Classification of Wildfire Spread Severity using Machine Learning Algorithm

MSc Research Project
Data Analytics

Amit Sahoo
X18188851

School of Computing
National College of Ireland

Supervisor: Dr. Catherine Mulwa

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Mr Amit Kumar Sahoo		
Student ID:	X18188851		
Programme:	Data Analytics	**Year:**	2019-2020
Module:	MSc Research Project		
Supervisor:	Dr. Catherine Mulwa		
Submission Due Date:	17th August 2020		
Project Title:	Classification of wildfire spread severity using machine learning algorithms		
Word Count:	9384	**Page Count**	29.

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: Amit Kumar Sahoo

Date: 17Aug 2020

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Classification of Wildfire Spread Severity using Machine Learning Algorithm

Amit Sahoo
x18188851

Abstract

Wildfire is one of the natural disasters, that can burn millions of acres of land at a very fast speed, almost burning everything that comes in the way. However, only a few of the wildfire occurs on their own, while majority are human caused. In this research, the size of the fire spread has been predicted, with respect to the weather details of the last five days of the outbreak. This research will help the Fire Fighting Department and the local governing body to predict the fire spread in advance and make decisions accordingly. Alaska location has been chosen specifically, for this research as there is a huge difference in temperature in summer and winter. Data has been collected from various sources and have been merged. At every stage of pre-processing, a Logistic Regression has been used as a baseline model. The technique that produces the highest accuracy has been carried forward to the next stage. Several Machine Learning algorithms have been performed, and it is observed that Artificial Neural Network, outperforms the other tree-based algorithms, ensembled algorithms and LSTM with an accuracy of 68%.

1 Introduction

Various countries have been battling with forest fires for a long time now. Such unfortunate incidents take place due to natural reasons like volcanic eruptions and lightning but mostly happens because of human activities like unextinguished cigarette butts, fire camping, and garbage deposit burning. Such incidents have become quite frequent over the last few years. As a result, a lot of flora and fauna species are getting disappeared, natural habitat and food chains are getting imbalanced. It also emits a lot of carbon emission to the environment, which leads to a green-house effect and climate change, soil erosion. Therefore, it will be useful if the wildfire spread can be predicted at its initial stage and classified as per the National Wildfire Coordinating Group's fire classification chart[1].

1.1 Motivation and Project Background

Several ways to prevent such wildfires and minimize the loss due to high voltage overhead electric powerlines in the Mediterranean forest have been proposed (Martinez-Canales, 1997).

[1] https://www.nwcg.gov/term/glossary/size-class-of-fire

A few of the previous studies on fire detection at the early stages is done with the help of a real-time TV camera (Cappellini, Mattii, and Mecocci, 1989). These camera helps to detect smoke during the day and flames during night. One more significant study on the detection of fire incidents in the boreal forest of Alaska (Bourgeau-Chavez, Kasischke, and French, 1993) using ERS-1 C-VV SAR imagery. It also determined the fire appearance based on elapsed time, geomorphology of the area under fire, and metrological parameters. Multi-dimensional satellite images are used for detecting forest fire in the eastern region of Russia (Kawano, Kudoh, and Makino, 1999).

The likelihood of severe fires is determined from fire indices (Minardi, Marchisio and Treder, 1999), which are extracted through ground and remote satellite observation. These indices are collected for a period of seven years and are interpolated across various grids of The United States to determine the density of the fire occurrence. Information about the land cover was derived by comparing the land cover maps with SPOT satellite images dataset (Lee et al., 1999). After superimposing the burn scar on a digitized land cover, it was evident that fire occurs mostly in easily accessible areas such as agricultural lands and plantation areas. Surveys were conducted to better understand the effect of fire and the characteristics of the fire signatures with the help of three remote sensing systems, operating in two different spectrums (French, Kasischke, and Bourgeau-Chavez, 1994).Geographical information system (GIS) is integrated with the Evidential Belief Function (EBF) to determine the probability of wildfire (Nami et al., 2018). Based on the derived probability, areas were classified into moderate, high, and very high zones. Forest department in few countries uses National Forest Fire Danger Rating System (NFFDRS) for predicting the severity of the fire.

For the last few years, there has been a lot of studies conducted in this field using Machine Learning (ML) and Deep Learning (DL). ML algorithms such as Logistic regression has been used to find the probability of lightning causing a sustainable ignition as well as the probability of the ignition being detected by the firefighting department (Wotton and Martell, 2005). Since most of the fires are caused by humans, a logistic generalized additive model (Vilar et al., 2010) is designed to find out the probability of ignition at a 1km^2 grid. DL has proved to be better in terms of accuracy when dealing with a huge amount of data for training and validation. Research by (Safi, Bouroumi and Bouroumi, 2011), was one of the oldest studies where the Neural Network (NN) approach is applied in predicting the size of the forest fire. DL models such as Long Short-term Memory (LSTM) has been used to predict the size of the fire at the beginning of its occurrence (Liang, Zhang, and Wang, 2019). This model has been implemented on a time series data, which was able to predict the occurrence trends with an accuracy of 90.9%. In this research, we have taken a step further in classifying the wildfire spread by using meteorological parameters of the past four days. ML models are built on these input features, considering them as independent attributes. A DL model has also been built which considers the sequence of weather parameters for the past five days to classify the severity of the fire spread.

1.2 Research Question

Prediction of forest fire is helpful for the local governing bodies and the fire-fighting department, so that adequate resources can be arranged before it gets too difficult to handle such untoward incidents.

> RQ: "*How well can classification (Logistic Regression, KNN, RF, NN, LSTM, Ensemble models) of forest fire spread enhance the prediction of fire spread (small, large or severity) to support Alaska fire department improve in saving life ?*"

Based on weather details of the past 5 days of fire incident, the fire spread severity can be predicted.

> Sub RQ: ''*Can a sequence of weather parameters (such as temperature, sea level pressure, dew point, precipitation, visibility, wind speed, preparedness level of the firefighting department and the number of fire incident) compared to the same features when considered independently for the past four days improve prediction of fire spread severity?*"

To solve the research question and sub question the following objectives were implemented:

1.3 Research Objective and Contribution

Obj1: Critical review of the literature on prediction of wildfire (2005-2020).
Obj2: Extracting the weather details of the day of fire outbreak and previous four days.
Obj3: Merging the fire dataset and the weather dataset and conducting initial pre-processing.
Obj4: Implementation of feature selection methods to determine the relevant predictors.
Obj5: Implementation of dimension reduction to reduce the number of dimensions.
Obj6: Implementation of SMOTE oversampling to handle the imbalanced data.
Ojb7: Implementation of classification of forest fire spread to enhance the prediction (for the main research question)

- Obj7_1: Implement and evaluate results of Decision Tree.
- Obj7_2: Implement and evaluate results of Bagging (Ensemble Technique).
- Obj7_3: Implement and evaluate results of K-Nearest Neighbour.
- Obj7_4: Implement and evaluate results of Random Forest.
- Obj7_5: Implement and evaluate the results of SVM Classifier.
- Obj7_6: Implement and evaluate results of Artificial Neural Network.

Obj8: Implementation of LSTM model using a sequence of weather parameters (for sub-research question).
Obj9: Comparison of developed model.

Contributions: The major contribution resulting from this project is classification and prediction models for forest fire. Accurately predicting the size of the forest fire at an early stage will help the firefighting team to strategize fire extinguishing operations, positioning the crew member and equipment in the best possible position. This can reduce the spread of the fire and help in saving the life of the local people, and the firefighters deployed on the site. This research will also help the local governing body to formulate rescue or evacuation operations for the people under the possible threat of fire. This will assist in granting permission to carry out the local business at the fire vicinity, restrict the tourist from entering the fire zone, and speculating the possible amount of carbon emission into the environment.

Section 2 provides a literature review on this field of study using fire indices and ML with different algorithms. Based on the findings from the previous section, Section-3 provides an overview of the methodology that has been used for this research. Data acquisition for this research, along with the pre-processing, data modelling done in section 4. In section 5, various algorithm that has been implemented and their corresponding evaluation has been explained. Section-6 explains the comparison of all the models used in the previous section based on a few evaluation parameters. Section 6 concludes this paper with the scope of future works.

2 Related Work

In this section, investigation of few of the already used models for predicting the size of a forest fire has been done. This section is further divided into subsections i.e. (1) Correlation between weather and wildfire occurrence (2) Predicting the size of a forest fire from fire indices and identified gaps (3) Review of ML in Predicting the Forest Fire Size and identified gaps (4) Review of NN in predicting the size of the forest fire and identified gaps.

2.1 Correlation between Weather and Wildfire Occurrence

A lot of research has been done to prove that wildfires and climate are highly correlated. A long-term association was proved between fire and weather (Koutsias et al., 2013) by analysing the data which spans over period of more than a century (1984-2010). The analysis revealed that statistically significant growth in fire occurrences have occurred after 1970. The number of fire occurrences is highly correlated with maximum air temperature and negatively correlated with precipitation. Thus, it establishes the fact that weather has a profound effect on the fire spread by directly controlling fuel moisture.

