

Failure Probability Metric by Machine Learning for Online Risk Assessment in Distribution Networks

Jonatas Boas Leite
Jose Roberto Sanches Mantovani
Dep. of Electrical Engineering
São Paulo State University - UNESP
Ilha Solteira, Brasil
mant@dee.feis.unesp.br

Tatjana Dokic, Qin Yan
Po-Chen Chen, Mladen Kezunovic
Dep. of Electrical and Computer Engineering
Texas A&M University
College Station, USA
kezunov@ece.tamu.edu

Abstract—The risk assessment approach is useful for monitoring and supervisory control because it provides distribution operator with the capability to quantify the tradeoff between reliability and economic performance. The risk assessment determines the likelihood of something going wrong in a distribution network through the failure probability metric. To deal with the massive variety of information required in the calculation of failure probability we propose a data mining approach. The proposed approach incorporates weather, asset and outage information for characterizing the risk in a distribution network section via GIS platform.

Keywords—electricity supply industry, failure probability, geographic information systems, power distribution, risk analysis;

I. INTRODUCTION

The online risk analysis for power system must be able to provide valuable knowledge that could increase the operation resilience against unpredictable rare events. This can be achieved by integrating the reliability analysis and distribution automation in a framework that aims to minimize the energy flow disruption impacts and reduce human intervention [1]. The quantitative and/or qualitative data, required by the decision-making process for risk control answer questions like: what is the likelihood that something will go wrong, and if it goes wrong, what are the consequences [2]?

The risk assessment is more widely employed in the power transmission system [3]-[4]. Most recent literature for the power distribution system also studies the risk as the main concern. For example, a risk-based security assessment is employed for the mid- and long-term planning for connecting the concentrated solar power in distribution networks in [5]. Authors, in [6], use long outage compensation and traditional reliability indices to perform the risk vulnerability analysis in order to effectively allocate both human and economic resources and equipment in power distribution system. For the purpose of investment allocation, [7] ranks risk values for prioritizing predictive actions on underground electricity distribution system using a multidimensional assessment. More recently, an investigation of the impact of severe weather condition on power distribution system is reported in [8], where the proposed risk assessment approach employs a combination of sequential Monte Carlo simulation with wind probability models from historical database and outage records. The mitigation of weather impacts against power distribution

system using risk analysis is proposed in [9] with a fuzzy logic approach for improving the decision-making process of outage management system (OMS) through a vulnerability analysis based on geographic information system (GIS) [10].

The application of risk assessment approach has improved the planning and operating procedures in power distribution system. In [11], a risk assessment methodology incorporates the weather impact into customer interruption cost (CIC) to differentiate categories of distribution network loads, such as health care centers and schools, in addition to traditional customers. To predict damages and its consequences due to weather impacts, the approach proposed in [12] integrates fault management events into interruption cost using a time series-based formulation for supporting the spatio-temporal data. The approach proposed in our current paper goes to the next step by addressing the failure probability, as illustrated in Fig. 1. Since the threats characterization involves processing of large volume of data coming from different sources, *i.e.* weather stations, lighting detection network, GIS, OMS and asset management system (AMS) [13]-[14], it is evident that the use of the big data analytics is required where the machine learning techniques demonstrate great efficiency in the knowledge extraction.

The paper organization is as follows. Section II provides the mathematical model of the probability estimator. Section III presents the data flow for implementing the machine learning technique as well as explanation of the vulnerability models. Section IV groups the failure probabilities in six classes using the Jenks natural breaks classification method, and then evaluates the proposed methodology using a real distribution network data. Section IV lists paper contributions.

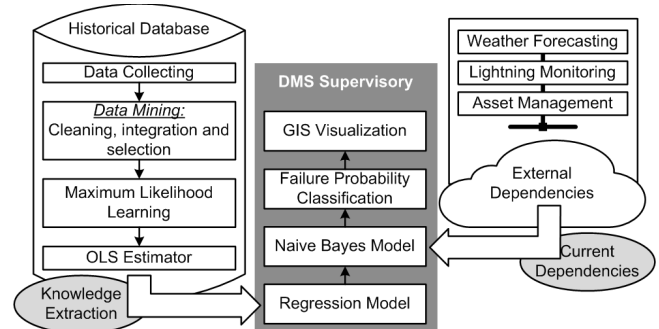


Fig. 1. Complete cycle of the failure probability metric.

This work was fully supported by the São Paulo Research Foundation – FAPESP (grant: 2015/17757-2 and 2015/21972-6), CAPES and CNPq (grant: 305371/2012-6) allowing Dr. Boas Leite to spend a year with Dr. Kezunovic's Texas A&M research team.

