Energy Infrastructure Risk Modeling

Assignment 6

ITWS 6600 – Data Analytics
Dec 8, 2023

Neelanga Thelasingha
Electrical Computer and Systems Engineering
Rensselaer Polytechnic Institute

Contents

Chapter 1

Introduction

Energy infrastructure consists of a complex network of components; any failure would have critical consequences for society. To properly assess and mitigate risks to energy infrastructure, it is essential to consider potential hazards [14, 15]. Severe weather events, such as hurricanes, tornadoes, snowstorms, earthquakes, floods, wildfires, volcanic eruptions, and landslides, pose substantial threats [10]. Additionally, non-natural hazards, including cyberattacks, physical attacks (such as terrorism), technical failures, human errors, and supply chain disruptions, further compound the challenges faced by the energy sector [9].

Understanding the interdependencies and interactions among these hazards is essential, as they often exhibit self-excitation and mutual excitation[13]. For instance, earthquakes increase the likelihood of aftershocks[11], and rainstorms elevate the probability of floods or landslides. Recognizing and modeling these dynamic relationships is critical for risk assessment.

To this end, Bayesian networks[12] are a powerful tool for modeling the vulnerability of energy infrastructure to various hazards and their subsequent influence on failures. Prior research has leveraged Bayesian networks as vulnerability models, presented by works such as [16] and [8]. This project builds upon this foundation, proposing a Bayesian network model that captures the influence flow of hazards through the vulnerability, subsequently triggering additional hazards.

Further, Vulnerability is defined as the conditional probability of a failure occurring given the presence of a hazard. By comprehensively modeling these vulnerabilities, we gain valuable insights into how the influence of hazards propagates through the infrastructure, potentially setting off a cascade of failures. In this report, we describe the methodology employed, present the results of our risk models, and compare them against state-of-the-art methods like decision trees and neural networks.

Chapter 2

Data and Model Descriptions

2.1 Geographical Region

In this work, we aim to construct the risk model by modeling the hazards and vulnerabilities present in the energy infrastructure. We selected the geographical area of the northeastern USA for our analysis as shown in Figure 2.1. The region represents the states of NY, NJ, PA, CT, MA, ME, VT, NH, RI.

We selected northeast US as it presents a mixture of energy resources and a diverse set of regions of socio-economical scale from metropolitan areas to rural regions. Further, the northeast is vulnerable to severe weather through snowstorms and rainstorms, while a minor earthquake risk is also present. The availability of verified publicly available data is also an important factor in region selection.

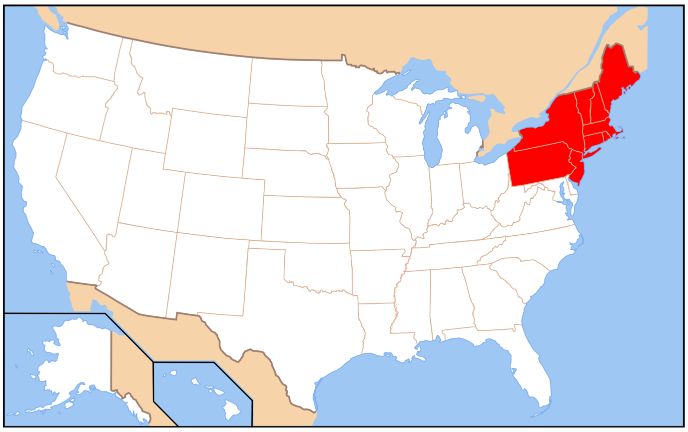

Figure 2.1: Geographical Area [1]

2.2 Data Sets

Given the area, we look at the following 6 datasetss related to hazards and vulnerabilities for our analysis.

2.2.1 Energy Infrastructure Data

We use Energy Infrastructure data to collect the power plant locations in the north-east region. This is obtained from **Homeland Infrastructure Foundation-Level Data** [2]. This consists of the Latitude and Longitutdes cooridantes of all power plants in north-east US including the generation type and more information as shown in Figure 2.2.

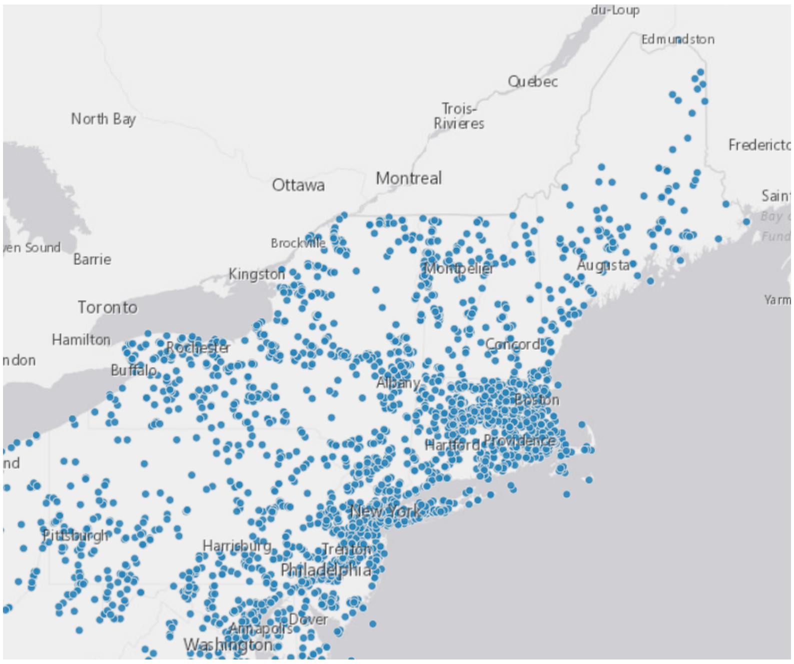

Figure 2.2: Power plant locations

2.2.2 Precipitation Data

We use Precipitation data to collect the historical rain storm and snow storm data in the north-east region. This is obtained from **NASA POWER DATA** [3]. This consists of the historical precipitation information for each 0.25 Latitude and 0.25 Longitutdes coordinates of the northeast USA for the past 4 years.

2.2.3 Earthquake Data

We use Earthquake Occurrence data to collect the historical intensity of earthquakes recorded in the north-east region. This is obtained from **U.S Geological Survey: Earthquake Hazard Program** [4]. This consists of the historical earthquake information for each occurred with relevant latitude and longitude coordinates of the northeast USA for the past 4 years.

2.2.4 Disaster Declarations Data

We use Disaster Declarations data to collect the historical declarations of statewide disasters recorded in the north-east region. This is obtained from **FEMA Web Disaster Declarations** [5]. This consists of the historical disaster declaration information for each that occurred with relevant diaster type at the northeast.

2.2.5 Power Disturbances

We use power disturbances data to analyze how the historical disturbances have occurred in north-east region. This data is obtained from **ISER Electric Disturbance Events** [6]. This consists of the historical Power Disturbances information for each event occurred.

2.2.6 Fire Occurrence

We use Fire Occurrence data to analyze how the historical wild and man-made fires have occurred in the north-east region. This data is obtained from **National Interagency Fire Center** [7]. This consists of information on the location, fire type and timeline for each event that occurred.

2.3 Models

In this project we employ 4 different models on the above data for modeling the risk as follows

2.3.1 Bayesian Network

Bayesian networks are probabilistic graphical models represented by directed acyclic graphs that express probabilistic dependencies among variables. Nodes represent random variables, and edges indicate conditional dependencies. They are applied in diverse fields such as medical diagnosis, risk assessment, and natural language processing. Learning algorithms enable the extraction of network structures and parameter estimations from data 2.3.

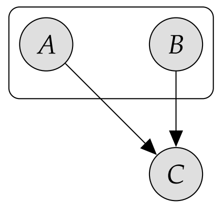

Figure 2.3: Bayesian Network Example

2.3.2 Decision Trees

Decision trees are represented by a tree structure. They make decisions based on features, splitting data to maximize information gain. Leaf nodes represent final outcomes and they are used mainly for classification tasks. Their versatility and visual simplicity make them widely applicable in various predictive modeling scenarios.