2.2 Predicting the Size of a Fire from Fire-Indices and Identified Gaps

The Canadian Fire Weather Index (FWI) is one the most popular and widely used tool to predict the severity of a fire hazard. It is one of the subsystems of Canadian Forest Fire danger (CFFDRS). FWI provides information on fire weather information, fuel moisture codes, and fire behaviour indexes (Vetrita *et al.*, 2012). An adaption of FWI is proposed by considering Mediterranean vegetation and climate (Chelli et al., 2015). The indexes are measured by fitting the collected data on fuel moisture content and the other values, as per the expected inputs for Canadian FWI. The results obtained can describe the fuel moisture dynamics despite small sample areas and time constraints for data collection. Another such index used for fire danger estimation is the McArthur Forest Fire Danger Index (FFDI). This index cannot be used in few regions due to the unavailability of instruments and human resources. This index is modified by introducing a Normalized Multi-brand Drought Index (NMDI) and achieved an overall accuracy of 82% (Suresh Babu *et al.*, 2017). Various fire danger rating system is being used in different part of the world, FFDI is used in Australia (McArthur et al. 1967), Forest Fire Behaviour Table (FFBT) in western Australia (Beggs, 1976), FWI in Canada (Wagner, 1987), National Fire Danger Rating System(NFDRS) in USA (Bradshaw *et al.*, 1984) and Nesterov Fire Danger Index System in Russia (V. G Nesterov, 1949). Most of these fire danger rating systems only use four weather parameters such as temperature, relative humidity, precipitation, and wind speed. However other factors such as topography, visibility and fuel properties, preparedness level of the fire-fighting department is assumed to be constant. Few of the other fire rating system uses ground data on a regular basis, which is not feasible all the time. Using machine learning techniques to tackle such problems can be a better solution.

2.3 Review of ML in Predicting Forest Fire Size and Identified gaps

Various machine learning algorithms such as Logistic regression, Support Vector Machine (SVM), Decision Tree, Random forest to classify a fire or a no-fire have been proposed (Molovtsev and Sineva, 2019) using a one of the most popular data source present in UCI repository i.e. data from The National park of Monteshino located in the northern part of Portugal. All these models for binary classification were evaluated with the help of a confusion matrix. RF algorithm proved to be superior among all the other models with respect to Recall, Precision, F1-Score, Accuracy. On the same dataset, the prediction on the fire size is also done (Cortez and Morais, 2007) using various data mining techniques such as Multiple Regression, DT, RF, NN and SVM on four different features i.e. spatial, temporal, FWI components, weather attributes. Low-cost real-time data is used instead of the satellite and scanner approaches. SVM proved to be the best among all other data mining techniques in terms of MAD, RMSE, and REC curve. However, the model has a very low predictive accuracy for large fires (> 1 hectare). Linear regression proved to be the most efficient machine learning technique in predicting the size of the forest fire with taking the area burnt as a dependent variable and the meteorological factors as a predictor (Afifuddin et al., 2019). Predictors such as high temperature and low humidity are found to be statistically significant in finding the exact amount of area burnt (Zhang, 2018). Pattern analysis has also been performed in the same research. However, the fire size usually depends on not only the current day but also on the previous days. Therefore, considering the previous weather details may produce more accurate results.

2.4 Review of NN in Predicting Forest Fire Size and Identified gaps

With the advancement of Neural Network over the last decade, the application is seen even in the field of natural disasters. Detection of forest fire has been done (Novac *et al.*, 2020) through an RGB-D enabled quadcopter which moves in a snake-like pattern and sweeps the targeted area. Once it senses the fire, it extracts the blob, and the same is then fed to a Deep Convolution Neural Network (DCNN). However, with this technique, only detection for fire can be done, but enough data cannot be extracted for the prediction of the fire size. Feedforward Neural network technique had been applied in predicting the size of a forest (Safi and Bouroumi, 2013). Hyperparameter such as the number of neurons in a layer and the number of hidden layers is determined heuristically. After a lot of hyperparameter fine-tuning, the architecture was finalized with one hidden layer, 36 neurons, and 10,000 iterations. Since optimizing such hyperparameters can take a lot of effort and calculation, a different approach was applied (Anshori *et al.*, 2019) called Extreme learning machine (ELM) which removed the way of learning from a slow gradient-based algorithm as well as the repetitive task of hyperparameter tuning. This technique uses a single hidden layer and is more stable in terms of accuracy and relatively faster in computation as compared to neural networks with backpropagation. ELM training includes several steps like initializing weights and bias, determining the output from the hidden layer, applying the required activation function, calculating the Moore-Penrose

Pseudo Inverse Matrix, deriving the output weights, and then the predicted results. Study (Liang, Zhang, and Wang, 2019) on the dataset collected from Canada National Fire Database (CNFDB) containing the wildfire and meteorological data for Alberta, Canada. Multi-collinearity and feature normalization are performed on the data and then fed to various deep learning models such as Backpropagation Neural Network (BPNN), Recurrent Neural Network (RNN), and Long-Short Term Memory (LSTM) for predicting the fire size. LSTM showed the best accuracy rate of 90.9%.

The result of the reviewed literature shows that creating a sequence of weather details for the previous four days of the fire incidents can better classifies the fire spread severity since the model memorizes the change in weather details for the previous days. This concludes Objective1 (mentioned in section 1.3).

3 Research Methodology and Design Specification

A data mining research is usually carried out in either KDD (Knowledge Discovery in Databases) or CRISP-DM (Cross-Industry Standard Process for Data Mining) methodology. CRISP-DM focuses on understanding the objective from a business perspective and then converts this knowledge into data mining problem designs a preliminary plan to achieve those objectives (Azevedo and Santos, 2008). KDD method has been used in this research as it extracts what is deemed knowledge using a selected dataset with pre-processing, transformation, data mining, and evaluation.

3.1 Alaska Fire Forest Methodology Approach

This modified KDD methodology for predicting the size of the forest fire consists of following stages (1) Data selection for forest fire in CSV format from Alaska Interagency Coordination Center (AICC) (2) Data selection for historic weather details for the day in which the fire incident takes place and the previous 4 days from The Old Farmer's Almanac[2] in form of CSV (3) Pre-processing and normalizing the input vectors and selecting the most important features (4) Reducing the dimension of the input features (5) Applying oversampling through SMOTE to make the data balanced (6) Models such as KNN, ANN, RNN, DT, SVM, RF, Ensembled Methods are trained on the training data (7) Models were evaluated based on few parameters.

3.2 Data Preparation Process Flow

The data preparation process (depicted in Figure1) consists of the following steps: (1) Downloading the fire data from AICC which contains the number of fire occurred, the total area impacted preparedness level and the date in which the fire incident was reported in form of a CSV. (2) This fire reported date is used to get the weather details from 'The Old Farmer's Almanac' website for that date as well as the previous four days. (3) The data is extracted (as a CSV file) with python libraries such as Beautifulsoup. (4) Python libraries for extraction of data is done in separate instances created in AWS cloud platform because it is a very time-

[2] https://www.almanac.com/weather/history/AK

consuming process and had to be executed in parallel systems. (5) The extracted weather data is finally merged with the fire data to obtain the final dataset for this research.(6) Fire spread area is divided into 3 bins named as 'Small', 'Large', 'Severe' which is a combination of Class A-C, D-E, F-G respectively[3](considering the unavailability of large volume of data).

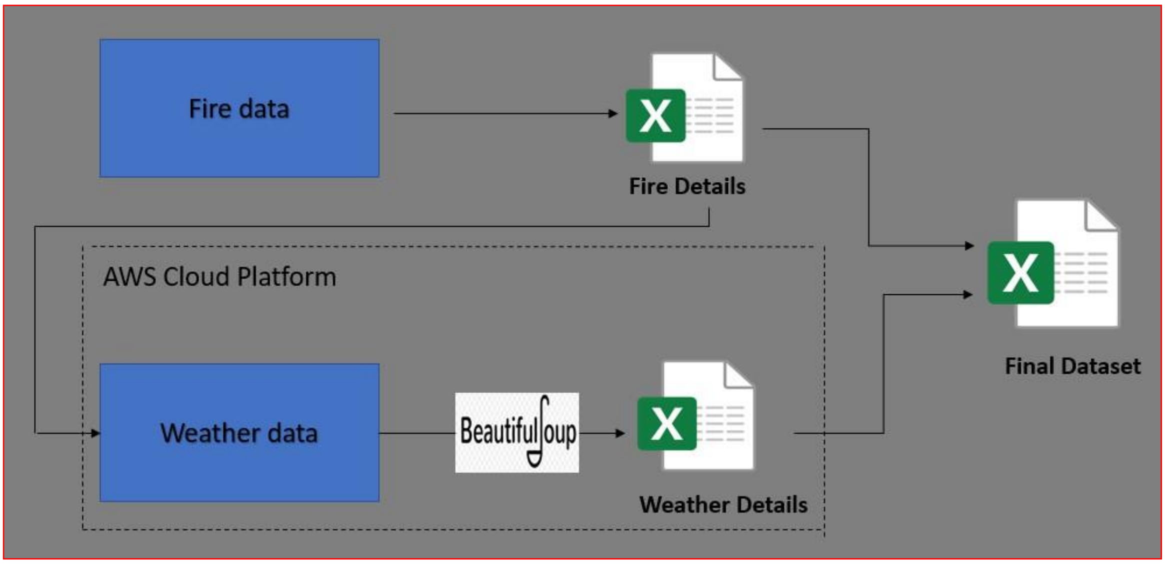


Figure 1: Data Creation Process

3.3 Project Design Process Flow

The project design phase (depicted in Figure 2) consists of three layers (1) Presentation layer (2) Business layer (3) Database layer. In presentation layer various visualization are done using Tableau, Python, R. In business layer, explanatory analysis, feature engineering, feature selection and feature reduction and model implementation has been done using various predefined packages and few custom functions in Python and R. Data resides in the form a CSV in the data layer.

