Analysis of Forest Fires using Bayesian Network

Kiana Alian, Luyao Ma, Arnab Mukhopadhyay
Master's Degree in Artificial Intelligence, University of Bologna
{ kiana.alian, luyao.ma, arnab.mukhopadhyay }@studio.unibo.it

February 12, 2025

Abstract

Forest fires pose a significant environmental challenge, with their occurrence and severity influenced by various meteorological and environmental factors. In this study, we leverage Bayesian Networks to model and analyze the probabilistic dependencies among key factors such as temperature, humidity, wind speed, and rainfall using the Forest Fires dataset from the UCI Machine Learning Repository. Four different Bayesian Network structures were evaluated: a manually-defined model, a tree search-based model, a hill-climbing with score-based model, and a combined hill-climbing, constraint-based, and score-based model. Each model was trained and assessed based on its ability to predict fire severity and its interpretability in capturing key dependencies. Our findings indicate that tree-based models provide simpler structures with lower accuracy, while score-based and hybrid methods yield higher accuracy at the cost of increased complexity. These insights contribute to improving predictive models for fire risk assessment and informing wildfire management strategies.

Introduction

Domain

In this study, we utilize the Forest Fires dataset from the UCI Machine Learning Repository, which was collected from Montesinho Natural Park in Portugal. This dataset records fire incidents alongside key meteorological and environmental variables, allowing for a detailed probabilistic analysis of wildfire risk.

The dataset contains 12 primary attributes, including temperature, humidity, wind speed, and rainfall, which are essential climatic variables affecting fire behavior. Additionally, it includes specialized fire weather indices used by fire management agencies:

- **Fine Fuel Moisture Code (FFMC):** Represents the moisture content of surface fuels, influencing fire ignition probability.
- **Duff Moisture Code (DMC):** Measures the moisture level in moderately compacted organic matter, affecting fire sustainability.
- **Drought Code (DC):** Reflects deep organic layer dryness, impacting fire intensity over prolonged periods.
- **Initial Spread Index (ISI):** An empirical metric that quantifies how fast a fire spreads under given wind and fuel conditions.

The dataset also includes geospatial coordinates of fire locations (X, Y), the month and day of each fire incident, and the burned area (ha) as the primary outcome variable. The goal of this study is to analyze the probabilistic dependencies between these factors using Bayesian Networks, providing insights into fire risk assessment and prediction.

Aim

To investigate the relationships between environmental and meteorological factors influencing forest fire occurrence and severity using Bayesian Networks. This analysis aims to provide probabilistic predictions, enable scenario simulations, and offer insights into key contributors to fire risk by capturing dependencies among variables such as temperature, humidity, wind speed, and rainfall.

Method

First, four different Bayesian Network models are constructed, either by explicitly defining the structure or by inferring it from data. For each model, parameters are learned from the available data. To analyze the impact of different factors on model accuracy, we vary evidence nodes, conditional probability distributions, and inference methods. We then apply different probabilistic inference techniques to assess how these variations influence the accuracy and performance of the models

Model

Custom Bayesian Network

The Bayesian Network structure was manually defined based on domain intuition and known dependencies among meteorological factors influencing forest fires. Temperature (`temp_bin`) directly affects the Fine Fuel Moisture Code (`FFMC_bin`), which represents the dryness of surface fuels. Similarly, relative humidity (`RH_bin`) influences the Duff Moisture Code (`DMC_bin`), affecting deeper organic layers. Wind speed (`wind_bin`) contributes to fire spread and impacts the Initial Spread Index (`ISI_bin`), which measures fire propagation potential. Rainfall (`rain_bin`) and

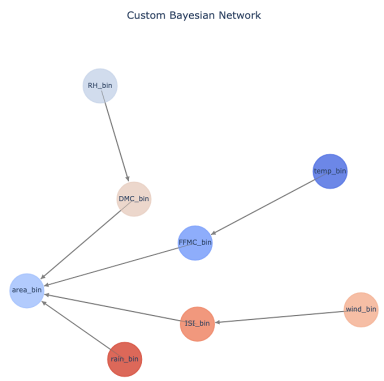

Figure 1: Custom Bayesian Network Structure

fuel moisture indices (`FFMC_bin`, `DMC_bin`) directly contribute to predicting the burned area (`area_bin`). These connections help model the causal relationships between weather conditions and fire behavior, allowing inference on fire severity based on observed environmental factors.