2.3.3 Deep Neural Network

Neural networks are computational models that consist of interconnected nodes, or neurons, organized in layers. These networks are trained on data to recognize patterns and make predictions. Deep neural networks have applications in various fields, including artificial intelligence, natural language processing, and computer vision. Their ability to learn and adapt from data makes them powerful tools for data analysis.

2.3.4 Naive- Bayes Classifier

Naive Bayes is a probabilistic learning algorithm used for classification, based on Bayes' theorem, and assumes independence between features. It calculates the posterior probability of a data point belonging to a certain class given the evidence and selects the class with the highest probability.

Chapter 3

Analysis

This section discusses how the datasets were pre-processed for the analysis. We create two specific datasets for analysis tasks.

3.1 Dataset 1

We use Dataset 1 to encode geographical features to enable predictions of vulnerability conditioned on the geo-location. In this dataset we have the following features.

3.1.1 Geographical Location

We consider a grid tiling of the north east region by 0.5×0.5 degrees of latitudes and longitudes. For each grid point, we evaluate the next features.

3.1.2 State

We use the lat, long coordinate of each grid square to assign them to the corresponding state and we create a column with state information for each point.

3.1.3 Generation Type

For each geo-location given by 0.5×0.5 square on the grid, we evaluate the power generation and assign the majority generation as the generation type of that block. The histogram of generation type distribution across the north east shown in Figure 3.1.

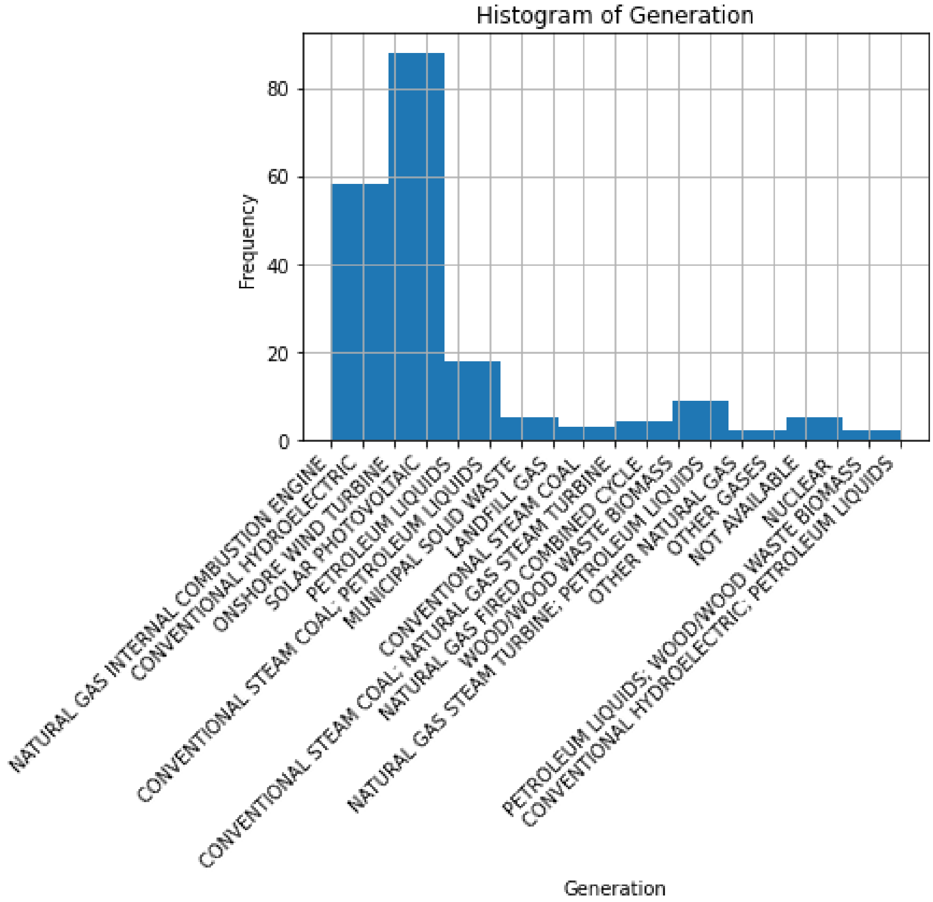

Figure 3.1: Majority generation distribution

3.1.4 Earthquake Risk

For each geo-location given by 0.5×0.5 square on the grid, we evaluate whether an earthquake has happened at this location and assign 1 or 0 to create a binary feature column. The distribution is shown in Figure 3.2.

3.1.5 Precipitation

For each geo-location given by 0.5×0.5 square on the grid, we evaluate the average daily precipitation to create a continuous feature vector. The initial distribution is shown in Figure 3.3. As the learning models we use are better suited for discrete variables, we quantize the continuous feature to bins created by the quartiles. This distribution is shown in the Figure 3.4.

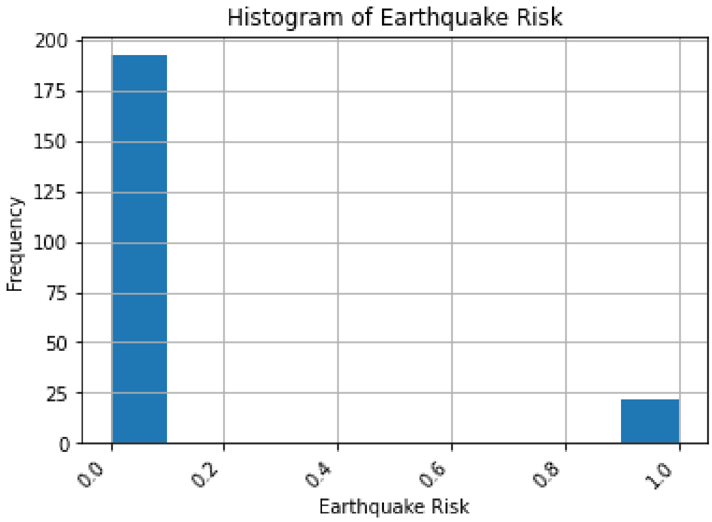

Figure 3.2: Earthquake risk distribution

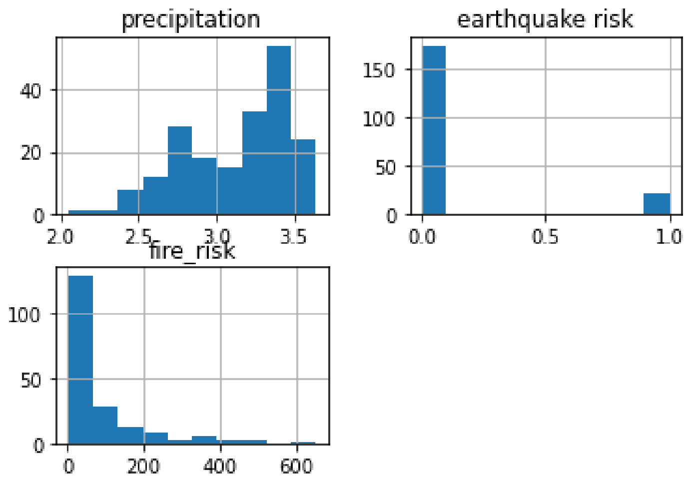

Figure 3.3: Continuous variable distribution- Dataset 1

3.1.6 Fire Risk

For each geo-location given by 0.5×0.5 square on the grid, we evaluate the fire risk by looking at the number of reported fires at each grid location historically. This creates a continuous feature vector. The initial distribution is shown in Figure 3.3. As the models we use are better suited for discrete variables, we quantize the continuous feature to bins created by the quartiles. This distribution is shown in the Figure 3.4.