Traditional KDD methodology has been modified as per the problem statement and requirement. Data has been collected from multiple sources and merged to build the final dataset. Data Acquisition and merging, initial pre-processing been done in the next section

[3] https://www.nwcg.gov/term/glossary/size-class-of-fire

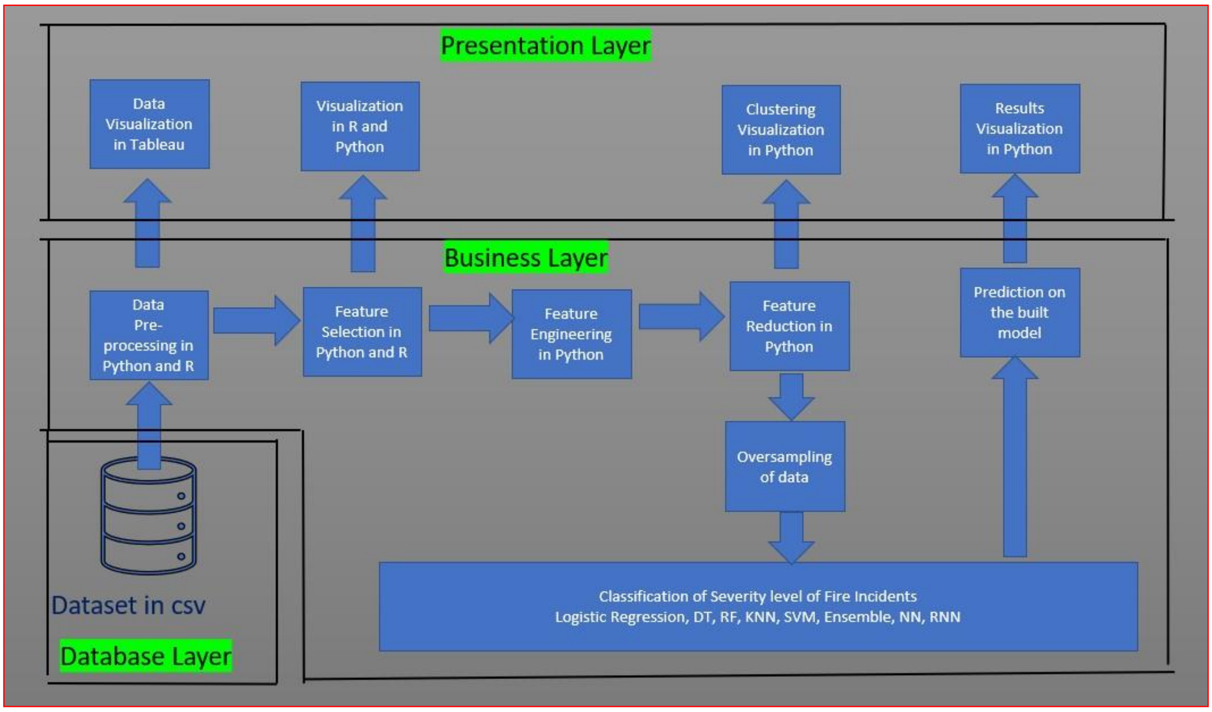


Figure 2: Project Design Process Flow

4 Data Acquisition and Data Pre-Processing

This section provides a detailed description on the process of creating the dataset from scratch, transforming the data as required, methods for selecting the important features, reducing it into minimal components and visualizing the data in reduced dimensions to check if there are any clusters formed.

4.1 Data Acquisition

Fire Incidents have been collected from Alaska Interagency Coordination Centre – Alaska Daily Stats Report[4], which contains the details of fire incident from 1993 to 2019 with the date of the incident, the total number of fires (cumulative over a year), total acres of area burnt (cumulative over a year), preparedness level of the Fire Department on the scale of 1-5, where 5 is the best preparedness level and 1 is the worst preparedness level, total number of fire incidents occurred due to human activities and lightening and their corresponding acres of area burnt. Analysis of classification of fire based on its reason (human activities or lightening) has been kept out of the scope for this research, thus related columns are dropped. Columns that are considered relevant for this research are as follows: 'ID', 'FireSeason', 'Month', 'Day', 'SitReportDate', 'TotalFires',' TotalAcres', 'PrepLevel'.
Historical weather data has been collected from 'THE OLD FARMER'S ALMANAC' website[5] which contains historical weather data for the USA and Canada. The inputs that need

[4] https://fire.ak.blm.gov/predsvcs/intel.php
[5] https://www.almanac.com/weather/history

to be provided for extracting the weather details are – (1) Area name (2) Date in the form of YYYY-MON-DD. The same details are also present in the URL of the resultant page. Akiachak area has been chosen for this research, and dates are selected from the fire incident data. URLs are created for scrapping the data for the fire incident day, as well as 4days before it. With the help of predefined packages in Python, an automated system is designed so that system will browse all the given URLs and extract the meteorological details such as minimum temperature in Fahrenheit(F), mean temperature(F), maximum temperature(F), mean sea level pressure in Inches of Mercury(IN), Mean due point(F), Total precipitation in inches(IN), visibility in Miles(MI), Mean wind speed in miles per hour(MPH), maximum sustained wind speed(MPH), maximum wind gust(MPH). All these details were extracted for the reported day for fire incident (Day0) and four days (Day-1, Day-2, Day-3, Day-4). The total number of features and observation extracted is shown in Table1.

Table 1: Dataset Description

Dataset	Record Count	Attribute count
Fire Dataset	3264	8
Day-0 Weather data	3264	10
Day-1 Weather data	3264	10
Day-2 Weather data	3264	10
Day-3 Weather data	3264	10
Day-4 Weather data	3264	10

Once the dataset is extracted for all five days, the next step is to merge them all based on the date. The same has been depicted in Figure3. In this case, the number of fire incidents and the total acres burnt is cumulative over a year. Thus, for simplicity, the cumulative numbers were converted into individual numbers. Further few duplicate records were removed from the dataset. Therefore, the number of records in the dataset got reduced from 3264 to 3249. This concludes objective2 (mentioned in section 1.3).

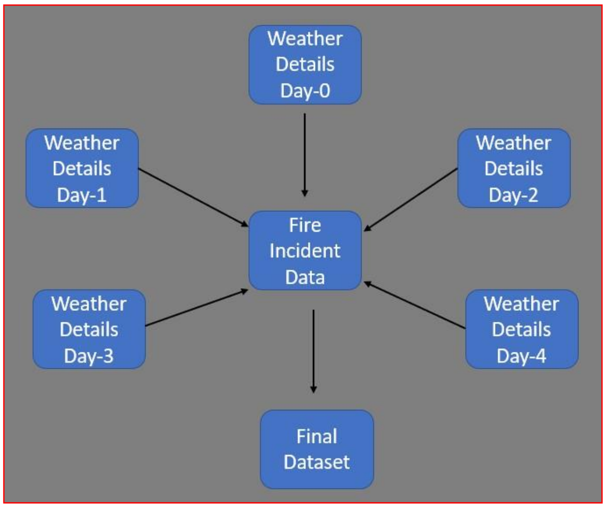


Figure 3: Merging of dataset

4.1.1 Handling of Missing values

Missing values is the first obstacle that needs to be dealt with, before building any model in machine learning. Missing values in this dataset were checked to find out the percentage of data missing for every column (whichever has at least one missing value). Figure 4 depicts the same information in form a of bar chart.

Figure 4: Precentage of Missing Values

Upon checking the distribution of the various attributes, it was observed that Maximum wind gust for all 5days (Day0, Day-1, Day-2, Day-3, Day-4) has outlier (999MPH) which constitute 33% of the data. All these outliers were replaced by Null which is further imputed in the next stage. There is no influential outlier in the remaining data. Package knowns as MICE (Multivariate Imputation via Chained Equation) in R has been used to impute the missing values as it creates multiple imputations as compared to the single imputation by mean. It imputes the missing value in a column based on other observed values and an imputation model (default is Predictive Mean Matching for numerical variables and Multinomial Logistic regression). This package generates three different datasets which are only different in missing values. The dataset whose column mean is closest to the corresponding original observed column has been accepted to be the final dataset. Figure5 represents the missing values in a column in red while the observed values in blue. The first block (row-1, column-1) represents the original dataset, while the other blocks represent the imputed values in different datasets generated through MICE.