Tree Search

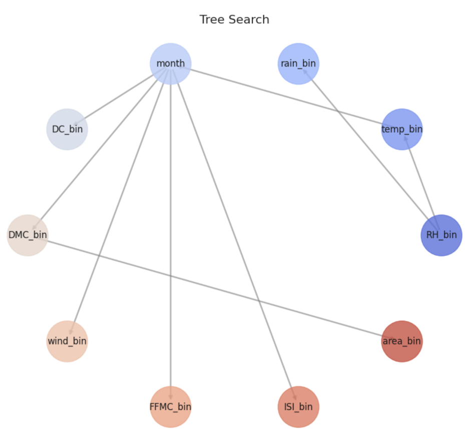

Figure 2: Tree Search

The Bayesian network structure was learned using a tree search algorithm to uncover the most significant dependencies within the fire dataset. 'Month' appears as the root node, indicating its importance in determining fire behavior. The connections represent the dependencies that the tree search algorithm found the most probable, given the trade-off between model fit and complexity.

Hill Climbing

This Bayesian network structure, refined through Hill Climbing, reveals complex inter-dependencies between weather and fuel conditions, indicating a sophisticated model reflecting intricate relationships, in order to predict fire area. Key variables like *month* and *rain_bin* have multiple influences. *ISI_bin* is also dependent on weather and previous variables. This model suggests that accurate fire prediction requires considering both direct and indirect effects of multiple factors.

Hill Climbing with Score Based

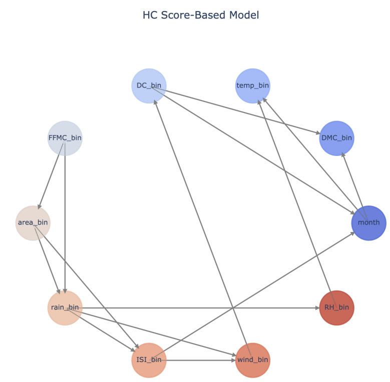

Figure 3: Hill climbing with score based

The Bayesian Network structure was learned using the Hill Climbing algorithm with the Bayesian Dirichlet Equivalent Uniform (BDeu) score as the evaluation metric. Hill Climbing is a heuristic search method that iteratively explores the space of possible network structures by adding, removing, or reversing edges to maximize the scoring function. This process continues until no further modifications improve the model's fit to the data. The learned structure captures complex dependencies between meteorological and environmental factors influencing fire occurrence. Compared to manually defined networks, this approach leverages data-driven optimization to discover relationships that may not be immediately intuitive, enhancing predictive accuracy.

Hill Climbing with Score and Constraint Based

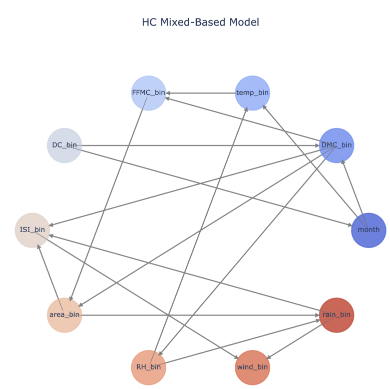

Figure 4: Hill climbing with score and constraint based

This "HC Constrained Model" illustrates a complex network of relationships among fire-related factors, learned with constraints during the Hill Climbing search. Each node represents a binned variable, such as FFMC, temperature, month, or area burned. The interconnected web of red edges signifies dependencies between these variables; the edges indicate possible direction that can flow from each node. This dense network suggests that many factors directly and indirectly affect fire behavior, representing the complexity inherent in fire prediction. Constraints imposed during Hill Climbing have led to a model that, while potentially complex, aims to balance accuracy with adherence to pre-defined rules or domain knowledge, resulting in an interconnected web of influences that suggests a holistic consideration of all factors for fire risk assessment.

