for classification.

K Nearest Neighbors (KNN) is a simple supervised machine learning model that was easy to implement and helped to solve both regression and classification problems. The data in supervised learning is labeled, the feedbacks and features are known. In other words, the KNN is the algorithm that trains function to get the appropriate result according to labeled input data that shall be implemented where new data is unlabeled. The KNN model is good solution as the algorithm assumes that similar parts exist in one place or near it, i.e., similar elements are near to each other.

Gaussian Naive Bayes (NB) is a classification technique based on the theorem with independent predictions. This classifier means that if we have some feature in a class, it will not give us any information about any other feature in the class. Naive Bayes is about conditional probability that something will happen, given that something else has already occurred. This model was used because the algorithm is fast and highly scalable. Usually this method is used for text classification problems, but in this research, we observed that it can be used with numbers as well. However, it considers all features to be unrelated, thus it cannot learn the relationship between features. Naive Bayes can learn importance of individual features, but cannot determine the relationship among features.

Random Forrest (RF) is large collection of data. As in SVM, the tasks in the Random Forrest method are classification and regression. In terms of efficiency, this method competes with SVM and neural networks. There are 2 stages in Random Forrest Algorithm: random forest creation and make a prediction from RF classifier. The prediction is made after forest classifier created in first stage with pseudo code. Take the test features by using the rules of randomly created decision tree to predict the target. Calculate the votes for each predicted target and finally high voted target will be the final prediction.

Decision tree is a graph that uses a branching method to illustrate every possible outcome of a decision. Decision trees are useful for focusing discussion when a group must make a decision. Programmatically, they can be used to assign monetary / time or other values to possible outcomes so that decisions can be automated.

Dataset

The news about fires in Australia forest have been spread fast, however, the same cannot be said about the datasets. The NASA FIRMS MODIS and VIIRS Fire provide an initial dataset, but it was not enough to make any predictions. It was necessary to use some other main parameters of weather, wind, sun hour, speed of wind in kilometers and in miles, pressure and etc. We took the dataset with 180 thousand rows from Earth data (earthdata.nasa.gov).

In [61]:
```
fires = pd.read_csv('datasets/fire_nrt_M6_96619.csv')
fires.head()
```

Out[61]:

	latitude	longitude	brightness	scan	track	acq_date	acq_time	satellite	instrument	confidence	version	bright_t31	frp	daynight
0	-14.281	143.636	323.9	1.7	1.3	2019-10-01	25	Terra	MODIS	70	6.0NRT	302.3	26.8	D
1	-14.284	143.532	343.5	1.7	1.3	2019-10-01	25	Terra	MODIS	90	6.0NRT	306.3	84.3	D
2	-14.302	143.706	320.2	1.7	1.3	2019-10-01	25	Terra	MODIS	30	6.0NRT	305.0	14.1	D
3	-14.283	143.652	320.4	1.7	1.3	2019-10-01	25	Terra	MODIS	57	6.0NRT	303.3	18.4	D
4	-14.285	143.521	349.4	1.7	1.3	2019-10-01	25	Terra	MODIS	94	6.0NRT	304.7	110.7	D

In [62]:
```
fires.describe()
```

Out[62]:

	latitude	longitude	brightness	scan	track	acq_time	confidence	bright_t31	frp
count	183593.000000	183593.000000	183593.000000	183593.000000	183593.000000	183593.000000	183593.000000	183593.000000	183593.000000
mean	-27.100821	141.939281	339.058568	1.602931	1.207766	811.910122	74.989738	303.325399	95.340657
std	8.172289	11.027220	28.605291	0.811106	0.247695	622.267670	25.041968	13.348698	241.045287
min	-43.116000	113.458000	300.000000	1.000000	1.000000	0.000000	0.000000	265.700000	0.000000
25%	-33.109000	131.570000	320.800000	1.000000	1.000000	345.000000	59.000000	293.800000	18.100000
50%	-30.132000	147.884000	334.300000	1.300000	1.100000	515.000000	82.000000	302.000000	36.400000
75%	-17.868000	150.650000	348.600000	1.800000	1.300000	1330.000000	99.000000	311.600000	82.400000
max	-9.387000	153.477000	507.000000	4.800000	2.000000	2355.000000	100.000000	400.100000	11164.100000

Image 1: The dataset "fire nrt M6 96619.csv" from NASA FIRMS MODIS and VIIRS Fire

The dataset “fire nrt M6 96619.csv” (please refer to image 1) contains several columns. Latitude - center of 1km fire pixel, but not necessarily the actual location of the fire as one or more fires can be detected within the 1km pixel. Longitude - center of 1km fire pixel, but not necessarily the actual location of the fire as one or more fires can be detected within the 1km pixel. Brightness temperature - channel 21/22 brightness temperature of the fire pixel measured in Kelvin. Along Track and Scan pixel sizes – the algorithm produces 1km fire pixels, but MODIS pixels get bigger toward the edge of scan. Acquisition Date and Time – Date and time of MODIS acquisition / overpass of the satellite (in UTC). Satellite: A = Aqua and T = Terra. Instrument – Constant value for MODIS. Confidence (0-100%): This value is based on a collection of intermediate algorithm quantities used in the detection process. It is intended to help users gauge the quality of individual hotspot / fire pixels. Confidence estimates range between 0 and 100 % and are assigned one of the three fire classes (low-confidence fire, nominal-confidence fire, or high-confidence fire). Version (Collection and source) – Version identifies the collection (e.g. MODIS Collection 6) and source of data processing: Near Real-Time (NRT suffix added to collection) or Standard Processing (collection only). “6.0NRT” - Collection 6 NRT processing. “6.0” - Collection 6 Standard processing. Brightness temperature 31 – Channel 31 brightness temperature of the fire pixel measured in Kelvin. Fire Radiative Power (frp) – Depicts the pixel-integrated fire radiative power in MW (megawatts). Day / Night: D = Daytime, N = Nighttime.