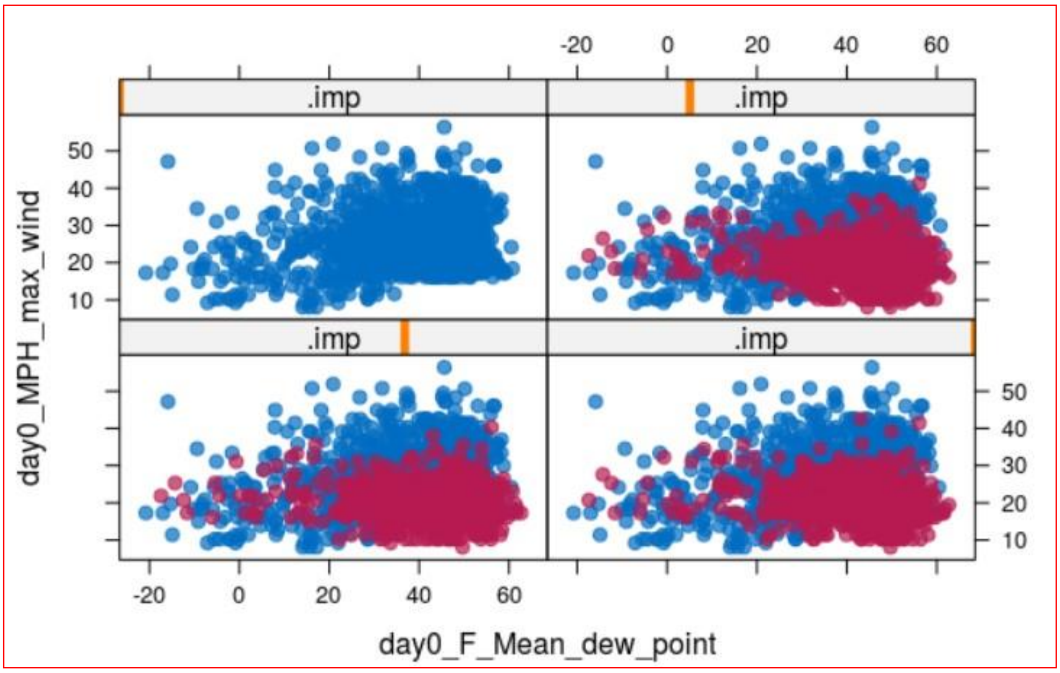


Figure 5: Missing Value Imputation

4.2 Feature Engineering

The final dataset contains both numerical and categorical values and each requires a different kind of treatment before applying machine learning models to train. Below are the few feature engineering techniques applied to this dataset.

4.2.1 One-Hot Encoding and Feature Scaling

Categorical values present in the dataset are Month, Year while Preparedness level for firefighters is ordinal data. One hot encoding has been applied to all the nominal and ordinal data, to transform each category to a different feature resulting in a binary transformation of 0 and 1. For example, the month column has different values (Jan – Dec), which are transformed into different columns with the name 'month_Jan' and so on. After one-hot encoding, the shape of the data increased from (3248,55) to (3248,95). Numerical values in the dataset are on different scales such as temperature feature is in Fahrenheit while visibility is in miles, precipitation is in inch, and wind speed is in miles per hour. All these measuring units are different, and so are the values. Machine learning algorithms only see numbers, where more significant numbers play a decisive role during training. Even Gradient-Descent in Neural Network converges faster in scaled data. Therefore, it is necessary to bring all the values into a common scale of 0-1. Normalization has not been performed for tree-based models such as DT and Random Forest. Data such as precipitation are highly skewed, with a maximum amount of rainfall only in few months and approximately zero in the remaining months. In such scenarios data has been transformed to a logarithmic scale to achieve a normal distribution. This concludes objective3 (mentioned in 1.3).

4.3 Feature Selection

Feature selection plays an important role in machine learning as it selects only key features to train the model. It is not advisable to use all the available features to build an algorithm. Multiple features often contain noise, which reduces the model's accuracy and complicates the computation. Thus, Feature selection eliminates data redundancy, avoid multi-collinearity, thus improves the model's accuracy, and reduces complexity.

4.3.1 Pearson's Correlation Coefficient

Pearson's Correlation Coefficient has been used to determine if there is any linear correlation between the input attributes. Correlated attributes (r > 0.8) are removed, resulting in a dataset of shape (3248,58) from (3248,95). A Multiple Logistic regression model has been implemented on uncorrelated data and evaluation parameters are checked (refer to Table2) as a baseline for feature selection algorithms. The correlation matrix is shown in Figure6.

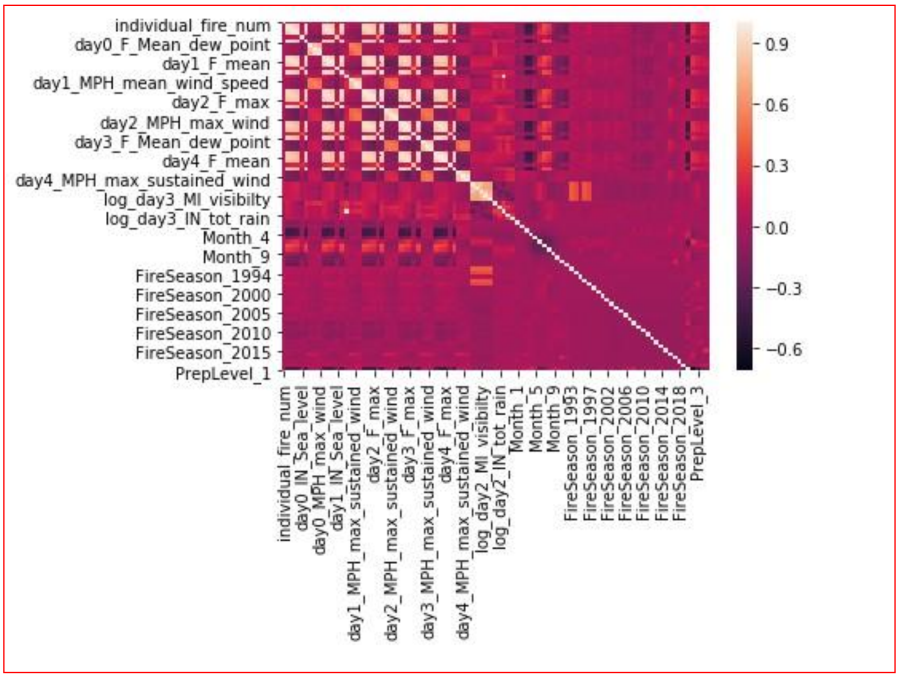


Figure 6: Correlation Matrix

4.3.2 Boruta Selection Algorithm

This feature selection algorithm works as a wrapper algorithm on random Forest. It retains all the features which are relevant to the output variable. This algorithm fits a Random Forest model on the dataset and recursively gets rid of the features that did not perform well in each iteration. Finally, the errors in the random forest model minimize which keeps only the minimal optimal subset of features. This algorithm has been applied in this dataset which retained 61 important features. The shape of the dataset got changed form (3248,95) to (3248,61). Multinomial logistic regression was built on the selected data and accuracy of the model is compared with the baseline feature selection model (refer to Table2).

4.3.3 Recursive Feature Elimination

Recursive Feature Elimination (RFE) is one of the traditional Feature selections used in ML. This method has been used in the same dataset with Random Forest as the underlying algorithm. RFE fits the model and removes the weakest feature (or features) until a specified number of features are retained. This algorithm retained only 8 important features on which a multinomial logistic regression was build and evaluated in terms of accuracy (refer to Table2).

4.3.4 Random Forest Feature Selection

Random Forest is one of the embedded methods used for feature selection, which is a combination of filter and wrapper methods (Kursa and Rudnicki, 2011). These kinds of feature selection are highly accurate, generalize better, and interpretable. This algorithm creates multiple decision trees with a random set of observations and features. The impurity is measured in the Gini Index or Information Gain. While training a tree, the feature that decreases the impurity more, is more important and is averaged out across all tress to determine the importance of the variable. This algorithm chose only 48 features to be important in this dataset. Several caveats concern the use of a random forest selection algorithm. These include concerns such as the assignment of equal importance to correlated features and assigning preferences to high cardinality features. Thus, another tree-based feature selection algorithm called as LGBMClassifier (from the Python package - lightgbm) derived 49 important features. Multinomial logistic regression is built on the selected features and the accuracy of the model is observed and compared with the baseline feature selection model (shown in Table2).

All the above Feature Selection has been applied on the dataset and a multinomial Logistic Regression is built on the important features derived by each of the above discussed algorithms. Table2 shows the comparison among all the used algorithms. It is evident that Boruta Feature Selection works the best among them. Hence the features that are considered important by the Boruta algorithms are finalized to be used in the next section (concludes Objective4, mentioned in section 1.3).