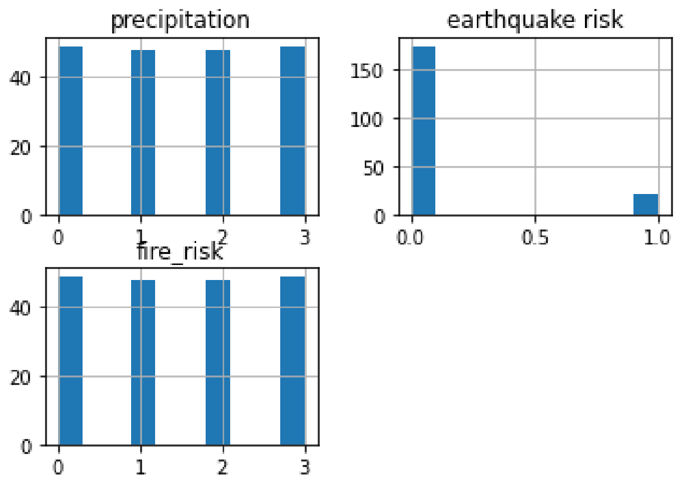

Figure 3.4: After quantization - Dataset 1

3.2 Dataset 2

We use Dataset 2 to encode time-dependent seasonality into features to enable predictions of hazards conditioned on the location given by state and the month of the year. In this dataset, we have the following features.

3.2.1 Month

We consider the month of the year as the anchor point to evaluate the rest of the features. This enables to look at the temporal and seasonal aspects of the data, while Dataset 1 looks at the spatial distribution.

3.2.2 State

For each of the following features, we use the corresponding state and create a column with state information for each point as information such as disaster declarations are mostly state-wide.

3.2.3 Disaster Types

From the disaster declrations from FEMA, we create the this feature. We also add the corresponding month and the state to the dataset as well. The histogram of Disaster types distribution across the north east shown in Figure 3.5. Further, Figure 3.6 Disaster declaration distribution across month and states.

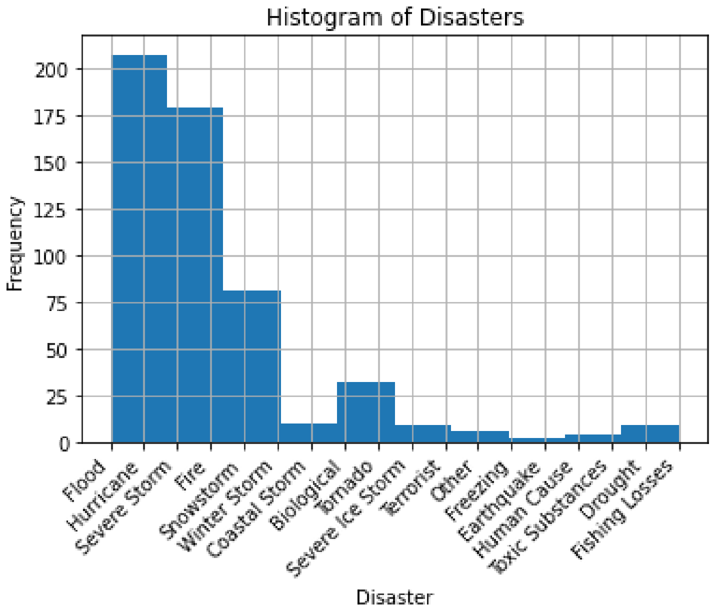

Figure 3.5: Disaster declaration distribution

3.2.4 Breakdown Rate

For each month and state combination, we evaluate the average breakdown rate and add this as a feature. The initial distribution is shown in Figure 3.7. As the learning models we use are better suited for discrete variables, we quantize the continuous feature to bins created by the quartiles. This distribution is shown in the Figure 3.8.

3.2.5 Precipitation

For each month and state combination, we evaluate the average daily precipitation to create a continuous feature vector. The initial distribution is shown in Figure 3.7. As the learning models we use are better suited for discrete variables, we quantize the continuous feature to bins created by the quartiles. This distribution is shown in the Figure 3.8.

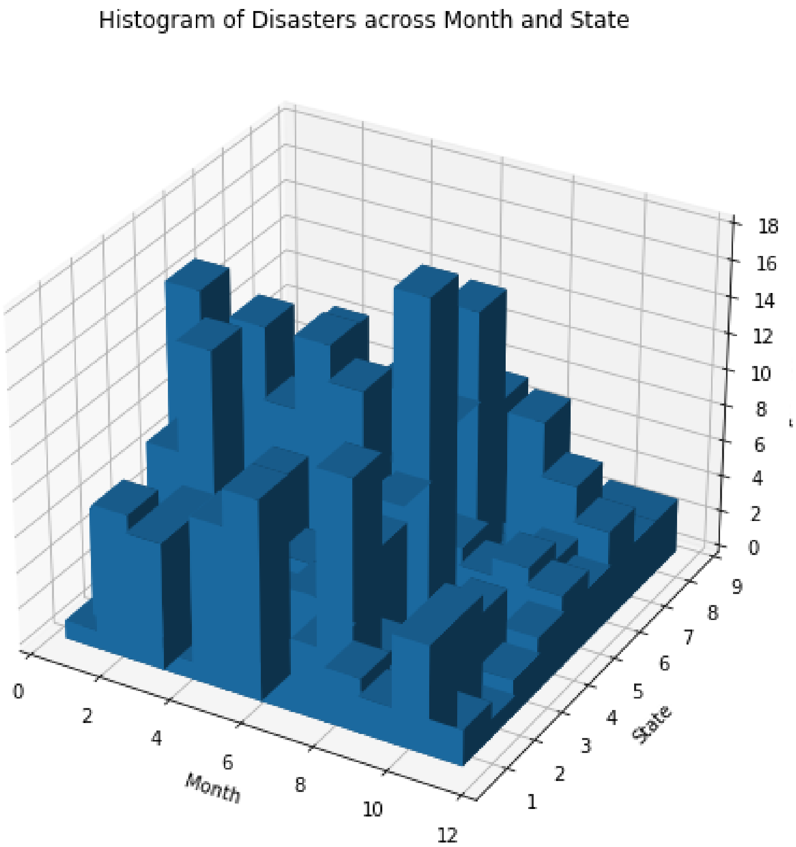

Figure 3.6: Disaster declaration distribution across month and state

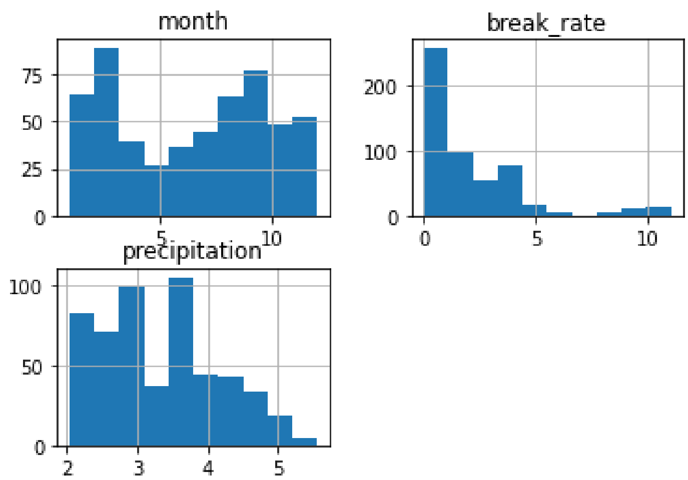

Figure 3.7: Continuous variable distribution - Dataset 2

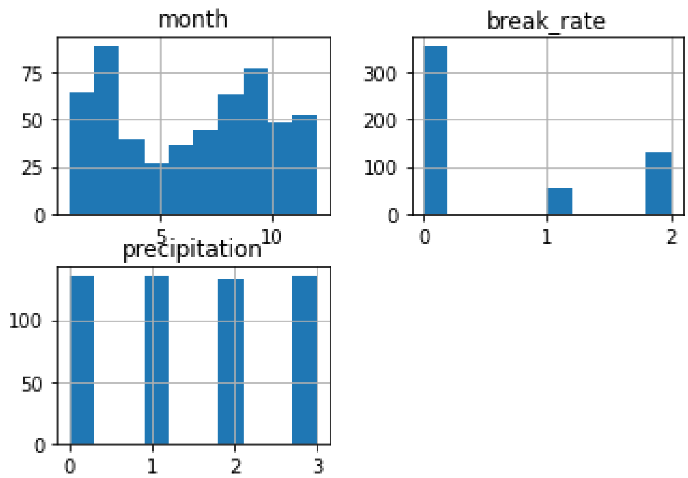

Figure 3.8: After quantization - Dataset 2

Chapter 4

Model Development

In this section we describe the model development for analysis. We use the following models for the two datasets.

