Application and Evaluation of Forest Fire Size Prediction Model Based on Machine Learning Algorithms

Executive Summary

By Ni Zikun

BEMM466J - Business Project

Professor Salimeh Pour Mohammad & Beth Kewell

Jan 2022

1. Subject Matter

This project proposes a method to build a forest fire size prediction model with higher accuracy. The prediction model is based on machine learning algorithms and forest fire related features. In the context of climate change, forest fires are becoming more frequent and serious, posing a great threat to the natural environment, property and human health. Forest fires in the tropics reduced forest cover by 12.2 million hectares in 2020, roughly half the size of the United Kingdom, an increase of 12% over the previous year. The fire size is an important index to measure the severity of forest fires. Predicting the size of the forest fire in the first place can help fire departments make effective rescue plans, such as how many areas to evacuate and how many supplies and firefighters will be needed. Therefore, the prediction model proposed in this project can help the government and fire departments reduce the loss and pollution caused by forest fires to some extent.

This project includes five major research procedures. The initial procedure reviews the literature and methods to identify the forest fire related factors and the effective classification algorithms. The following procedure collects the historical forest fire data and related factors through public databases. The third procedure focuses on feature selection based on different machine learning algorithms. The fourth procedure establishes prediction models using each machine learning algorithm and its key features. In the last procedure, the most effective model is chosen by comparing models' predictive performance evaluation.

2. Methods of Analysis

The experimental datasets used in this project contain the historical ignition point dataset, historical ignition point humidity dataset, historical ignition point weather dataset, and ignition point elevation dataset. The historical ignition point dataset of Alberta from 1950 to 2019 is downloaded from Canada's Natural Resources Datamart. The historical ignition point humidity dataset is downloaded from Canadian Wildland Fire Information System Datamart. The historical ignition point weather dataset is

crawled from the Canadian Weather website. The ignition point elevation dataset is collected from the Google Elevation API.

The modelling algorithms in this project include decision tree, random forest, support vector machine and logistic regression algorithm. Since forest fire size can be classified into seven sizes based on the U.S. Forest Fire Classification System, forest fire prediction is essentially a multi-classification problem. According to the literature and methods review, the above four classification algorithms may have better performance in forest fire prediction.

Precision, Recall, and F1-Score are used to evaluate the predictive performance in this project due to the imbalanced number of samples (Class A forest fire accounts for 74%). Precision is a metric to understand how well the model correctly predicts positive observations, while Recall measures how well the model correctly predicted all possible positive observations. The F1-Score takes the weighted average of Precision and Recall. Therefore, Precision, Recall and F1-Score can effectively evaluate the prediction performance of classification models in this project.

3. Findings and Recommendations

Increasing the number of features in the model improves its predictive ability, but once the number of features has passed its peak, the predictive performance begins to decline. Feature selection is a method for determining which features are most useful in predicting the objective. In the feature selection research of different algorithms, this project finds that the key features of tree-based algorithms are elevation, temperature, duff moisture code, speed of maximum gust and drought code. The key features of the logistic regression algorithm are fine fuel moisture code, relative humidity, duff moisture code, and temperature. The key features of the support vector machine algorithm are fine fuel moisture code, temperature, relative humidity, and duff moisture code.

After establishing the prediction models based on their key features, this project evaluates and compares the predictive performance of these models. Through the evaluation of model prediction performance, this project finds that the random forest

model had a better prediction performance in general, with an F1-score of 0.84, followed by the support vector machine model with an F1-score of 0.81. The predictive performance evaluation on each fire size also shows that the random forest model has the higher Precision, Recall and F1-score in all classes of fire size. However, the predictive performance of the logistic regression model is lower, indicating that the correlation between related features and forest fire size is not linear. Therefore, it can be concluded that both the random forest algorithm and support vector machine algorithm can predict the final fire size based on fire-related features with reasonable accuracy.

Based on the findings in predictive performance evaluation, it is recommended to use random forest algorithm combined with elevation, temperature, duff moisture code, speed of maximum gust and drought code to predict the size of forest fires. The combination of these factors and the algorithm is effective in predicting the size of forest fires in Alberta. This project is of great significance for forest fire protection. Through this forest fire size prediction model, the fire department can get an idea of the final size of a forest fire in the early stage, and then take appropriate measures to reduce the loss caused by forest fires.

4. Ethics and Limitations

The ethics of Internet crawlers are considered in this project. When crawling historical weather dataset, this project complied with the requirements of the website's robots exclusion standard and Internet ethics.

This project has several limitations. There is a discrepancy between the nearest weather station's historical weather data and the actual weather at the forest fire point. More forest fire related factors, such as population density, vegetation type and slope aspect, could be explored and incorporated into the model. In addition to the four machine learning algorithms studied in this project, more artificial intelligence algorithms can be experimented to predict forest fire size.