The information received from satellites is not enough, so we intended to collect the weather forecast data by ourselves. For the purpose of the current research work the coordinates and dates were taken from observation of the satellite dataset. It is important to know that the weather tracker adapters are not available where people do not live. And we have chosen the cities in Australia where there was at least one weather forecast adapter track. The studied time period was from December 11, 2019 to January 11, 2020.

Here is the algorithm of collecting datas:

In [206]:
```
cities = pd.read_excel('cities.xlsx')
cities.head()
```

Out[206]:

	Name	Latitude	Longitude
0	Tamworth, NSW, Australia	-31.083332	150.916672
1	Queanbeyan, NSW, Australia	-35.353333	149.234161
2	Penrith, NSW, Australia	-33.758011	150.705444
3	Newcastle, NSW, Australia	-32.916668	151.750000
4	Liverpool, NSW, Australia	-33.920921	150.923141

In [207]:
```
cities.describe()
```

Out[207]:

	Latitude	Longitude
count	89.000000	89.000000
mean	-32.461098	140.917670
std	5.796728	12.924641
min	-42.880554	115.345833
25%	-35.549999	138.503052
50%	-33.483334	145.981674
75%	-31.747000	150.705444
max	-12.462827	153.399994

In [212]:
```
import requests
listof = []

for index, row in cities.iterrows():
    link = "http://api.worldweatheronline.com/premium/v1/past-weather.ashx?key=api-key&format=json&tp=24&q=" + row.Name + "&date

    response = requests.get(link)
    if response.status_code == 200:
        listof.append(response.json())
```

Image 2: The algorithm of data collection

Weather Dataset

It was crucial to add data on weather. We have already mentioned that the research work is closely interacting with weather and its influence on fires and fire radiation.

As there was lack of necessary information, in order to obtain appropriate results, we have collected some additional information. DewPointC and DewPointF - the atmospheric temperature (varying according to pressure and humidity) below which water droplets begin to condense and dew can form, in Celsius and in Fahrenheit. FeelsLikeC and FeelsLikeF - The Feels Like Temperatures show what the outdoor temperature will feel like for the current day, in Celsius and in Fahrenheit. HeatIndexC and HeatIndexF - The Heat Index is a measure of how hot it really feels when relative humidity is factored in with the actual air temperature, in Celsius and in Fahrenheit.

In [213]:
```
weathers = pd.read_excel('datasets/weatherll.xlsx')
weathers.head()
```

Out[213]:

	DewPointC	DewPointF	FeelsLikeC	FeelsLikeF	HeatIndexC	HeatIndexF	WindChillC	WindChillF	WindGustKmph	WindGustMiles	...	pressure	sunHour	total!
0	7	45	22	71	22	71	21	70	15	9	...	1013	14.5	
1	7	44	21	70	21	70	21	70	10	6	...	1013	14.5	
2	9	47	17	62	18	64	17	62	10	6	...	1013	12.4	
3	8	47	20	68	20	68	20	67	13	8	...	1011	14.5	
4	7	44	20	68	20	68	20	68	15	9	...	1009	14.5	

5 rows × 30 columns

In [214]:
```
weathers.describe()
```

Out[214]:

	DewPointC	DewPointF	FeelsLikeC	FeelsLikeF	HeatIndexC	HeatIndexF	WindChillC	WindChillF	WindGustKmph	WindGustMiles	...	precipMM
count	8272.000000	8272.000000	8272.000000	8272.000000	8272.000000	8272.000000	8272.00000	8272.000000	8272.000000	8272.000000	...	8272.000000
mean	9.908486	49.812863	20.356625	68.639144	20.857471	69.534574	19.89132	67.797510	21.547268	13.385638	...	1.442481
std	5.513481	9.915008	6.815690	12.273600	6.280577	11.311903	6.37749	11.474888	8.506184	5.285785	...	5.813696
min	-7.000000	20.000000	3.000000	37.000000	5.000000	41.000000	3.00000	37.000000	5.000000	3.000000	...	0.000000
25%	6.000000	43.000000	15.000000	59.000000	16.000000	61.000000	15.00000	59.000000	15.000000	9.000000	...	0.000000
50%	9.000000	48.000000	21.000000	69.000000	21.000000	70.000000	20.00000	68.000000	20.000000	13.000000	...	0.000000
75%	13.000000	56.000000	25.000000	78.000000	25.000000	78.000000	24.25000	76.000000	27.000000	17.000000	...	0.500000
max	25.000000	77.000000	42.000000	108.000000	42.000000	108.000000	41.00000	105.000000	62.000000	39.000000	...	231.000000

8 rows × 27 columns

Image 3: Weather database

WindChillC and WindChillF - Wind-chill or windchill (popularly wind chill factor) is the lowering of body temperature due to the passing-flow of lower- temperature air, in Celsius and in Fahrenheit. WindGustKmph and WindGustMile - Wind speed, or wind flow speed, is a fundamental atmospheric quantity caused by air moving from high to low pressure, usually due to changes in temperature, in KM/H and in Mile/H. avgtempC and avgtempF - Average temperature, in Celsius and in Fahrenheit. The city where located current tracker adapter. Cloud cover (also known as cloudiness, cloudage, or cloud amount) refers to the fraction of the sky obscured by clouds when observed from a particular location. Weather data on current date. Humidity is the concentration of water vapor present in the air. Water vapor, the gaseous state of water, is generally invisible to the human eye. maxtempC and maxtempF - Maximum temperature, in Celsius and in Fahrenheit. mintempC and mintempF - Minimum temperature, in Celsius and in Fahrenheit. precipMM - in meteorology, precipitation is any product of the condensation of atmospheric water vapor that falls under gravity from clouds, in Millimeters. By definition, atmospheric or air pressure is the force per unit of area exerted on the Earth's surface by the weight of the air above the surface. How many hours sun will rise, total snow in centimeters, the ultraviolet index or UV Index is an international standard measurement of the strength of sunburn-producing ultraviolet (UV) radiation at a particular place and time. In meteorology, visibility is a measure of the distance at which an object or light can be clearly discerned. The information on weather in current location is also very important and useful. windspeedKmph and windspeedMiles - Wind speed, or wind flow speed, is a fundamental atmospheric quantity caused by air moving from high to low pressure, usually due to changes in temperature, in KM/H and in Miles/H. Latitude and longitude of location of tracker adapter.