For Dataset 1:

1. Learn conditional probability Distributions - Bayesian Network
2. Classify fire risk - Decision Tree
3. Classify fire risk - Naive Bayes Classifier
4. Classify fire risk - Deep Neural Network
5. Classify earthquake risk - Decision Tree
6. Classify earthquake risk - Naive Bayes Classifier
7. Classify earthquake risk - Deep Neural Network

For Dataset 2:

1. Learn conditional probability Distributions - Bayesian Network
2. Classify breakdown rate - Decision Tree
3. Classify breakdown rate - Naive Bayes Classifier
4. Classify breakdown rate - Deep Neural Network

4.1 Dataset 1

4.1.1 Learn conditional probability Distributions - Bayesian Network

We design the structure of the Bayesian network and then employ parameter estimation. The used structure is shown in Fig. 4.1.

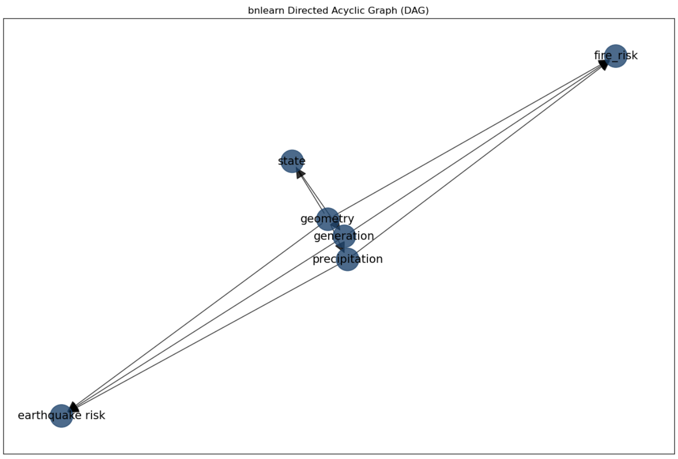

Figure 4.1: Bayesian network structure - Dataset 1

The conditional probability distributions are shown in the following tables.

geometry	geometry(0)	...	geometry(186)
state(0)	0.1204	...	0.1204
state(1)	0.1204	...	0.1204
state(2)	0.2779	...	0.1204
state(3)	0.1204	...	0.1204
state(4)	0.1204	...	0.1204
state(5)	0.1204	...	0.2779
state(6)	0.1204	...	0.1204

state	...	state(6)
generation(0)	...	0.1821
generation(1)	...	0.2459
generation(2)	...	0.1821
generation(3)	...	0.1821
generation(4)	...	0.2077

geometry	...	geometry(186)
precipitation(1)	...	0.3682
precipitation(2)	...	0.2106
precipitation(3)	...	0.2106
precipitation(4)	...	0.2106

generation	...	generation(4)
geometry	...	geometry(186)
precipitation	...	precipitation(4)
fire risk(1)	...	0.25
fire risk(2)	...	0.25
fire risk(3)	...	0.25
fire risk(4)	...	0.25

generation	...	generation(4)
geometry	...	geometry(186)
precipitation	...	precipitation(4)
earthquake risk(0)	...	0.5
earthquake risk(1)	...	0.5

4.1.2 Classify fire risk - Decision Tree

We use a decision tree to learn the classification of fire risk variable. The trained structure is shown in Fig. 4.2. The table 4.1 shows the classification result.

Table 4.1: Classification Report – Decision Tree – fire risk – Dataset 1

	Precision	Recall	F1-Score	Support
0	0.29	0.25	0.27	8
1	0.31	0.45	0.37	11
2	0.20	0.25	0.22	8
3	0.83	0.42	0.56	12
Accuracy			0.36	39
Macro Avg	0.41	0.34	0.35	39
Weighted Avg	0.44	0.36	0.38	39

4.1.3 Classify fire risk - Naive Bayes Classifier

We use a Naive Bayes Classifier to learn the classification of fire risk variable. The pair plot is shown in 4.3. The table 4.2 shows the classification result.

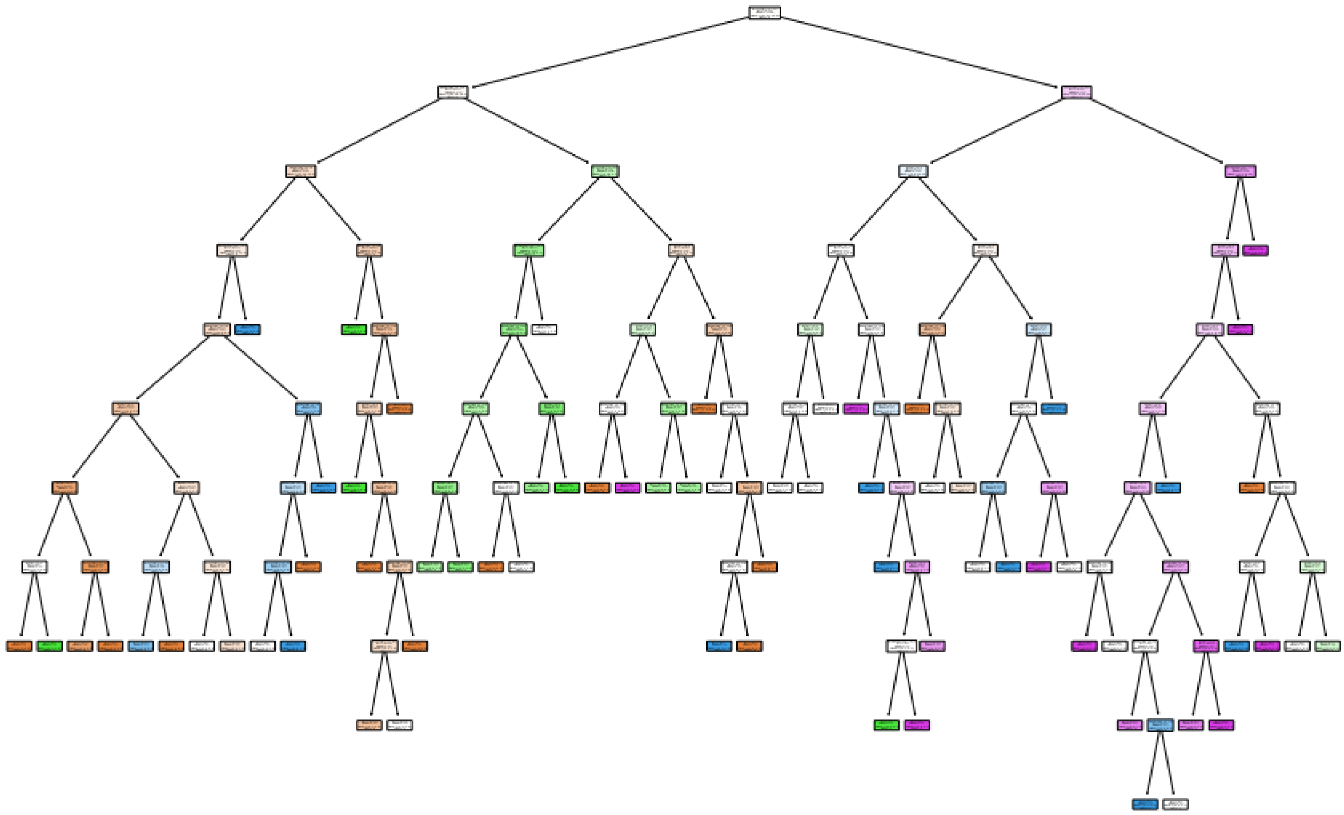

Figure 4.2: Learned Decision Tree – fire risk – Dataset 1

4.1.4 Classify fire risk - Deep Neural Network

We use a Deep Neural Network to learn the classification of fire risk variable. The neural net structure is the same as shown in Fig. 4.4. The table 4.3 shows the classification result.

4.1.5 Classify Earthquake Risk - Decision Tree

We use a decision tree to learn the classification of Earthquake Risk variable. The trained structure is shown in Fig. 4.5. The table 4.4 shows the classification result.