Table 2: Comparison of different Feature Selection Methods

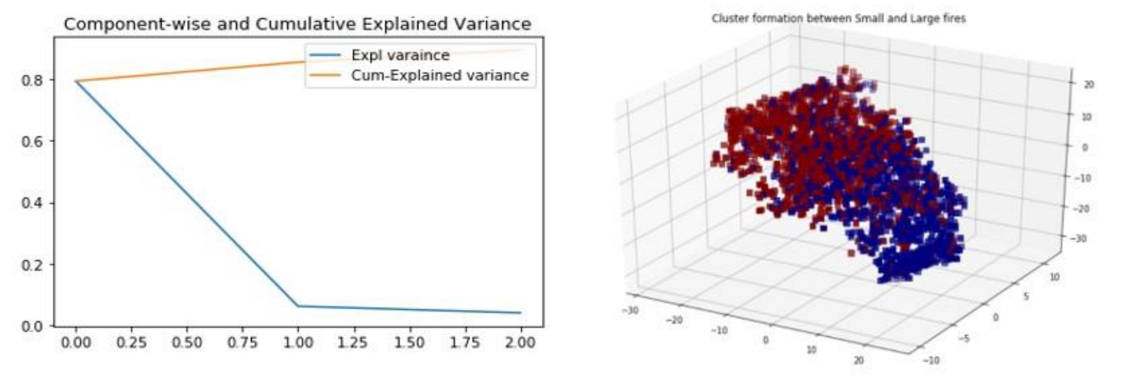
Feature Selection Methods	Number of variables retained	Type of Model Built	Accuracy Score in %
Pearson Correlation Coefficient R	58	Multinomial Logistic Regression	64.6
Recursive Feature Elimination	8	Multinomial Logistic Regression	64.6
Random Forest Feature selection	48	Multinomial Logistic Regression	64.7
Gradient Boosting Feature Selection	49	Multinomial Logistic Regression	60.9
Boruta Feature Selection	61	Multinomial Logistic Regression	65.8

4.4 Dimension Reduction

Dimension Reduction is a process of reducing the number of features in the dataset. The dataset after the initial pre-processing and feature selection has 61 features. More the number of features, more is the number of samples required, which would not only make the model complex but also increases the probability of overfitting and adds redundancy and noise.

4.4.1 Principal Component Analysis and T-distributed Stochastic Neighbour Embeddings

Principal Component Analysis (PCA) is one of the fundamental and traditional methods to derive a new set of components from the existing features, called 'Principal Components'. These principal components explain most of the variance in the dataset. The first Component explains most of the variance, followed by the second component and so on. The components are orthogonal to each other meaning, there is no correlation among the extracted features. PCA has been implemented on this dataset which shows that, there is one major component which is explaining 80% of the variance in the dataset. The same has been demonstrated in the elbow plot in Figure7(a). In the graph, the blue line represents the component-wise explained variance while the orange line represents the cumulative explained variance. The problem with PCA is that the extracted features are the linear combination of the original features which are not interpretable. Therefore, there is a need for few advanced dimension reduction techniques. T-distributed Stochastic Neighbour Embeddings (t-SNE) is one of the modern techniques for dimension reduction, which was implemented in 2008. Unlike PCA, it is a non-linear method that preserves the local (cluster) structure which helps to give a better visualization of the clusters. In this research, the dataset of 61features is reduced to 3dimension, and clusters are created with the severity of the fire as labels. Two separate clusters for fire severity can be seen with blue as 'small scale' fires and marron colour indicating 'large and severe' fire incidents (shown in Figure 7b).



(a)Elbow plot for PCA (b) Clusters on t-SNE

Figure 7:Dimension Technique in PCA and t-SNE

Similar clusters are created using other non-linear dimension reduction algorithms, such as Isomap, Umap to check if any distinguished clusters are visible in reduced dimensions. A Simple multinomial logistic regression is built on the reduced dimension, derived from all the techniques applied and accuracy scores were compared. It is noticed that accuracy for t-SNE is highest as compared to other dimension reduction techniques (shown in Table3). Thus, the features extracted from t-SNE has been used to train the models in the next section (Objective5 of section1.3 is achieved).

Table 3: Comparison of Dimension Reduction Techniques

Dimension Reduction Methods	Number of Components Retained	Type of Model Built	Accuracy Score in %
Principal Component Analysis	2	Multinomial Logistic Regression	61.69
Factor Analysis	3	Multinomial Logistic Regression	60.46
Independent Component Analysis	3	Multinomial Logistic Regression	61.69
t- Distributed Stochastic Neighbor Embedding	3	Multinomial Logistic Regression	62.3
Uniform Manifold Approximation and Projection	3	Multinomial Logistic Regression	61.5
Isometric Mapping	3	Multinomial Logistic Regression	61.38

5 Implementation, Evaluation and Results of Fire Forest Classification Models

This chapter presents implementation, evaluation of the classification of forest fire spread through (1) Data preparation for ML models and implementation of SMOTE oversampling (2) Implementation and evaluation of Decision tree classifier (3) Implementation and evaluation of bagging classifier (4) Implementation and evaluation of KNN (5) Implementation and evaluation of Random Forest (6) Implementation and evaluation of Support Vector Machine (7) Implementation and evaluation of ANN (8) Implementation and evaluation of RNN on sequence of weather parameters.

Implementation:

The implementation is divided into three important phases, namely: sampling of data, model selection and hyperparameter tuning based on gridsearch optimization algorithm. Implementation has been carried out in two processes. In the first process, all the input weather features for different days are considered independent of each other and predict the severity of the fire spread (implemented from section 5.2-5.7). Derived features (after Dimension Reduction mentioned in 4.5.1) have been used to train the traditional machine learning methods for classification. In the second process, Neural Network has been implemented on all the available features (implemented in section 5.7) of the dataset and an LSTM model on a sequence of all input weather features for each day has been built to predict the severity of the fire spread (implemented in section 5.8). Grid-search algorithm has been implemented on all

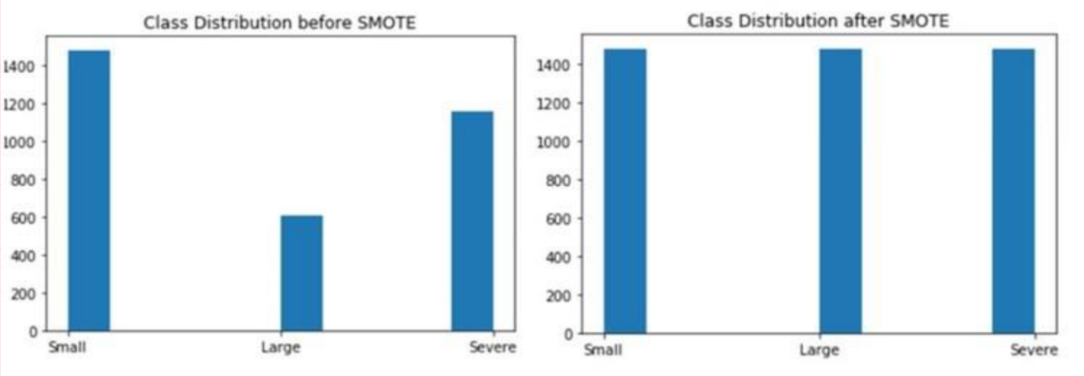
ML algorithms to obtain the best possible hyperparameters for tuning. Python package- sklearn has been used to instantiate various models and evaluation matrices.

Evaluation Parameters:

In the case of imbalanced data, accuracy metrics cannot be used to judge a classifier algorithm. For this research, Confusion Matrix has been used to derive a few other parameters to evaluate all the machine learning models such as Precision, Recall, F1-Score. Precision is the number of True positives against the sum of True Positives and False Positives. It is also called as classifier's exactness. A low precision value will indicate many False Positives cases. On the other hand, Recall is the number of True Positives divided by the number of True Positives and False Negatives. It is also called the classifier's completeness, where low recall value indicates, high False Negatives. The F1-Score is 2*((Precision*Recall)/(Precision+Recall)). It conveys the balance between Precision and Recall.

5.1 Data Preparation and SMOTE-Oversampling

The final Dataset has 3248 observations and 3 derived components from t-SNE. The data is then split into 80% training and 20% validation data. From Figure8(a), it is evident that the dataset is not balanced with respect to the dependent variable, in other words, all the classes are not represented equally. Therefore, machine learning models tend to ignore the minority class. SMOTE is one of the most common oversampling methods to solve this class imbalance problem. It oversamples the minority class by taking a minority sample and introducing synthetic examples along the line segment joining any of the K minority class nearest neighbours. Python package 'imblearn' has been used for this kind of oversampling. In this case, the number of small fires is more than large and severe fires. A simple multinomial logistic regression model was built to check the Precision, Recall, and F1-Score. Since the distribution of the data is more inclined to small and severe fires, the algorithm learns to classify these fires correctly, however, due to the presence of very few 'Large' fires, it fails to classify the 'Large' fires correctly. The same can be seen with the help of evaluation parameters (shown in % in Table4(a)). However, after applying SMOTE the distribution of all the classes becomes identical (shown in Figure 8(b)) and the evaluation parameters for the minority class have also improved (shown in % in Table4(b)). With this section objective6 (in section 1.3) is achieved.



(a) Distribution of data before SMOTE (b) Distribution of data after SMOTE

Figure 8: Implementation of SMOTE

Table 4: Evaluation of Oversampled data

	Evaluation for Multinomial Logistic Regression		
Actual Data	Precision	Recall	F1-Score
Large	0	0	0
Severe	57	65	61
Small	63	81	71

(a)Evaluation before SMOTE

	Evaluation for Multinomial Logistic Regression		
Oversampled Data	Precision	Recall	F1-Score
Large	47	30	36
Severe	50	63	56
Small	54	63	58

(b)Evaluation after SMOTE

5.2 Implementation and Evaluation of Decision Tree Classifier

Implementation:

As the name suggests, Decision Tree (DT) algorithm follows a tree-like structure where features are represented an internal node, branch as decision rule and leaf node as outcomes. Recursive partitioning happens based on attribute values. Because of its tree-like structure, it is easy to understand and interpret. Since it is a non-parametric method, it does not depend on the probability distribution assumption. Grid-search technique has been used to find the best hyperparameters for the model. Models shows maximum accuracy with maximum_number of leaf node as 99, minimum sample split as 13.