Analysis

Experimental setup

We employed Bayesian network modeling using the pgmpy library to analyze the probabilistic dependencies among environmental and meteorological factors influencing forest fires. Instead of assuming feature independence as required by Naïve Bayes, we explored different Bayesian network structures to better capture dependencies in the data.

For Bayesian network learning, we implemented four distinct models: a manually defined structure based on domain knowledge, a structure learned using a tree search algorithm, a network optimized using hill climbing with the Bayesian Dirichlet Equivalent Uniform (BDeu) score, and an extension of hill climbing that integrates both score-based and constraint-based structure learning to balance accuracy and interpretability.

For parameter estimation, we applied maximum likelihood estimation to learn the conditional probability distributions, and inference was performed using the variable elimination algorithm. Model performance was evaluated using accuracy, recall, precision, and F1-score, computed via scikit-learn. To visualize network structures, networkx and seaborn were utilized, while plotly was employed for data exploration and graphical analysis. The evaluation framework compared the effectiveness of each Bayesian network in predicting fire severity, assessing structural complexity against predictive performance.

Results

Model	Accuracy	Recall	F1 Score	Precision
CBN	0.9808	0.9808	0.9808	0.9808
TSB	**0.9904**	**0.9904**	**0.9856**	**0.9905**
HCSB	**0.9904**	**0.9904**	**0.9856**	**0.9905**
HCSCB	**0.9904**	**0.9904**	**0.9856**	**0.9905**

Table 1: Performance comparison of Bayesian Network models on fire prediction.

The Tree Search-Based (TSB) and Hill Climbing methods, including both Score-Based (HCSB) and Constraint & Score-Based (HCSCB) variations, demonstrate superior performance compared to the Custom Bayesian Network (CBN). Specifically, these models achieve an accuracy of **0.9904**, recall of **0.9904**, F1 score of **0.9856**, and precision of **0.9905**, whereas the CBN attains an accuracy of 0.9808, recall of 0.9808, F1 score of 0.9808, and precision of 0.9808.

The nearly identical metrics among the TSB, HCSB, and HCSCB methods suggest that the additional constraints in HCSCB do not significantly impact model performance, yielding results comparable to its score-based counterpart. Therefore, TSB and Hill Climbing-based approaches (HCSB and HCSCB) are preferable for fire prediction, as they provide more accurate results than the manually defined CBN.

Conclusion

In this study, we initially considered using Naïve Bayes for fire prediction but found that its assumption of feature independence was too restrictive for our dataset. To better capture dependencies between variables, we designed a custom Bayesian Network based on domain knowledge. We then applied Tree Search to construct an initial structure and refined it using Hill Climbing. Additionally, we explored two variations of Hill Climbing: one purely score-based and another combining score and constraint-based optimization. The results demonstrate that these structured approaches improve predictive accuracy compared to a manually defined Bayesian Network.

Links to external resources

`https://archive.ics.uci.edu/ml/machine-learning-databases/forest-fires/forestfires.csv`

References

Pacheco, F.; Fernandes, P. M.; Moreira, F.; Rossa, C. G.; Vallejo, V. R. (2015). Assessing the probability of wildfire occurrence in Portugal using a Bayesian Network approach. *Environmental Management*, **55**(2):397–410. https://doi.org/10.1007/s00267-014-0413-5.

Cortez, P.; Morais, A. D. J. R. (2007). A data mining approach to predict forest fires using meteorological data. *New Trends in Artificial Intelligence*, **3**:512–523. https://doi.org/10.1007/s00521-008-0236-3.