4.1.6 Classify Earthquake Risk - Naive Bayes Classifier

We use a Naive Bayes Classifier to learn the classification of Earthquake Risk variable. The pair plot is shown in 4.6. The table 4.5 shows the classification result.

Table 4.2: Classification Report – fire risk – Naive Bayes Classifier– Dataset1

	Precision	Recall	F1-Score	Support
0	0.40	0.50	0.44	8
1	0.33	0.36	0.35	11
2	0.00	0.00	0.00	8
3	0.69	0.75	0.72	12
Accuracy			0.44	39
Macro Avg	0.36	0.40	0.38	39
Weighted Avg	0.39	0.44	0.41	39

Table 4.3: Classification Report – fire risk - Deep Neural Network – Dataset 1

	Precision	Recall	F1-Score	Support
0	0.24	0.50	0.32	8
1	0.33	0.27	0.30	11
2	0.22	0.25	0.24	8
3	0.75	0.25	0.38	12
Accuracy			0.31	39
Macro Avg	0.39	0.32	0.31	39
Weighted Avg	0.42	0.31	0.31	39

Table 4.4: Classification Report – Decision Tree – Earthquake Risk – Dataset 1

	Precision	Recall	F1-Score	Support
0	0.91	0.83	0.87	35
1	0.14	0.25	0.18	4
Accuracy			0.77	39
Macro Avg	0.52	0.54	0.52	39
Weighted Avg	0.83	0.77	0.80	39

Table 4.5: Classification Report – Earthquake Risk – Naive Bayes Classifier– Dataset1

	Precision	Recall	F1-Score	Support
0	0.90	1.00	0.95	35
1	0.00	0.00	0.00	4
Accuracy			0.90	39
Macro Avg	0.45	0.50	0.47	39
Weighted Avg	0.81	0.90	0.85	39

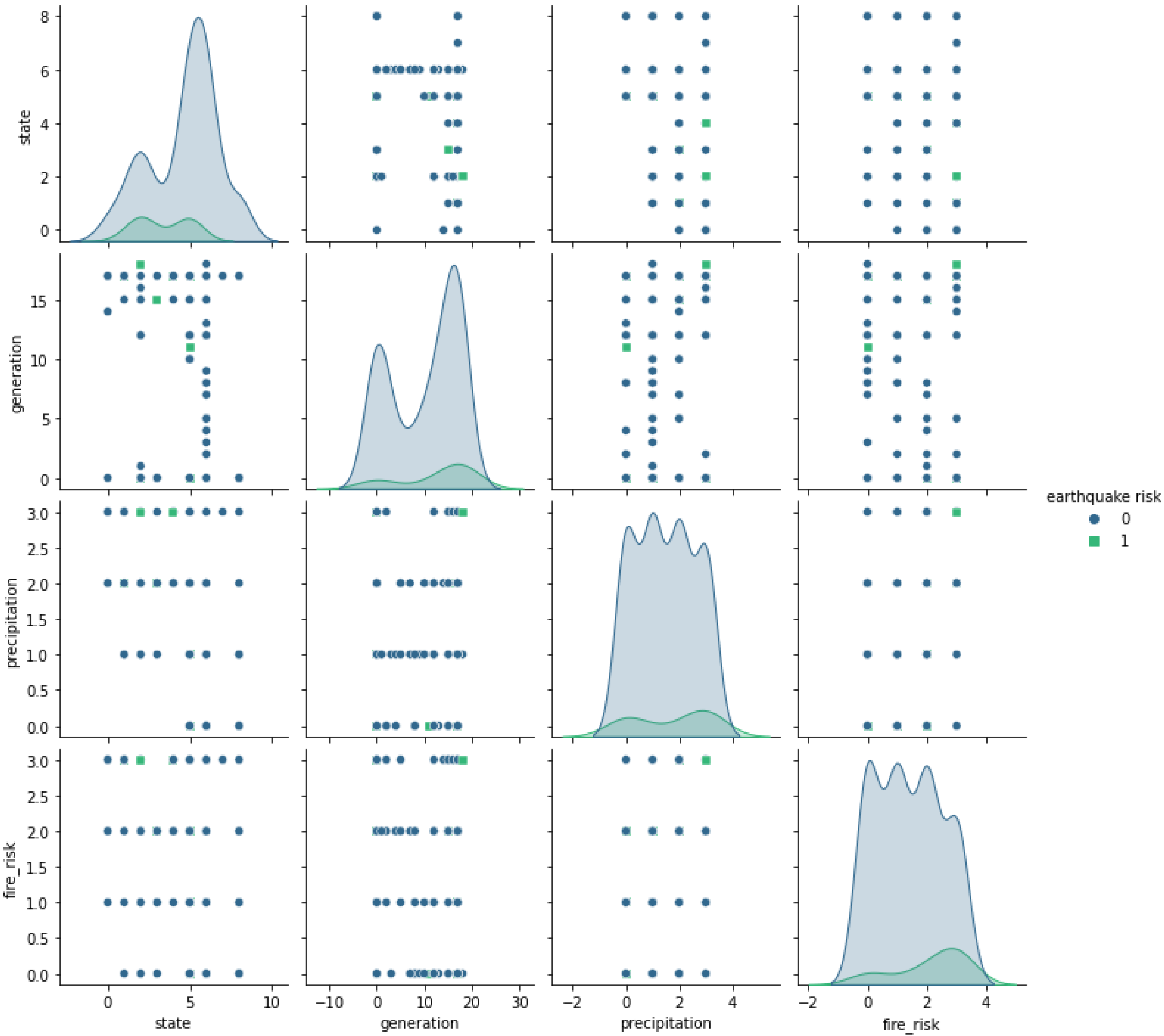

Figure 4.3: Pairs Plot – fire risk – Dataset 1

4.1.7 Classify Earthquake Risk - Deep Neural Network

We use a Deep Neural Network to learn the classification of Earthquake Risk variable. The neural net structure is shown in Fig. 4.4. The table 4.6 shows the classification result.

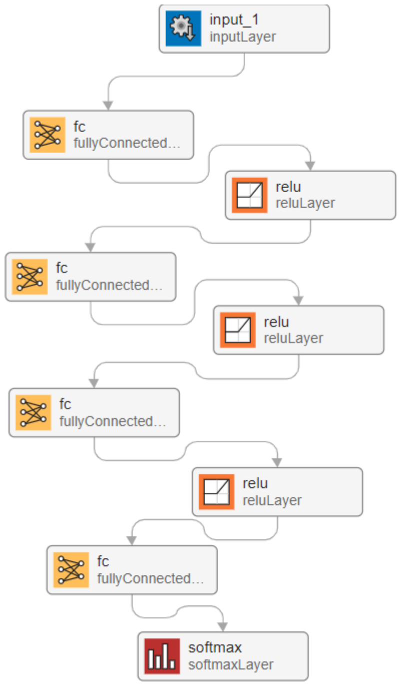

Figure 4.4: Deep Neural Network Structure

4.2 Dataset 2

4.2.1 Learn conditional probability Distributions - Bayesian Network

We design the structure of the Bayesian network and then employ parameter estimation. The used structure is shown in Fig. 4.7.