Proposed method

Within the current research work, we have focused on and implemented 3 main components of the work process.

1. Monitoring. The information produced from monitoring can help the firefighters to understand the fire behavior such as point of ignition, the spread speed and the direction of maximum spread. These parameters can be used as input for fire simulation program to help extinguishing fire and provide safety to firefighter.

2. Analysis. Predicting fire behavior is an art as much as it's a science. Even seasoned firefighters have trouble reading fire behavior and predicting fire's potential threat to property and lives. When they can't predict, the result may very well the behavior, this can lead to tragedy. Up-to-the-minute satellite mapping and weather information, remote sensing

technologies, data evaluation, computer modelling, and internet communications have changed the face of fire behavior analysis, thus putting the task of fire suppression into a new dimension.

3. Prediction. Applied to generate predictions in real forest fire situations, using historical data both to train the system and to check the results. Results have demonstrated that the system accurately predicts the ability of forest fires. It has been demonstrated that using a distributed architecture enhances the overall performance of the system.

There are approximately 40 000 rows in the data received from satellite within the period of Dec 11, 2019 and Jan 11, 2020. Merging logic: If weather long/lat is nearest/near to long/lat for fire area in satellite dataset, it says that weather matches with fire area (+-1 C). After merging information, the dataset is ready for following processes.

A scatter plot is a type of plot or math diagram using Cartesian coordinates to show values for typically two variables for a set of data. The data are displayed as a collection of points, each having the value of one variable determining the position on the horizontal axis and the value of the other variable determining the position on the vertical axis.

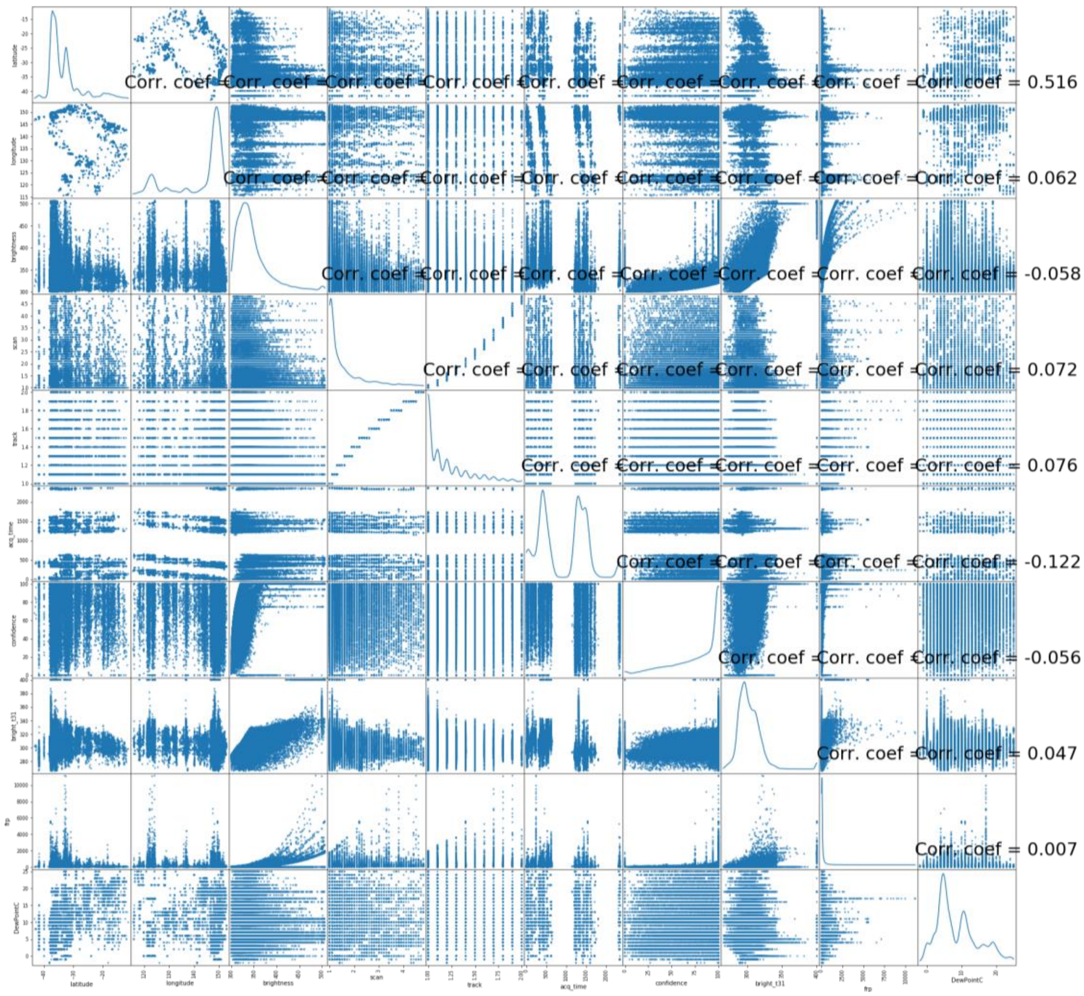

Figure 1: Scatter and Density plot

A correlation matrix is the result of calculating correlations of the same type for each pair of a set of R variables measured on a quantitative scale on a single sample. The main task of the correlation matrix analysis is to identify the structure of relationships of a set of features. In this case, visualization analysis of correlation points is possible – a graphical representation of the structure of statistically significant relationships, as you can mention by the graph of correlation matrix.