Evaluation and Results:

Confusion Matrix has been generated to validate the model prediction using the validation data and compare it with the ground truth data. The same is shown in Table5. Evaluation parameter such as Precision, Recall, F1-Score for all the classes has been mentioned (in %) in the Table5. The model has an overall accuracy of 56.4%. Objective 7_1 (in section 1.3) is achieved.

Table 5: Confusion Matrix and Evaluation for Decision Tree Classifier

Decision Tree Classifier		Predicted		
		Large	Severe	Small
Actual	Large	180	75	60
	Severe	53	177	69
	Small	69	62	145

Evaluation for Decision Tree Classifier			
Oversampled Data	Precision	Recall	F1-Score
Large	60	57	58
Severe	56	59	58
Small	53	53	53
Accuracy- 56.4%			

5.3 Implementation and Evaluation of Bagging Classifier

Implementation:

This is one of the ensemble techniques that fit random subsets of the original dataset to base classifiers and then average out (or maximum vote criteria) their predictions to result in the final prediction. Thus, it combines several weak learners to form a single strong learner, thereby reduces variance and brings stability. Decision tree has been chosen to be the base learners in this research.

Evaluation and Results:

Confusion Matrix to validate the model has been shown in Table6. The model works better in terms of Accuracy, Precision, Recall and F1-score (shown in %) in comparison with the previous model. The same is evident from Table6. Objective7_2 (in section 1.3) is achieved.

Table 6: Confusion Matrix and Evaluation for Bagging Classifier

Bagging Classifier		Predicted		
		Large	Severe	Small
Actual	Large	206	58	51
	Severe	43	193	63
	Small	69	51	156

Evaluation for Bagging Classifier			
Oversampled Data	Precision	Recall	F1-Score
Large	65	65	65
Severe	64	65	64
Small	58	57	57
Accuracy- 62.3%			

5.4 Implementation and Evaluation of KNN Classifier

Implementation:

K-nearest Neighbour (KNN) is non-parametric method where an object is classified by the number of votes from its K-nearest neighbours, where K is a positive odd integer number. Normalized data has been used for training as it improves the accuracy. Model is run on various values of K and it was observed that model runs with optimum performance with K values as 5, and distance metric as 'Euclidean'.

Evaluation and Results:

Once the model is trained with the training data, validation data is used to make prediction through the model, and this predicted output is then compared with the original data to check Precision, Recall and F1-score(shown % in Table7) for each of the class and the overall accuracy. Corresponding Confusion Matrix and evaluation parameters can be seen in Table7. The Accuracy is found to be around 61%. Objective7_3 (in section 1.3) is achieved.

Table 7: Confusion Matrix and Evaluation for KNN

K-Nearest Neighbour Classifier		Predicted: Large	Predicted: Severe	Predicted: Small
Actual	Large	208	58	49
	Severe	57	183	59
	Small	77	48	151

Evaluation for K-Nearest Neighbour Classifier			
Oversampled Data	Precision	Recall	F1-Score
Large	61	66	63
Severe	63	61	62
Small	58	55	56
Accuracy- 60.8%			

5.5 Implementation and Evaluation of Random Forest Classifier

Implementation:

Random Forest (RF) basically consists of many individual decision trees that operate together. Each tree determines a class prediction and the class with maximum votes becomes the model prediction. Randomized search with cross-validation enabled, has been used to find the best hyperparameters for tuning the model. Model runs with maximum accuracy on n_estimator as 733, min_sample_split as 2, min_sample_leafs as 1, max_dept as 100, max_features as 'auto' and bootstrap as 'True'.

Evaluation and Results:

Confusion Matrix has been used to compare the predicted output and observed validation data, which can be seen in Table8. Evaluation parameters such as Precision, Recall, F1-Score (in %) of all the output class and model accuracy (64.2%) can be seen in Table8.

Table 8: Confusion Matrix and Evaluation for Random Forest

Random Forest Classifier		Predicted: Large	Predicted: Severe	Predicted: Small
Actual	Large	210	50	55
	Severe	48	198	53
	Small	63	49	164

Evaluation for Random Forest Classifier			
Oversampled Data	Precision	Recall	F1-Score
Large	65	67	66
Severe	67	66	66
Small	60	59	60
Accuracy- 64.2%			

5.6 Implementation and Evaluation of SVM Classifier

Implementation:

Support Vector Machine (SVM) is one of the supervised models, which can perform linear classification as well non-linear classification with the help of a kernel trick (mapping the input features into a high dimensional feature space). It uses a subset of observations in the decision function, which makes this algorithm memory efficient. Gridsearch has been used to find the optimal parameters such as C=1, gamma=1 and kernel as 'rbf'.

Evaluation and Results:

Correlation Matrix and evaluation parameters (in %) has been used to evaluate the model (shown in Table9). Accuracy of the model is mere 53%. Objective7_5 (in section 1.3) is achieved.

Table 9: Confusion Matrix and Evaluation for SVM

Support Vector Machine Classifier		Predicted		
		Large	Severe	Small
Actual	Large	165	107	43
	Severe	96	164	39
	Small	90	43	143

Evaluation for Support Vector Machine Classifier			
Oversampled Data	Precision	Recall	F1-Score
Large	47	52	50
Severe	52	55	54
Small	64	52	57
Accuracy- 53.0%			

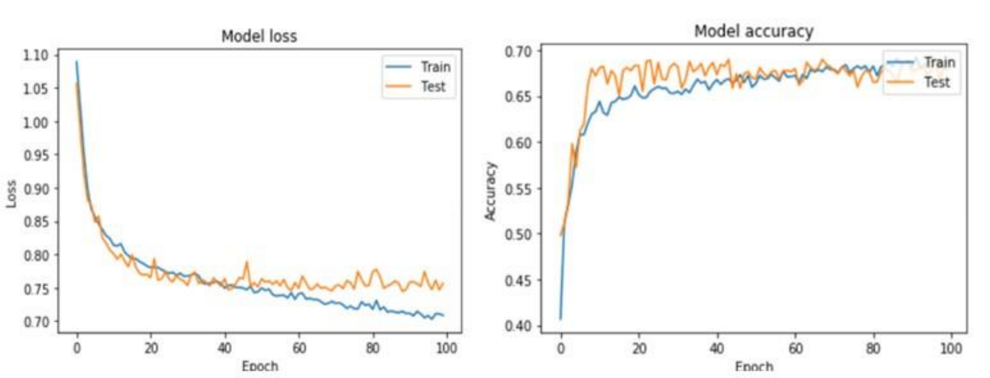
5.7 Implementation and Evaluation of Artificial Neural Network

Implementation:

Artificial Neural Network (ANN) works like a human brain. It consists of a single input layer, single/multiple hidden layers, and an output layer. The number of neurons in the input layer is the same as the number of input features (68 in this case), while the number of neurons in the hidden layers is determined heuristically. In this study, keras library in Python has been used to implement ANN with the number of layers and the number of neurons in each layer, defined with the help of a Dense constructor. Error minimization has been done using 'adam' optimizer, loss function as 'Categorical CrossEntropy' and 'accuracy' as the evaluation metrics.

Evaluation and Results:

From Figure 9(a), it is observed that the loss function of the training data approaches towards zero while the loss function of the validation of data decreases rapidly till 45 epochs, but later tends to gradually increase. Figure9(b) shows the accuracy graph of training and validation data. Model's accuracy reaches the highest value of 67% after 45 epochs. The Confusion Matrix and other evaluation parameters (in %) have been mentioned in Table10. This concludes objective 7.6 (mentioned in section 1.3).



(a)Loss Function (b) Accuracy graph

Figure 9: ANN Model Performance

Table 10: Confusion matrix and Evaluation of ANN

Artificial Neural Network		Predicted		
		Large	Severe	Small
Actual	Large	232	18	45
	Severe	102	128	62
	Small	47	12	244

Evaluation for Artificial Neural Network			
Oversampled Data	Precision	Recall	F1-Score
Large	61	79	69
Severe	81	44	57
Small	70	81	75
Accuracy- 67.86%			

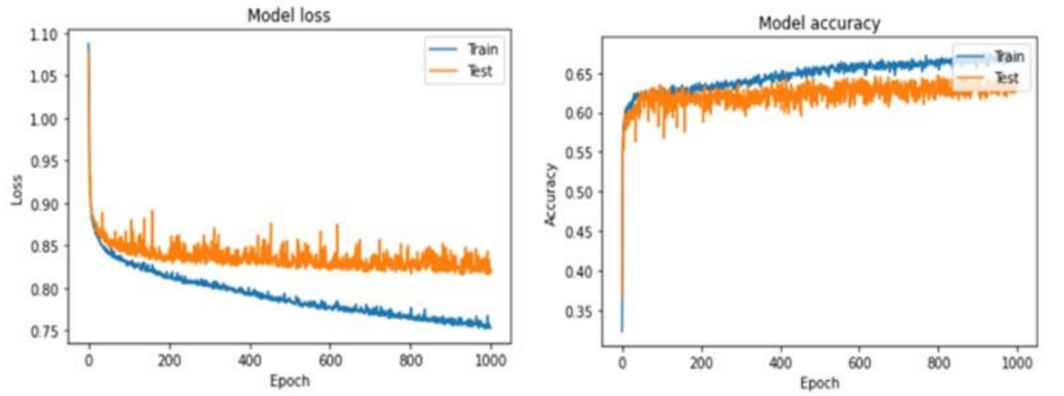
5.8 Implementation and Evaluation of LSTM

Implementation:

Long-Shot Term Memory (LSTM) is a special kind of RNN that are naturally suited to sequential data using some hidden units where the output depends on the previous computations. In this case, the weather details of each day (total 5 days) create a sequence to classify the severity of the fire. Therefore, the maximum number of time steps in this research is five. In each time steps, 12 features were given as input sequentially. A 3-Dimensional array (-1,5,12) has been created to feed the same into an LSTM.