The conditional probability distributions are shown in the following tables. For month,

month(1)	0.0788391
month(2)	0.0853368
month(3)	0.0697423
month(4)	0.0723413
month(5)	0.0710418
month(6)	0.111977
month(7)	0.0794888
month(8)	0.0716916
month(9)	0.0775395
month(10)	0.0827377
month(11)	0.0950834
month(12)	0.10418

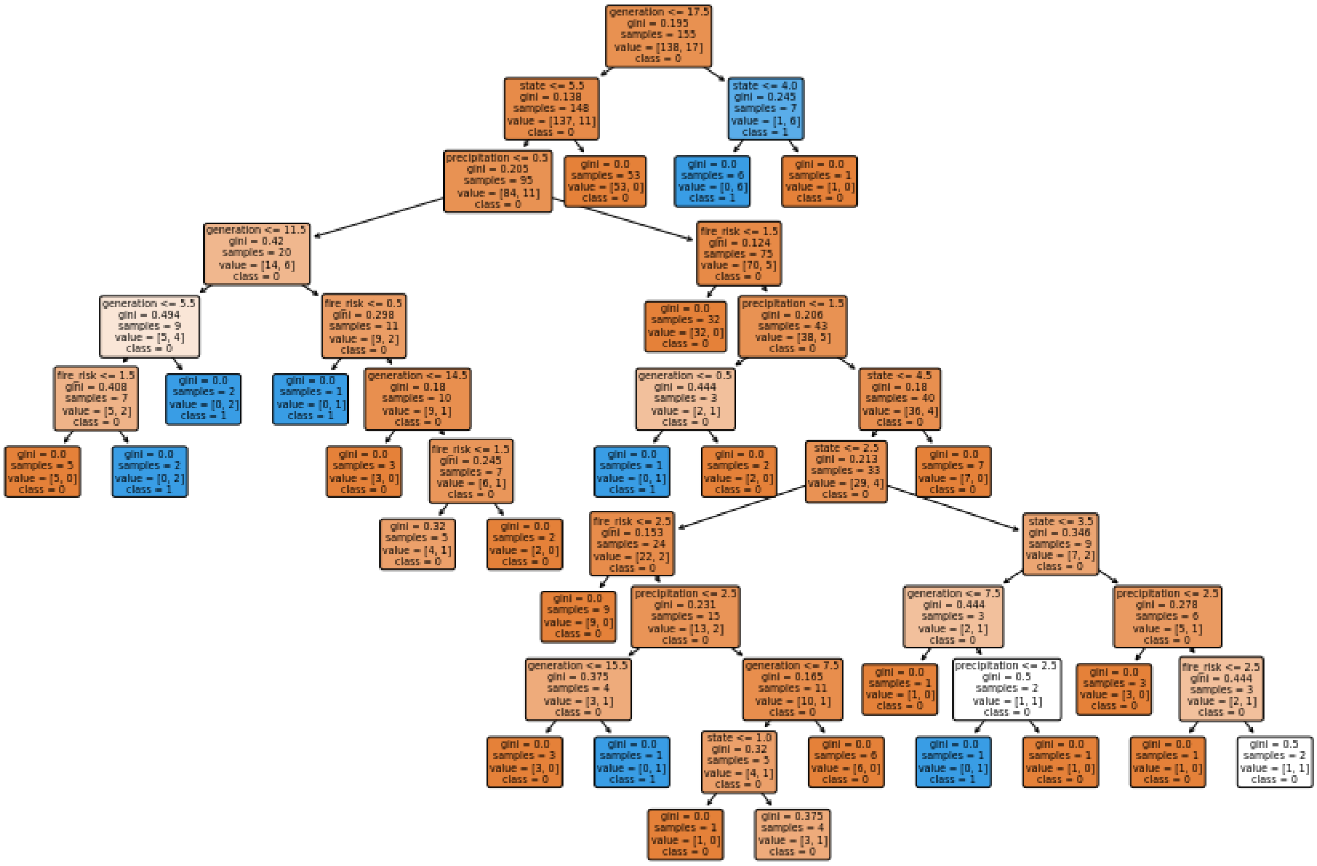

Figure 4.5: Learned Decision Tree – Earthquake Risk – Dataset 1

For disaster,

month	state(1)	...	state(9)
disaster(0)	0.09441087613293052	...	0.13268792710706148
disaster(1)	0.09441087613293052	...	0.07118451025056946
disaster(2)	0.09441087613293052	...	0.07118451025056946
disaster(3)	0.09441087613293052	...	0.07118451025056946
disaster(4)	0.09441087613293052	...	0.19419134396355348
disaster(5)	0.17598187311178248	...	0.13268792710706148
disaster(6)	0.09441087613293052	...	0.2556947608200455
disaster(7)	0.25755287009063443	...	0.07118451025056946

For state

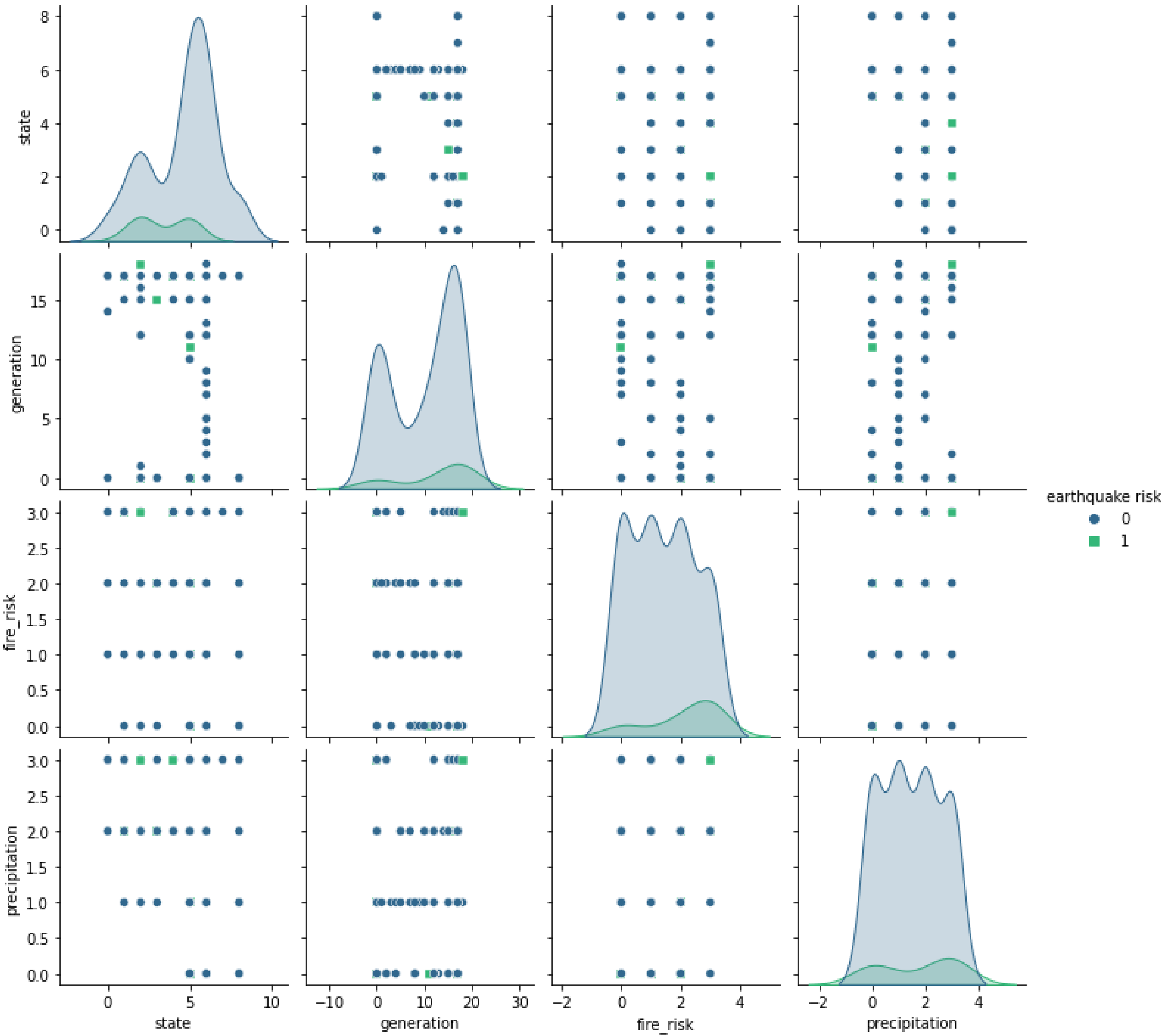

Figure 4.6: Pairs Plot – Earthquake Risk – Dataset 1

state(1)	0.0975381
state(2)	0.109234
state(3)	0.116381
state(4)	0.110534
state(5)	0.109884
state(6)	0.144322
state(7)	0.113133
state(8)	0.0903906
state(9)	0.108584

Table 4.6: Classification Report – Earthquake Risk - Deep Neural Network – Dataset 1