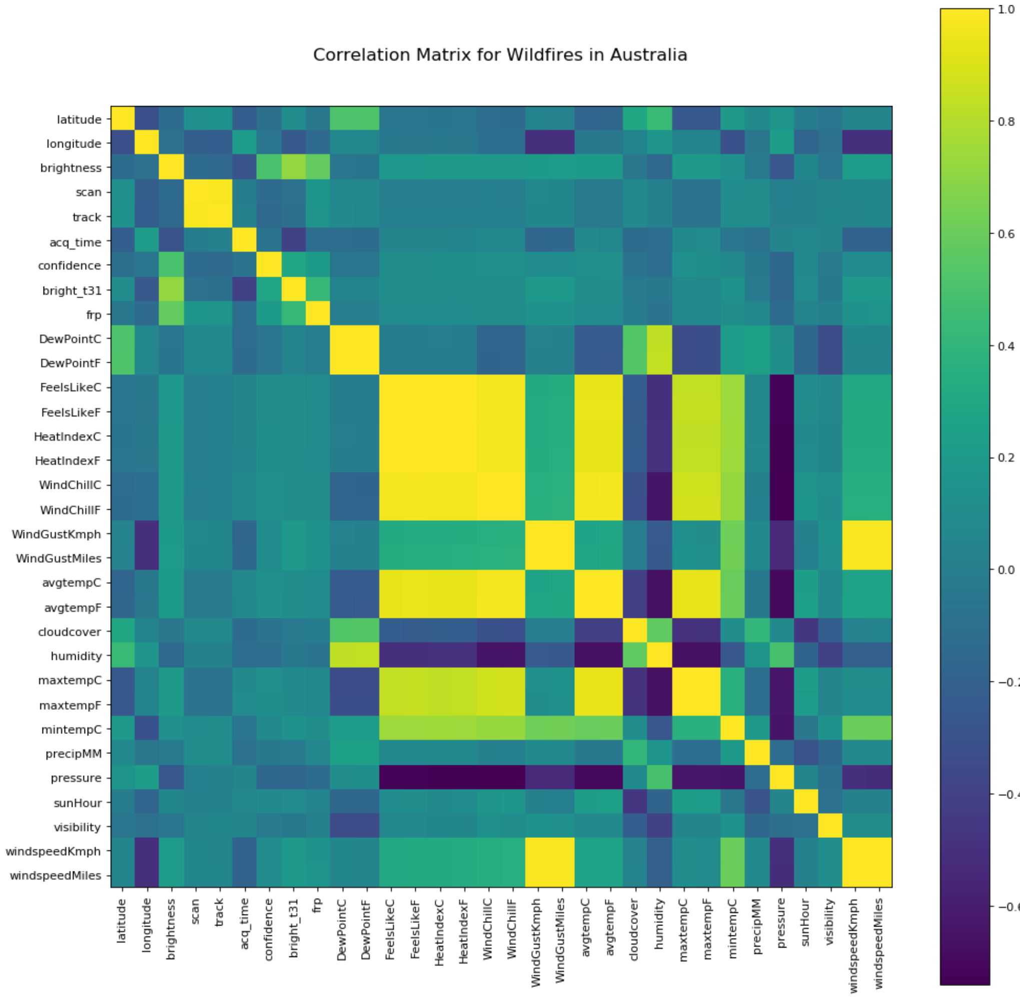

Figure 2: Correlation Matrix of Bushfires in Australia

Results and Discussions

It was considered two experiments during the current research work.

1. In first main experiment, we classified our dataset to identify whether it is fire or not. The shape of our dataset is 31 thousand features. Used features are: brightness, track, FeelsLikeC, HeatIndexC, WindGustKmph, avgtempC, maxtempC, mintempC, precipMM, windspeedKmph, pressure, humidity, DewPointC. According to satellite and weather forecast information, the data is classified as fire or not fire.

- Use Scikit Learn algorithms for doing classification the dataset to create the models.
- Find Root Mean Square Error and accuracy for every model.

- Plot ‘Relative performance of various models’ (REC curves).

2. In the second experiment, we are considering the influence of weather on the spread of fires. According to satellite and weather forecast information, the data is classified to predict FIRE RADIATIVE POWER of fire place with the probability of fire occurrence for 100 %.

- Use Scikit Learn algorithms for doing classification the dataset.
- Find Root Mean Square Error and accuracy for every model.
- Plot ‘Relative performance of various models’ (REC curves).

Regression Error Characteristic (REC) estimation

Receiver Operating Characteristic (ROC) curves provide a powerful tool for visualizing and comparing classification results. Regression Error Characteristic (REC) curves generalize ROC curves to regression. REC curves plot the error tolerance on the x-axis versus the percentage of points predicted within the tolerance on the y-axis. The resulting curve estimates the cumulative distribution function of the error. The REC curve visually presents commonly-used statistics. The area-over-the-curve (AOC) is a biased estimate of the expected error. The R2 value can be estimated using the ratio of the AOC for a given model to the AOC for the null-model. It is possible to quickly assess the relative merits of many regression functions by examining the relative position of their REC curves. The shape of the curve reveals additional information that can be used to guide modeling.