Evaluation and Results:

Like ANN, Keras has been implemented, with a LSTM layer and two dense layers with 4 and 3 neurons. The model is compiled with ‘adam’ optimizer and ‘accuracy’ as the evaluation metrics and the number of epochs as 1000. The loss function (shown in Figure 10(a)) and the accuracy curve (shown in Figure 10(b)) remains constant after 400epochs. The Confusion Matrix and evaluation parameter (in %) have been shown in Table11. This concludes objective 8 (mentioned in section 1.3).



(a) Loss Function of LSTM (b) Accuracy Graph of LSTM

Figure 90: LSTM Model's Performance

Table 101: Confusion matrix and Evaluation of ANN

LSTM		Predicted Large	Predicted Severe	Predicted Small
Actual	Large	173	59	41
Actual	Severe	62	215	22
Actual	Small	72	64	183

Evaluation for LSTM			
Oversampled Data	Precision	Recall	F1-Score
Large	56	63	60
Severe	64	72	68
Small	74	57	65
Accuracy- 64.0%			

6 Comparison of Developed Models

Since the dataset is made balanced with the help of SMOTE-oversampling, accuracy can be used for model evaluation. Comparing all the developed models, in the previous section in terms of accuracy, it is observed that ANN with only 4layers and 100epochs works best in comparison to all other developed models. Even the sequence of inputs over the last 5days of fire incident did not produce better results than ANN. The same can be shown in the form of a graph in Figure11. The precision value of severe fires is 81%, which means, out of fires classified as 'Severe', 81% of them were actually 'Severe'. Similarly, the precision of small fires is 70%, which means out of the fires classified as 'Small', 70% of them are actually 'Small'. Likewise, precision of 'large' fires is 61%, which means out of the fires classified as 'Large', 61% of them are actually large. On the other hand, recall value of 'Large' fires is 79, which means out of the fires that are actually large, 79% of them are classified as 'Large'. Similarly, recall of 'Small' and 'Severe' fires is 44 and 81% which means, out of fires that are actually small and severe, 44 and 81% of them are classified as 'Small' and 'Severe' respectively. F1 score is a harmonic mean of precision and recall, which means it treats both precision and recall equally important. F1 score of Large, Severe, and Small fires are 69, 57, 75% respectively. This concludes the final objective of this research.

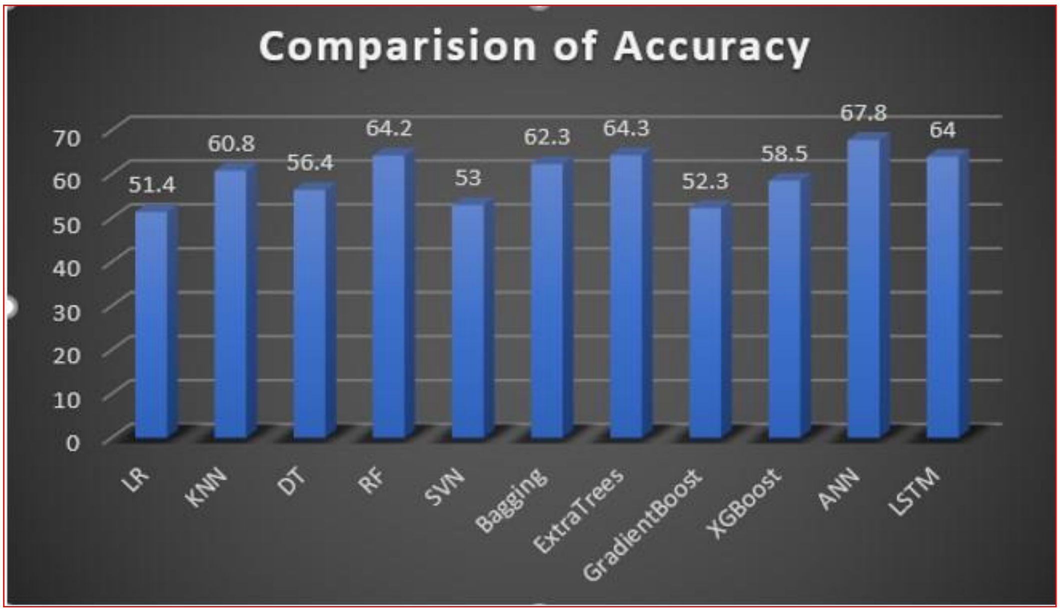


Figure 101: Comparison of Developed Models

7 Discussion

Collecting weather data through web-scrap became a tedious task because few of the weather parameters were missing in the website and had to be handled correctly. Web scraping around 3500 records took around 6-7hrs of execution time. And the same cycle was repeated five times to collect data for five days. Enough sleep time was introduced, to avoid multiple hits within a small interval of time. Python files were executed in a virtual system made in the AWS cloud platform. Once the data is collected, all the datasets were merged with the date as the key (similar to lookup in xls). Popular packages in R and Python are explored to leverage the best use of them for implementation. For example, packages in R such as mice for missing value imputation, Boruta for feature selection has been used, while for model implementation sklearn package in python has been used. For, visualization, python packages such as Plotly and Seaborn have been used, while EDA is performed on Tableau. For Neural Network and DL, Google Colab has been used to leverage the power of GPUs.Various techniques for missing value imputation were tried, such as mean imputation, median imputation, and mice imputation. The later one has been implemented as it provides more meaningful information. During feature selection, a lot of advanced techniques were applied and tested on a multinomial logistic regression, to pick the best selection algorithm. A similar approach has been performed during feature reduction. SMOTE- oversampling helped not only to increase the number of observations but also to balance the data. Therefore, accuracy has been the primary focus while evaluating any model. Various tree-based and non-tree-based algorithms have been applied. It was ensured that data points are scaled between 0-1 before applying any non-tree-based algorithm. Out of all the applied algorithm, ANN proved to be best on accuracy. It also outperformed the sequential classification models like LSTM.

8 Conclusion

The study started with understanding the relationship of fire with respect to local weather. But the hypothesis that was kept for this research was that the weather does not depend entirely on the day on which the fire takes place, instead it depends on the weather details of the previous few days. But the number of previous days that needs to be considered was unclear. Therefore, it was assumed that the weather details of the past five days, might help to predict the size of a forest fire. By considering the weather details of the past five days, and an accuracy of 67% has been achieved through ANN. Even fire classification on a sequence (with memory enabled) of weather inputs could not classify the fire spread better than ANN. The assumption for this research is that the weather of Alaska is the same all over the counties. It also assumed that if the fire incident has happened in a specific month, the past four days also falls in the same month. The preparedness levels of the firefighters also assumed to be the same on the day of the outbreak of fire and its previous four days.

As a part of future work, a sequence of weather details of the last 10-15days can be fed into a Recurring Neural Network and check if it performs better as compared to the last 4days data. In addition to the implementation of this research, a bi-directional LSTM or a GRU can be used to check if it outperforms the basic ANN. Other factors such as local population, topology, road connectivity, availability of water sources, type of soil, type of vegetation can be considered as input features.

Acknowledgement

I am sincerely thankful to Dr. Catherine Mulwa, who guided me during the research with her valuable inputs. I will also be thankful to my younger brother Mr. Sujit Sahoo who has kept me motivated all through the project duration.

References

Anshori, M. *et al.* (2019a) 'Prediction of Forest Fire using Neural Network based on Extreme Learning Machines (ELM)', in *2019 International Conference on Sustainable Information Engineering and Technology (SIET). 2019 International Conference on Sustainable Information Engineering and Technology (SIET)*, Lombok, Indonesia: IEEE, pp. 301–305. doi: 10.1109/SIET48054.2019.8986106.

Anshori, M. *et al.* (2019b) 'Prediction of Forest Fire using Neural Network based on Extreme Learning Machines (ELM)', in *2019 International Conference on Sustainable Information Engineering and Technology (SIET). 2019 International Conference on Sustainable Information Engineering and Technology (SIET)*, pp. 301–305. doi: 10.1109/SIET48054.2019.8986106.

Assessment of forest fire danger using automatic weather stations and MODIS TERRA satellite datasets for the state Madhya Pradesh, India - IEEE Conference Publication (no date). Available at: https://ieeexplore.ieee.org/document/8126118 (Accessed: 2 July 2020).

Azevedo, A. and Santos, M. (2008) 'KDD, semma and CRISP-DM: A parallel overview', in, pp. 182–185.

Beggs, B. J. (1976) *Forest fire behaviour tables for Western Australia*. Forest Dept., Western Australia. Available at: https://agris.fao.org/agris-search/search.do?recordID=US201300061773 (Accessed: 2 July 2020).