	Precision	Recall	F1-Score	Support
0	0.90	1.00	0.95	35
1	0.00	0.00	0.00	4
Accuracy			0.90	39
Macro Avg	0.45	0.50	0.47	39
Weighted Avg	0.81	0.90	0.85	39

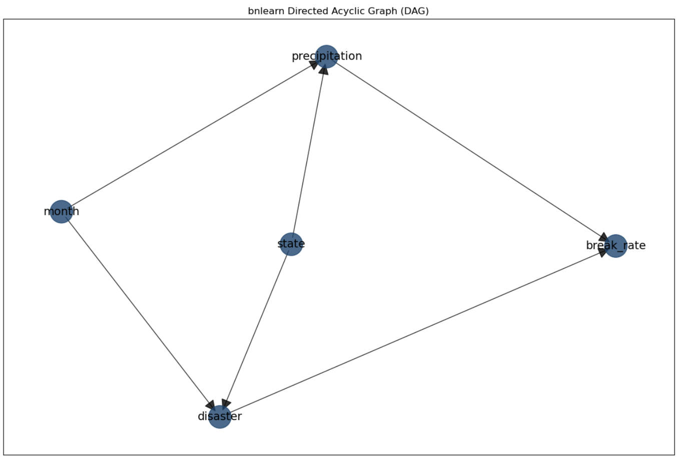

Figure 4.7: Bayesian network structure - Dataset 2

For precipitation

month	...	month(12)
state	...	state(9)
precipitation(1)	...	0.14236902050113895
precipitation(2)	...	0.14236902050113895
precipitation(3)	...	0.5728929384965832
precipitation(4)	...	0.14236902050113895

For the Breakdown rate,

Disaster	...	Disaster(7)
Precipitation	...	Precipitation(4)
Break_Rate(1)	...	0.36253041362530414
Break_Rate(2)	...	0.30413625304136250
Break_Rate(3)	...	0.3333333333333333

4.2.2 Classify Breakdown Rate - Decision Tree

We use a decision tree to learn the classification of Breakdown Rate variable. The trained structure is shown in Fig. 4.8. The table 4.7 shows the classification result.

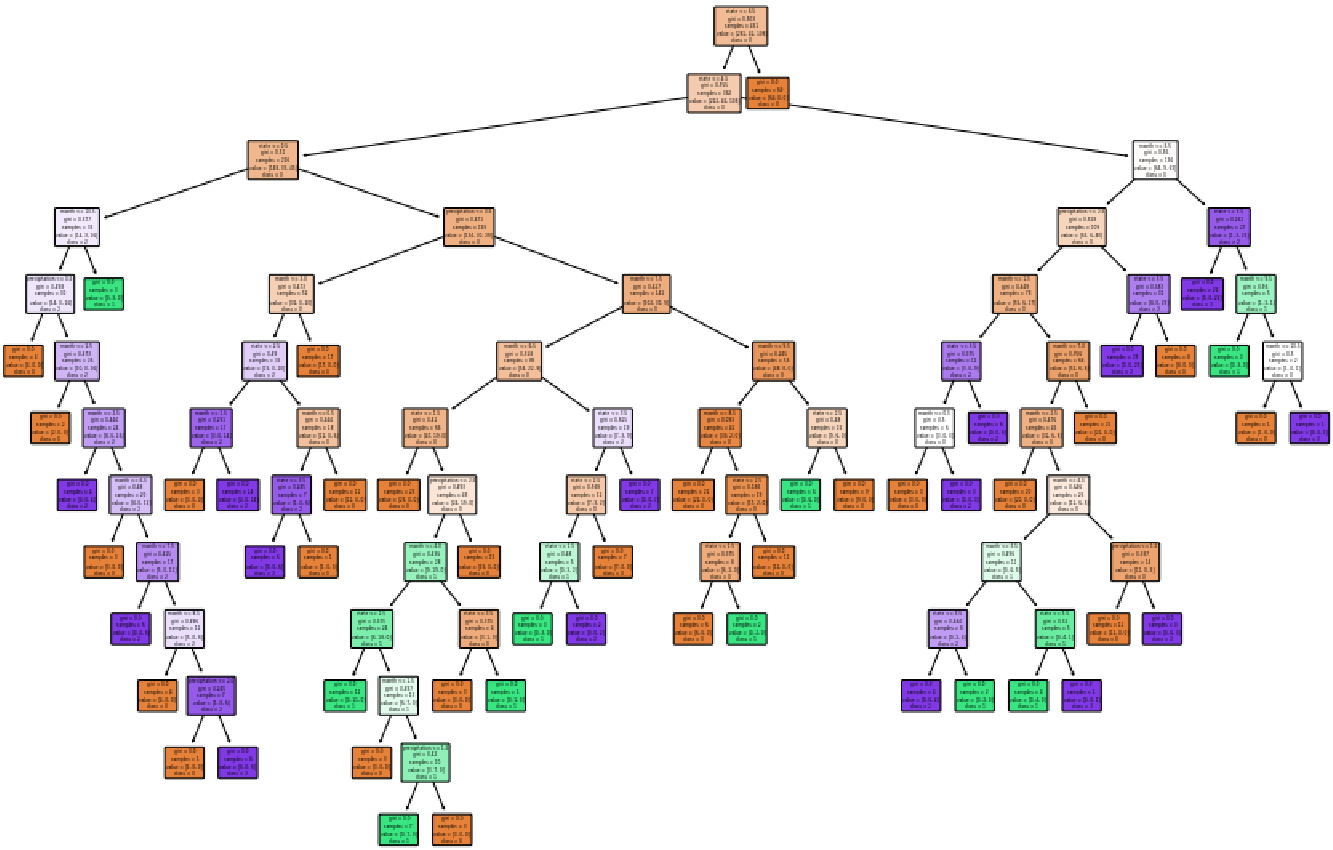

Figure 4.8: Learned Decision Tree – Breakdown Rate – Dataset 2

Table 4.7: – Breakdown Rate – Decision Tree – Dataset2

	Precision	Recall	F1-Score	Support
0	1.00	0.99	0.99	75
1	0.92	1.00	0.96	11
2	1.00	1.00	1.00	22
Accuracy			0.99	108
Macro Avg	0.97	1.00	0.98	108
Weighted Avg	0.99	0.99	0.99	108

4.2.3 Classify Breakdown Rate - Naive Bayes Classifier

We use a Naive Bayes Classifier to learn the classification of Breakdown Rate variable. The pair plot is shown in 4.9. The table 4.8 shows the classification result.

Table 4.8: Classification Report – Breakdown Rate – Naive Bayes Classifier– Dataset 2

	Precision	Recall	F1-Score	Support
0	0.69	0.99	0.81	75
1	0.00	0.00	0.00	11
2	0.00	0.00	0.00	22
Accuracy			0.69	108
Macro Avg	0.23	0.33	0.27	108
Weighted Avg	0.48	0.69	0.56	108

4.2.4 Classify Breakdown Rate - Deep Neural Network

We use a Deep Neural Network to learn the classification of Breakdown Rate variable. The neural net structure is the same as shown in Fig. 4.4. The table 4.9 shows the classification result.