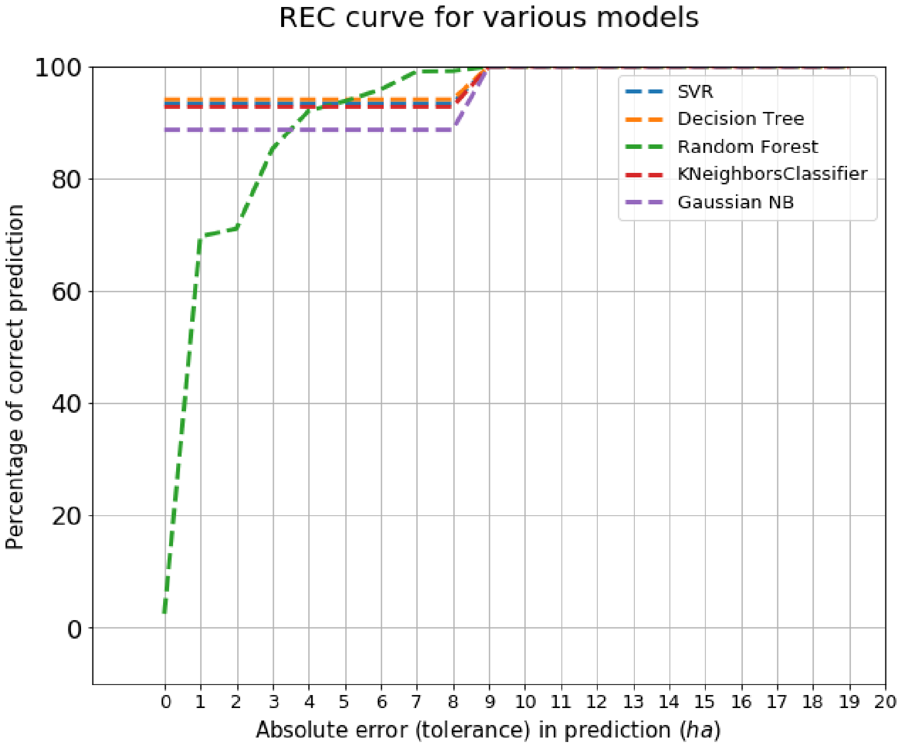

Figure 3: REC curve for various models (experiment 1)

The current research work focuses on comparison between models that shows the differences between them. It was considered five models and compared the accuracy of their predictions:

- Support Vector Regressor with 93.4% accuracy (root error is 25%);
- Decision Tree Regressor with 94% accuracy (root error is 24%);
- Random Forest Regressor with 76% accuracy (root error is 21%);
- KNeighborsClassifier with 92.8% accuracy (root error is 26%);
- Gaussian NB with 88.7% accuracy (root error is 33%).

It is obvious that the most accurate is the Decision Tree Repressor Method with the 94 % accuracy.

Fire Radiative Power

```
In [331]: frp = real[ (real['conf'] == 1)]
          frp.shape

Out[331]: (28374, 43)

In [332]: def convert_frp(frp):
              if frp <= 50.0:
                  return 0
              elif (frp > 50.0) & (frp <= 100.0):
                  return 1
              elif (frp > 100.0) & (frp <= 150.0):
                  return 2
              elif (frp > 150.0) & (frp <= 200.0):
                  return 3
              elif (frp > 200.0) & (frp <= 250.0):
                  return 4
              elif (frp > 250.0) & (frp <= 300.0):
                  return 5
              elif (frp > 300.0) & (frp <= 350.0):
                  return 6
              elif (frp > 350.0) & (frp <= 400.0):
                  return 7
              elif (frp > 400.0) & (frp <= 450.0):
                  return 8
              return 9
```

Image 4: Fire Radiative Power

Fire Radiative Power (FRP) is the rate of emitted radiative energy by the fire at the time of the observation, measured in megawatts. The main issue of the second experiment we consider – 'Is there any influence of weather conditions on the spread of fires?'. To be able to answer this question, we divided the parameter of FRP into some intervals, to make the categorization (minimum 50, maximum 500). It is useful to group the data as it is impossible to use regression and consequently the results of the experiment.

KNeighborsClassifier

```
In [360]: classifier = KNeighborsClassifier(n_neighbors = 7)
          classifier.fit(X_train, y_train)
          predicted = classifier.predict(X_test)

          print('Accuracy: ', accuracy_score(predicted, y_test))
          print("RMSE for KNeighborsClassifier:",np.sqrt(np.mean((y_test-predicted)**2)))

          Accuracy:  0.7011629272876777
          RMSE for Decision Tree: 1.8635466981679825
```

Image 5: Accuracy of KNeighborsClassifier

If the accuracy of prediction is less than 50 %, it means that there is no influence of weather on the fires. According to the data of weather, the accuracy result we got makes 70 %. Thus, we can conclude that there was influence of weather on bushfires in Australia in the beginning of 2020.