Bourgeau-Chavez, L. L., Kasischke, E. S. and French, N. H. F. (1993) 'Detection And Interpretation Of Fire-disturbed Boreal Forest Ecosystems In Alaska Using Spacebarne SAR Data', in *Proceedings of IEEE Topical Symposium on Combined Optical, Microwave, Earth and Atmosphere Sensing*. *IEEE Topical Symposium on Combined Optical, Microwave, Earth and Atmosphere Sensing*, Albuquerque, NM, USA: IEEE, pp. 236–239. doi: 10.1109/COMEAS.1993.700229.

Bradshaw, L. S. *et al.* (1984) *The 1978 National Fire-Danger Rating System: technical documentation*. INT-GTR-169. Ogden, UT: U.S. Department of Agriculture, Forest Service, Intermountain Forest and Range Experiment Station, p. INT-GTR-169. doi: 10.2737/INT-GTR-169.

Cappellini, V., Mattii, L. and Mecocci, A. (1989) 'An intelligent system for automatic fire detection in forests', in Cantoni, V. et al. (eds) *Recent Issues in Pattern Analysis and Recognition*. Berlin, Heidelberg: Springer Berlin Heidelberg (Lecture Notes in Computer Science), pp. 351–364. doi: 10.1007/3-540-51815-0_67.

Chelli, S. *et al.* (2015) 'Adaptation of the Canadian Fire Weather Index to Mediterranean forests', *Natural Hazards*, 75(2), pp. 1795–1810. doi: 10.1007/s11069-014-1397-8.

French, N. H. F., Kasischke, E. S. and Bourgeau-Chavez, L. L. (1994) 'Multi-sensor analysis of the effects of fire in the Alaskan boreal forest', in *Proceedings of National Aerospace and Electronics Conference (NAECON'94)*. *Proceedings of National Aerospace and Electronics Conference (NAECON'94)*, pp. 1085–1089 vol.2. doi: 10.1109/NAECON.1994.332922.

Kawano, K., Kudoh, J. and Makino, S. (1999) 'Forest fire detection in Far East region of Russia by using NOAA AVHRR images', in *IEEE 1999 International Geoscience and Remote Sensing Symposium. IGARSS'99 (Cat. No.99CH36293)*. *IEEE 1999 International Geoscience and Remote Sensing Symposium. IGARSS'99*, Hamburg, Germany: IEEE, pp. 858–860. doi: 10.1109/IGARSS.1999.774465.

Koutsias, N. *et al.* (2013) 'On the relationships between forest fires and weather conditions in Greece from long-term national observations (1894–2010)', *International Journal of Wildland Fire*, 22(4), p. 493. doi: 10.1071/WF12003.

Kursa, M. and Rudnicki, W. (2011) 'The All Relevant Feature Selection using Random Forest'.

Lee, G. *et al.* (1999) 'Analysis of vegetation types affected by the 1997 forest fires in Sumatra', in *IEEE 1999 International Geoscience and Remote Sensing Symposium. IGARSS'99 (Cat. No.99CH36293)*. *IEEE 1999 International Geoscience and Remote Sensing Symposium. IGARSS'99*, Hamburg, Germany: IEEE, pp. 738–740. doi: 10.1109/IGARSS.1999.774424.

Liang, H., Zhang, M. and Wang, H. (2019a) 'A Neural Network Model for Wildfire Scale Prediction Using Meteorological Factors', *IEEE Access*. *IEEE Access*, 7, pp. 176746–176755. doi: 10.1109/ACCESS.2019.2957837.

Liang, H., Zhang, M. and Wang, H. (2019b) 'A Neural Network Model for Wildfire Scale Prediction Using Meteorological Factors', *IEEE Access*. *IEEE Access*, 7, pp. 176746–176755. doi: 10.1109/ACCESS.2019.2957837.

Liang, H., Zhang, M. and Wang, H. (2019c) 'A Neural Network Model for Wildfire Scale Prediction Using Meteorological Factors', *IEEE Access*. *IEEE Access*, 7, pp. 176746–176755. doi: 10.1109/ACCESS.2019.2957837.

Martinez-Canales, J. F. (1997) 'A review of the incidence of medium and high voltage overhead electric power lines in causing forest fires', in *14th International Conference and Exhibition on Electricity Distribution (CIRED 1997 - Distributing Power for the Millennium)*. *14th International Conference and Exhibition on Electricity Distribution (CIRED 1997 - Distributing Power for the Millennium)*, Birmingham, UK: IEE, pp. v3-27-v3-27. doi: 10.1049/cp:19970533.

Minardi, J., Marchisio, G. B. and Treder, R. P. (1999) 'Spatial linear modeling and forecasting of forest fires across the United States', in *IEEE 1999 International Geoscience and Remote Sensing Symposium. IGARSS'99 (Cat. No.99CH36293)*. *IEEE 1999 International Geoscience and Remote Sensing Symposium. IGARSS'99*, Hamburg, Germany: IEEE, pp. 861–863. doi: 10.1109/IGARSS.1999.774466.

Molovtsev, M. D. and Sineva, I. S. (2019) 'Classification Algorithms Analysis in the Forest Fire Detection Problem', in *2019 International Conference 'Quality Management, Transport and Information Security, Information Technologies' (IT&QM&IS)*. *2019 International Conference 'Quality Management, Transport and Information Security, Information Technologies' (IT&QM&IS)*, Sochi, Russia: IEEE, pp. 548–553. doi: 10.1109/ITQMIS.2019.8928398.

Nami, M. H. *et al.* (2018) 'Spatial prediction of wildfire probability in the Hyrcanian ecoregion using evidential belief function model and GIS', *International Journal of Environmental Science and Technology*, 15(2), pp. 373–384. doi: 10.1007/s13762-017-1371-6.

Novac, I., Geipel, K., *et al.* (2020) 'A Framework for Wildfire Inspection Using Deep Convolutional Neural Networks', in, pp. 867–872. doi: 10.1109/SII46433.2020.9026244.

Novac, I., Geipel, K. R., *et al.* (2020a) 'A Framework for Wildfire Inspection Using Deep Convolutional Neural Networks', in *2020 IEEE/SICE International Symposium on System Integration (SII)*. *2020 IEEE/SICE International Symposium on System Integration (SII)*, pp. 867–872. doi: 10.1109/SII46433.2020.9026244.

Novac, I., Geipel, K. R., *et al.* (2020b) 'A Framework for Wildfire Inspection Using Deep Convolutional Neural Networks', in *2020 IEEE/SICE International Symposium on System Integration (SII)*. *2020 IEEE/SICE International Symposium on System Integration (SII)*, pp. 867–872. doi: 10.1109/SII46433.2020.9026244.

Patel, M., Patel, A. and Ghosh, D. R. (no date) 'Precipitation Nowcasting: Leveraging bidirectional LSTM and 1D CNN', p. 7.

Patel, M., Patel, A. and Ghosh, R. (2018) 'Precipitation Nowcasting: Leveraging bidirectional LSTM and 1D CNN', *ArXiv*.

Peet, G. B. (no date) 'FOREST FIRE BEHAVIOUR TABLES', p. 50.

Rizzo, R. *et al.* (2016) ‘A Deep Learning Approach to DNA Sequence Classification’:, in, pp. 129–140. doi: 10.1007/978-3-319-44332-4_10.

Safi, Y., Bouroumi, A. and Bouroumi, A. (2011) ‘A neural network approach for predicting forest fires’, in *2011 International Conference on Multimedia Computing and Systems*. *2011 International Conference on Multimedia Computing and Systems*, pp. 1–5. doi: 10.1109/ICMCS.2011.5945716.

Size Class of Fire | NWCG (no date). Available at: /term/glossary/size-class-of-fire (Accessed: 9 August 2020).

Suresh Babu, K. V. *et al.* (2017) ‘Assessment of forest fire danger using automatic weather stations and MODIS TERRA satellite datasets for the state Madhya Pradesh, India’, in *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*. *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pp. 1876–1881. doi: 10.1109/ICACCI.2017.8126118.

UCI Machine Learning Repository: Forest Fires Data Set (no date). Available at: http://archive.ics.uci.edu/ml/datasets/Forest+Fires (Accessed: 2 July 2020).

Vetrita, Y. *et al.* (2012) ‘Drought and fine fuel moisture code evaluation: An early warning system for forest/land fire using remote sensing approach’, *International Journal of Remote Sensing and Earth Sciences*, 9, pp. 140–147. doi: 10.30536/j.ijreses.2012.v9.a1841.

Vilar, L. *et al.* (2010) ‘A model for predicting human-caused wildfire occurrence in the region of Madrid, Spain’, *International Journal of Wildland Fire*, 19(3), p. 325. doi: 10.1071/WF09030.

Wagner, C. E. V. (1987) *Development and structure of the Canadian Forest Fire Weather Index System*. Available at: /paper/Development-and-structure-of-the-Canadian-Forest-Wagner/5145e78acc8fabac336cee7e51dc0c01ec7654e5 (Accessed: 2 July 2020).

Wotton, B. M. and Martell, D. L. (2005) ‘A lightning fire occurrence model for Ontario’, *Canadian Journal of Forest Research*, 35(6), pp. 1389–1401. doi: 10.1139/x05-071.

Zhang, Z. (2018) *Multivariate Time Series Analysis in Climate and Environmental Research*. doi: 10.1007/978-3-319-67340-0.