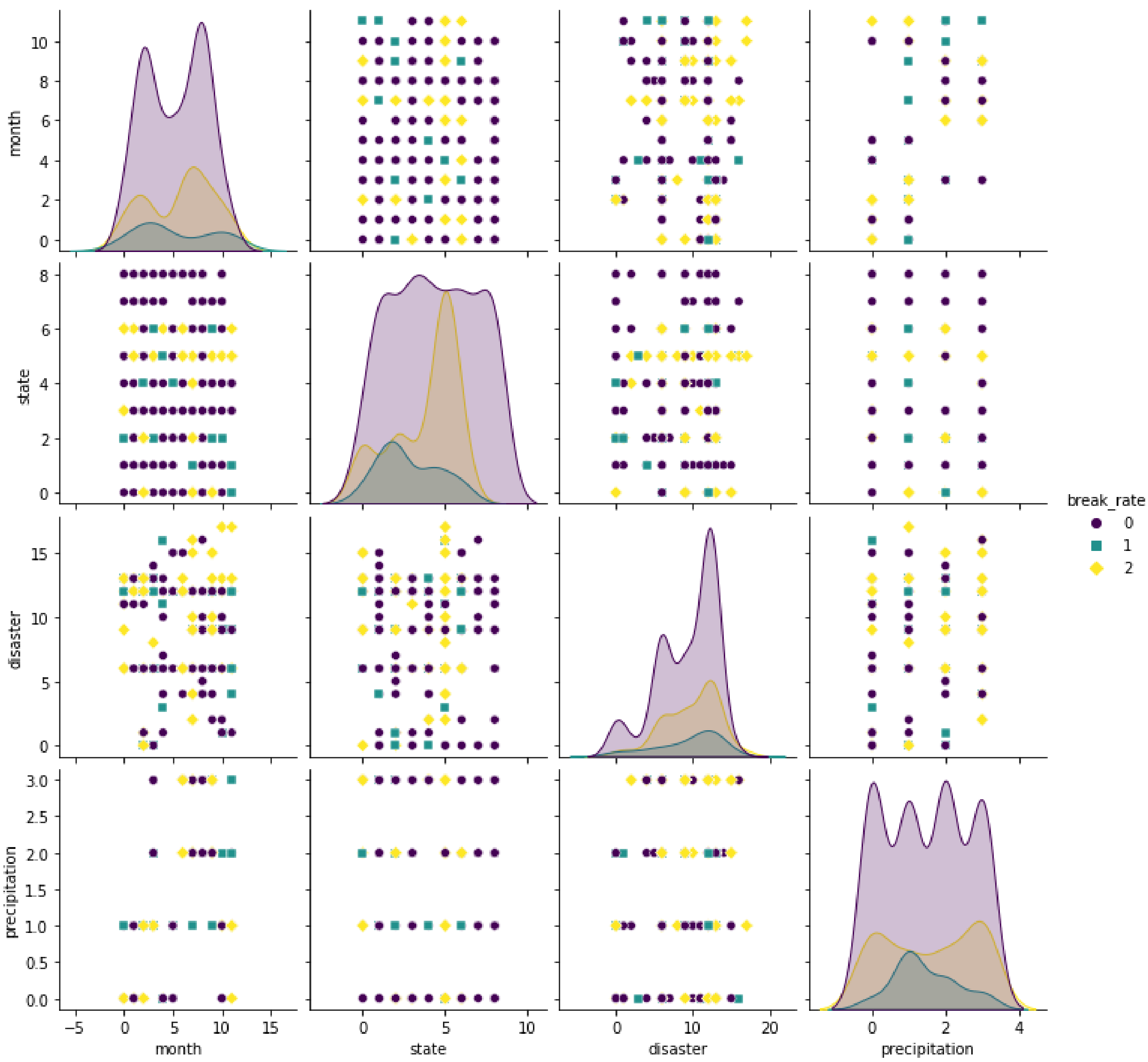

Figure 4.9: Pairs plot – Breakdown Rate – Dataset 2

Table 4.9: Classification Report – Breakdown Rate - Deep Neural Network – Dataset 2

	Precision	Recall	F1-Score	Support
0	0.69	0.99	0.81	75
1	0.00	0.00	0.00	11
2	0.00	0.00	0.00	22
Accuracy			0.69	108
Macro Avg	0.23	0.33	0.27	108
Weighted Avg	0.48	0.69	0.56	108

Chapter 5

Conclusions

Through this work, we were able to explore the impact of hazards affecting critical infrastructure through seasonal severe weather and other events. Further, the vulnerability presented by geo-locations was analyzed as well.

The analysis of Dataset 1 yields the conditional probability distributions of earthquake and fire risk through the Bayesian network. Thus, it can be used as an inference method to evaluate the vulnerability of a geo-location to the energy infrastructure.

Further, the analysis on Dataset 2 gives the ability to infer break down rate conditional on the month of the year, precipitation and disaster classifications through the Bayesian network. This provides useful insights into hazard occurrences and vulnerabilities present in the temporal dimension.

The classification models developed for both data sets present insights into dependencies present in the variables and their predictability. While DNN provided acceptable accuracies, the decision tree model performed well with explainable properties.

We believe this work can be further improved by incorporating the data on transmission lines in the region as well as shown in Fig. 5.1. Further, increasing the spatial and temporal resolution will also contribute to higher accuracy.

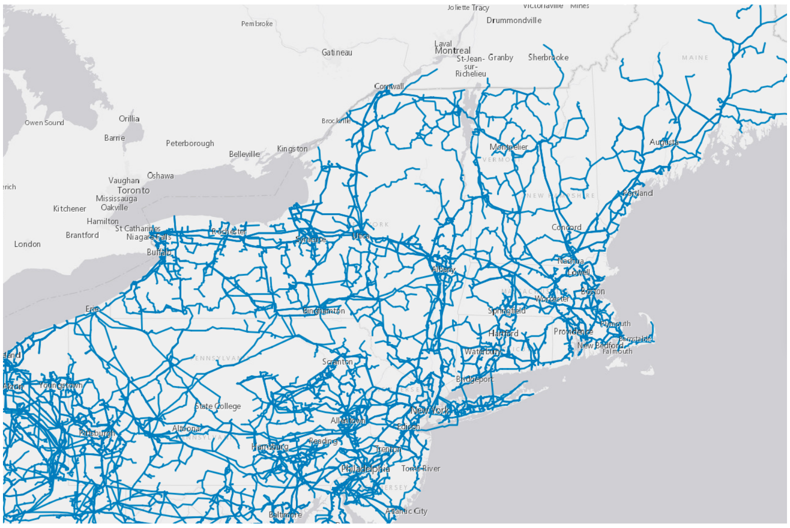

Figure 5.1: Critial Transmission lines

Bibliography

[1] URL: https://commons.wikimedia.org/wiki/File:Map_of_the_North_Eastern_United_States.svg.

[2] URL: https://hifld-geoplatform.opendata.arcgis.com/datasets/geoplatform::power-plants-2/about.

[3] URL: https://power.larc.nasa.gov/.

[4] URL: https://earthquake.usgs.gov/earthquakes/search/.

[5] URL: https://www.fema.gov/openfema-data-page/fema-web-disaster-declarations-v1.

[6] URL: https://www.oe.netl.doe.gov/oe417.aspx.

[7] URL: https://data-nifc.opendata.arcgis.com/datasets/nifc::inform-fire-occurrence-data-records/explore?location=43.590069%2C-71.436469%2C6.78&showTable=true.

[8] Mauricio Gallardo. “Measuring vulnerability to multidimensional poverty with Bayesian network classifiers”. In: *Economic Analysis and Policy* 73 (2022), pp. 492–512.

[9] Oleg Ivanchenko et al. “Risk assessment of critical energy infrastructure considering physical and cyber assets: methodology and models”. In: *2018 IEEE 4th International Symposium on Wireless Systems within the International Conferences on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS-SWS)*. IEEE. 2018, pp. 225–228.

[10] Mirjam Merz et al. “A composite indicator model to assess natural disaster risks in industry on a spatial level”. In: *Journal of Risk Research* 16.9 (2013), pp. 1077–1099.

[11] Yosihiko Ogata. “Statistics of earthquake activity: Models and methods for earthquake predictability studies”. In: *Annual Review of Earth and Planetary Sciences* 45 (2017), pp. 497–527.

[12] Judea Pearl. “Bayesian networks”. In: (2011).

[13] Ling Tan et al. "Can we detect trends in natural disaster management with artificial intelligence? A review of modeling practices". In: *Natural Hazards* 107 (2021), pp. 2389–2417.

[14] Jing Wang et al. "Literature review on modeling and simulation of energy infrastructures from a resilience perspective". In: *Reliability Engineering & System Safety* 183 (2019), pp. 360–373.

[15] Ype C Wijnia and Paulien M Herder. "Modeling Interdependencies in electricity infrastructure risk". In: *1st Annual CZAEE International Conference "Critical Infrastructure in the energy sector: Vulnerabilities and protection", Prague*. 2004.

[16] Shuai Zhang et al. "Quantification of human vulnerability to earthquake-induced landslides using Bayesian network". In: *Engineering Geology* 265 (2020), p. 105436.