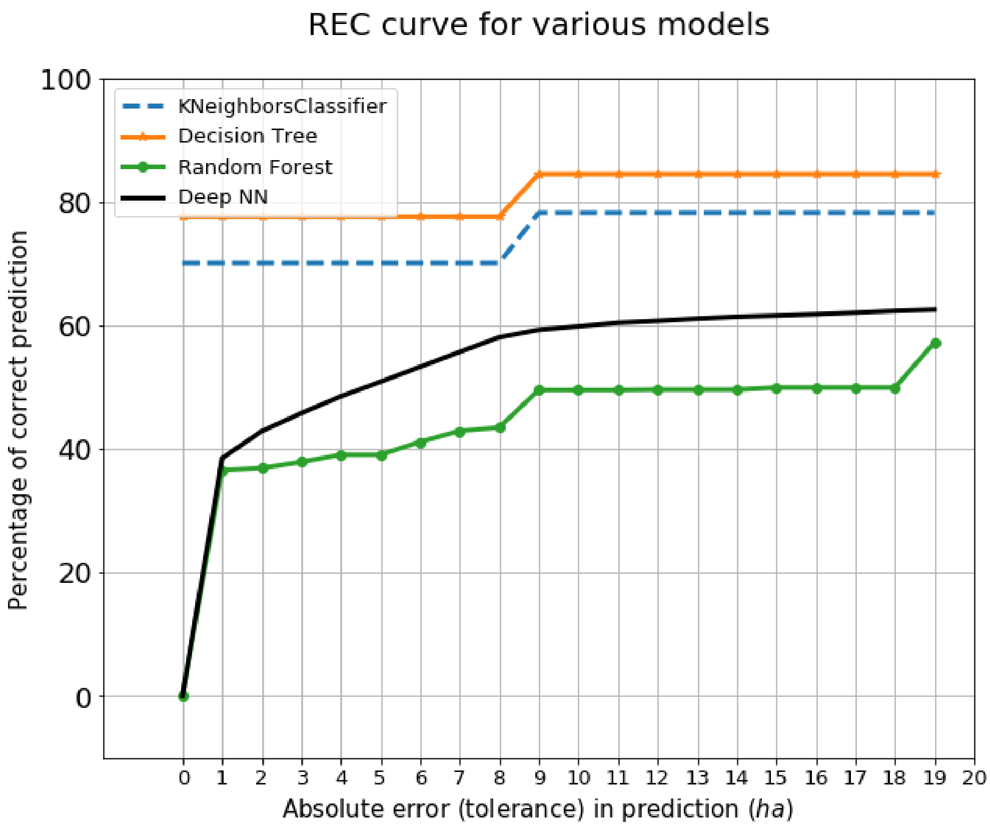

Figure 4: REC curve for various models (experiment 2)

Conclusion

Machine Learning models and algorithms can be applied for predicting the forest fires by using data obtained from satellite and weather forecast sources. The results of these analysis and monitoring shall be used by firefighters and emergency authorities for prediction the bushfires. We believe that these measures will prevent from destructive disaster and minimize the number of victims and value of damages. The aim of the current paper was to emphasize the role of artificial intelligence in the fire fighting. There are researchers and scientists who have already started deep studying of this issue and we believe that the relevance of the problem will push young scientists to consider bushfires prediction. Surely parameters and features will be broader.

One of the most important conclusions was that with the help of methods and algorithms of Machine Learning we can understand the tendency and use results in our everyday life. These tools can also be implemented for detection of center of ignition and the fire radiative power. The timely acquired information can save human lives, animals and avoid ecological and economic disasters. The regular information like weather forecast, wind speed, UV rays, humidity can help to prevent from the disasters.

Every single research work brings to better understanding on model and method that shall be used for bushfire problem solving. With the development in this issue, the scientists and researchers succeed in predicting the outcomes more accurately than the previous ones. The more additional data parameters and features are used, the more accurately results acquired. During the current work we got the accuracy level of 94 %. And we believe that further the accuracy level can achieve the result of 95 % accuracy.

References

Paradis C. (2020). Fires from Space: Australia. *NASA Satellite Data MODISC6 and VIIRS 375m from 2019-08-01 to 2020-01-11*

Resnick B., Irfan U., Samuel S. (2020). 8 things everyone should know about Australia's wildfire disaster. Retrieved from: www.vox.com

Sayad Y. O., Mousannif H., Al Moatassime H. (2019). Predictive modeling of wildfires: A new dataset and machine learning approach. *Fire Safety Journal*, Pages 130 – 146.

Jaafari A., Pourghasemi H. R. (2019). Factors Influencing Regional-Scale Wildfire Probability in Iran: An Application of Random Forest and Support Vector Machine.

Breiman L. (2001). Random forests. Machine Learning. 45(1), 5 - 32.

Granwal L. (2020). Bushfires in Australia - Statistics & Facts. *Statista*. Retrieved from https://www.statista.com/topics/6125/bushfires-in-australia/

Demenko S. (2019). Оракул огня. Компьютер прогнозирует возникновение пожара. *Российская газета - Федеральный выпуск № 178(7936).* Retrieved from https://rg.ru/2019/08/13/reg-sibfo/kak-rabotaet-sistema-prognozirovaniia-pozharov.html

Images title:

In [61]:
```python
fires = pd.read_csv('datasets/fire_nrt_M6_96619.csv')
fires.head()
```

Out[61]:

	latitude	longitude	brightness	scan	track	acq_date	acq_time	satellite	instrument	confidence	version	bright_t31	frp	daynight
0	-14.281	143.636	323.9	1.7	1.3	2019-10-01	25	Terra	MODIS	70	6.0NRT	302.3	26.8	D
1	-14.284	143.532	343.5	1.7	1.3	2019-10-01	25	Terra	MODIS	90	6.0NRT	306.3	84.3	D
2	-14.302	143.706	320.2	1.7	1.3	2019-10-01	25	Terra	MODIS	30	6.0NRT	305.0	14.1	D
3	-14.283	143.652	320.4	1.7	1.3	2019-10-01	25	Terra	MODIS	57	6.0NRT	303.3	18.4	D
4	-14.285	143.521	349.4	1.7	1.3	2019-10-01	25	Terra	MODIS	94	6.0NRT	304.7	110.7	D

In [62]:
```python
fires.describe()
```

Out[62]:

	latitude	longitude	brightness	scan	track	acq_time	confidence	bright_t31	frp
count	183593.000000	183593.000000	183593.000000	183593.000000	183593.000000	183593.000000	183593.000000	183593.000000	183593.000000
mean	-27.100821	141.939281	339.058568	1.602931	1.207766	811.910122	74.989738	303.325399	95.340657
std	8.172289	11.027220	28.605291	0.811106	0.247695	622.267670	25.041968	13.348698	241.045287
min	-43.116000	113.458000	300.000000	1.000000	1.000000	0.000000	0.000000	265.700000	0.000000
25%	-33.109000	131.570000	320.800000	1.000000	1.000000	345.000000	59.000000	293.800000	18.100000
50%	-30.132000	147.884000	334.300000	1.300000	1.100000	515.000000	82.000000	302.000000	36.400000
75%	-17.868000	150.650000	348.600000	1.800000	1.300000	1330.000000	99.000000	311.600000	82.400000
max	-9.387000	153.477000	507.000000	4.800000	2.000000	2355.000000	100.000000	400.100000	11164.100000

Image 1. The dataset “fire nrt M6 96619.csv

Here is the algorithm of collecting datas:

In [206]:
```python
cities = pd.read_excel('cities.xlsx')
cities.head()
```

Out[206]:

	Name	Latitude	Longitude
0	Tamworth, NSW, Australia	-31.083332	150.916672
1	Queanbeyan, NSW, Australia	-35.353333	149.234161
2	Penrith, NSW, Australia	-33.758011	150.705444
3	Newcastle, NSW, Australia	-32.916668	151.750000
4	Liverpool, NSW, Australia	-33.920921	150.923141

In [207]:
```python
cities.describe()
```

Out[207]:

	Latitude	Longitude
count	89.000000	89.000000
mean	-32.461098	140.917670
std	5.796728	12.924641
min	-42.880554	115.345833
25%	-35.549999	138.503052
50%	-33.483334	145.981674
75%	-31.747000	150.705444
max	-12.462827	153.399994

In [212]:
```python
import requests
listof = []

for index, row in cities.iterrows():
    link = "http://api.worldweatheronline.com/premium/v1/past-weather.ashx?key=api-key&format=json&tp=24&q=" + row.Name + "&date

    response = requests.get(link)
    if response.status_code == 200:
        listof.append(response.json())
```

Image 2. The algorithm of collecting the data

```
In [213]: weathers = pd.read_excel('datasets/weatherll.xlsx')
          weathers.head()
Out[213]:
```

	DewPointC	DewPointF	FeelsLikeC	FeelsLikeF	HeatIndexC	HeatIndexF	WindChillC	WindChillF	WindGustKmph	WindGustMiles	...	pressure	sunHour	total!
0	7	45	22	71	22	71	21	70	15	9	...	1013	14.5	
1	7	44	21	70	21	70	21	70	10	6	...	1013	14.5	
2	9	47	17	62	18	64	17	62	10	6	...	1013	12.4	
3	8	47	20	68	20	68	20	67	13	8	...	1011	14.5	
4	7	44	20	68	20	68	20	68	15	9	...	1009	14.5	

5 rows × 30 columns

```
In [214]: weathers.describe()
Out[214]:
```

	DewPointC	DewPointF	FeelsLikeC	FeelsLikeF	HeatIndexC	HeatIndexF	WindChillC	WindChillF	WindGustKmph	WindGustMiles	...	precipMM
count	8272.000000	8272.000000	8272.000000	8272.000000	8272.000000	8272.000000	8272.00000	8272.000000	8272.000000	8272.000000	...	8272.000000
mean	9.908486	49.812863	20.356625	68.639144	20.857471	69.534574	19.89132	67.797510	21.547268	13.385638	...	1.442481
std	5.513481	9.915008	6.815690	12.273600	6.280577	11.311903	6.37749	11.474888	8.506184	5.285785	...	5.813696
min	-7.000000	20.000000	3.000000	37.000000	5.000000	41.000000	3.00000	37.000000	5.000000	3.000000	...	0.000000
25%	6.000000	43.000000	15.000000	59.000000	16.000000	61.000000	15.00000	59.000000	15.000000	9.000000	...	0.000000
50%	9.000000	48.000000	21.000000	69.000000	21.000000	70.000000	20.00000	68.000000	20.000000	13.000000	...	0.000000
75%	13.000000	56.000000	25.000000	78.000000	25.000000	78.000000	24.25000	76.000000	27.000000	17.000000	...	0.500000
max	25.000000	77.000000	42.000000	108.000000	42.000000	108.000000	41.00000	105.000000	62.000000	39.000000	...	231.000000

8 rows × 27 columns

Image 3. Weather database

```
In [331]: frp = real[ (real['conf'] == 1)]
          frp.shape
Out[331]: (28374, 43)
```

```
In [332]: def convert_frp(frp):
              if frp <= 50.0:
                  return 0
              elif (frp > 50.0) & (frp <= 100.0):
                  return 1
              elif (frp > 100.0) & (frp <= 150.0):
                  return 2
              elif (frp > 150.0) & (frp <= 200.0):
                  return 3
              elif (frp > 200.0) & (frp <= 250.0):
                  return 4
              elif (frp > 250.0) & (frp <= 300.0):
                  return 5
              elif (frp > 300.0) & (frp <= 350.0):
                  return 6
              elif (frp > 350.0) & (frp <= 400.0):
                  return 7
              elif (frp > 400.0) & (frp <= 450.0):
                  return 8
              return 9
```

KNeighborsClassifier

```
In [360]: classifier = KNeighborsClassifier(n_neighbors = 7)
          classifier.fit(X_train, y_train)
          predicted = classifier.predict(X_test)

          print('Accuracy: ', accuracy_score(predicted, y_test))
          print("RMSE for KNeighborsClassifier:",np.sqrt(np.mean((y_test-predicted)**2)))

          Accuracy:  0.7011629272876777
          RMSE for Decision Tree: 1.8635466981679825
```

Image 4. Fire Radiative Power

Image 5. Accuracy of KNeighborsClassifier

Figures title:

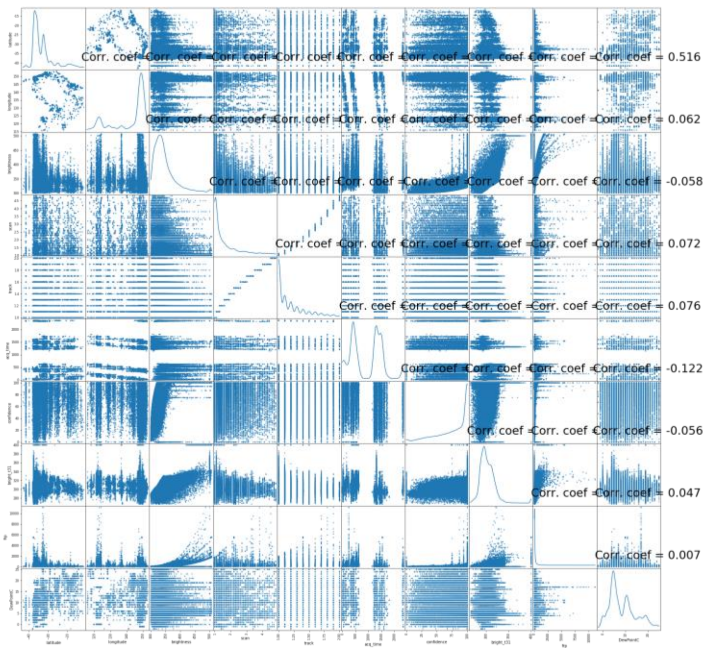

Figure 1. Scatter and Density plot

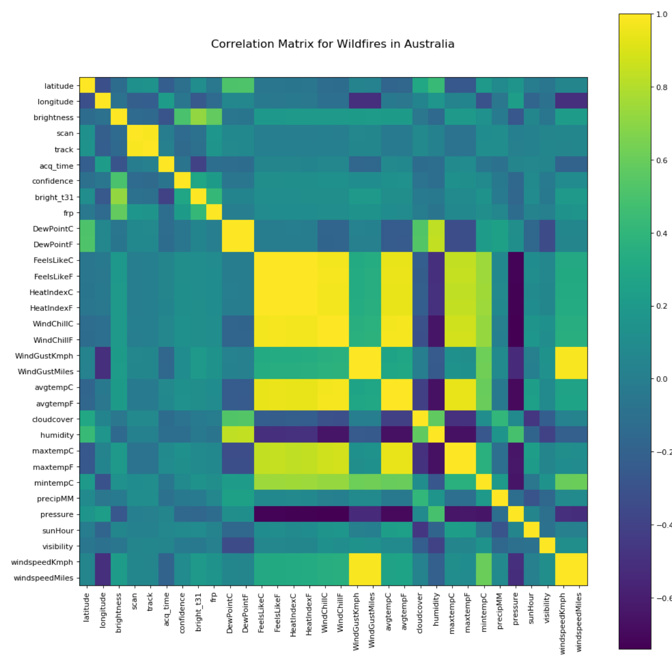

Figure 2. Correlation Matrix of Bushfires in Australia

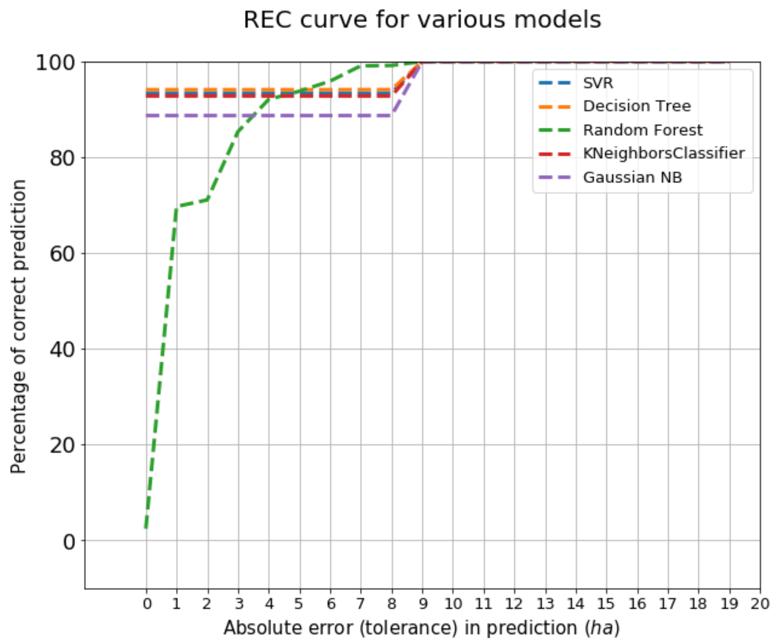

Figure 3. REC curve for various models (experiment 1)

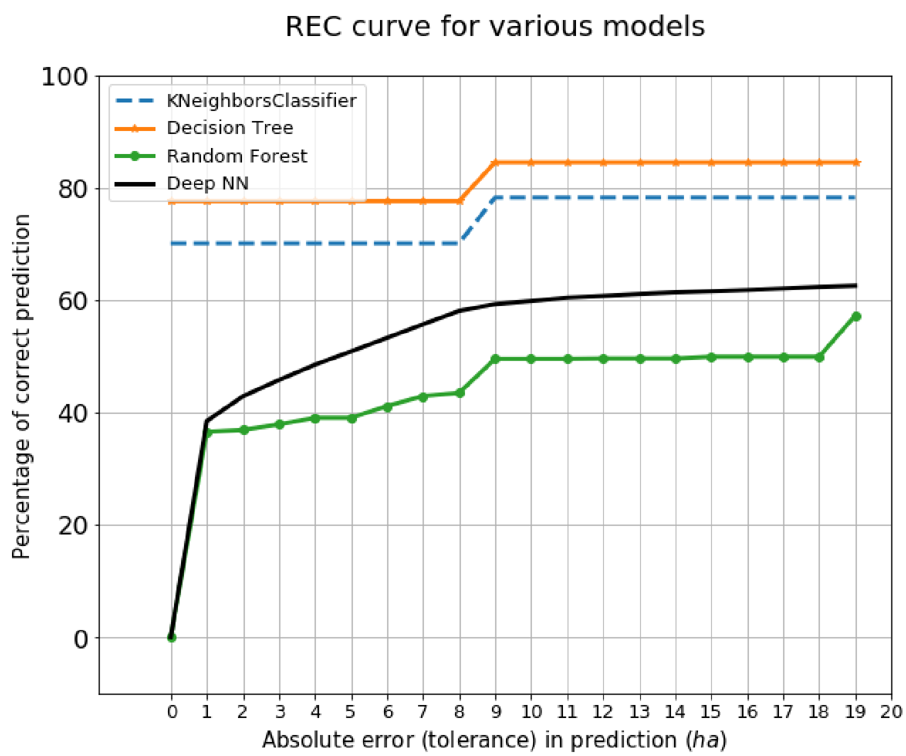

Figure 4. REC curve for various models (experiment 2)