II. NAIVE BAYES PROBABILITY ESTIMATION

The Naive Bayes is one of the most popular density estimation method which reached its fame because it is very simple to program and intuitive, fast to train and very easy to deal with missing attributes [15]. This method is a special form of Bayesian network where the value of a particular feature is conditionally independent of the value of any other feature of interest. Although Naive Bayes models are frequently applied for clustering and classification, its application for other probabilistic models, such as general probabilistic learning and inference, remains unexploited [16]. In practice, the weak correlation among the features of interest that form the independence model may be quite a realistic model in many real situations [17]. In power distribution system, the features of interest are external dependences, *i.e.* different types of threats that can jointly provide the failure probability using a Naive Bayes model as given by (1) and (2).

$$p(f | \mathbf{X}) = \frac{\tilde{p}(\mathbf{X} | f) \tilde{p}(f)}{\sum_f \tilde{p}(\mathbf{X} | f) \tilde{p}(f)} \quad (1)$$

$$\tilde{p}(\mathbf{X} | f) = \prod_{i=1}^D (\tilde{\theta}_i^f)^{x_i} (1 - \tilde{\theta}_i^f)^{1-x_i} \quad (2)$$

The failure feature has two states, $\text{dom}(f) = \{0,1\}$, thus the value of $p(f=1 | \mathbf{X})$ is the conditional probability of failure occurrence subject to $\mathbf{X} = \{x_i | \text{dom}(x_i) = \text{dom}(f) \wedge i = 1, \dots, D\}$ of external dependencies that are listed in Table I.

Due to Naive Bayes conditional independence assumption, the probability of observing the vector \mathbf{X} can be compactly written as in (2) where $p(x_i=1 | f) \equiv \theta_i^f$ and, consequently, $p(x_i=0 | f) \equiv 1 - \theta_i^f$. The conditional independence also allows the application of the maximum likelihood learning to the Naive Bayes model.

$$p(x_i=1 | f) = \frac{\text{number times } x_i=1 \text{ for } f}{\text{number of data points in } f} \quad (3)$$

Equation (3) is used in the proposed model for extracting knowledge from the available databases by obtaining the monthly likelihood of each feature of interest. Fig. 1 shows the medium wind speed probability for failure occurrence in January and April during seven years, from 2009 to 2015. In January, the maximum likelihood learning provides annual probabilities surrounding the average value while, in April, the provided annual probabilities have decreasing behavior along years. This behavior suggests that the average probability is insufficient to reflect all the information from the observed data. The ordinary least squares (OLS) estimator is applied for providing the regression model able to predict the probability value in the current year of analysis as shown in Fig. 2 where the year of analysis is 2016. The regression model resulting from the OLS estimator is given by (4) and (5).

$$\tilde{\theta}_i^f = \mathbf{B}_i^T \mathbf{T} \quad (4)$$

$$\mathbf{B}_i = \left(\sum_y \mathbf{T}_y (\mathbf{T}_y)^T \right)^{-1} \sum_y \mathbf{T}_y \theta_{i,y}^f \quad (5)$$

TABLE I. DESCRIPTION OF EXTERNAL DEPENDENCIES.

x_i	Feature of interest	x_i	Feature of interest
x_1	Low wind speed	x_6	Rainy weather
x_2	Medium wind speed	x_7	Thunderstorm weather
x_3	High wind speed	x_8	Lightning incidence
x_4	Good weather	x_9	Vegetation over height
x_5	Misty weather	x_{10}	Ageing degradation

The matrix $\mathbf{T}_y = (1 \ t_y)^T$ has the past years of observed probabilities and, consequently, the matrix $\mathbf{B}_i = (\beta_{0,i} \ \beta_{1,i})^T$ has prediction parameters. The determination of \mathbf{B}_i represents the knowledge extraction from available databases through stored procedures routines on database server [18].

III. DATA FLOW AND MODEL IMPLEMENTATION

The proposed approach for failure probability estimation based on the Naive Bayes model is evidentially a cross platform configuration that requires both information and data processing from different sources.

Fig. 3 presents the information flow involved in the characterization of distribution network sections according to their failure probability using the *web app* that is a supervisory application running in a workstation by the distribution network operator [19] and [20]. The added object :*Threat* obtains failure probabilities by sequentially running the methods *runEstimator()* and *calcSFP(X)*. The first one estimates the probability values for features of interest using eq. (4) where elements of \mathbf{B}_i come from the output table :*Likelihood* provided by the stored procedure :*GetLikelihood* that is executed in the database server (DB SRV) after the supervisory application to request its execution using the query command "EXEC SP ...". The second method is then performed to calculate the failure probability for each distribution network section using eq. (1) and (2) that require the vector of current dependencies, *i.e.* the current status of features of interest, provided by weather forecasting and lightning monitoring at external http servers as well as by vulnerability models for tree growth and ageing degradation.

In this configuration, there are some important functions that should be explained in more details, such as the obtaining the annual probabilities via the maximum likelihood learning and vulnerability models.

A. Obtaining the Weather-Related Probabilities

Correlations among historical databases, for example, between outage and weather databases should provide useful

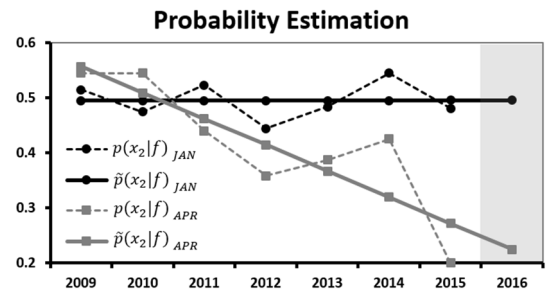


Fig. 2. Comparison in the estimation of medium wind speed probabilities.

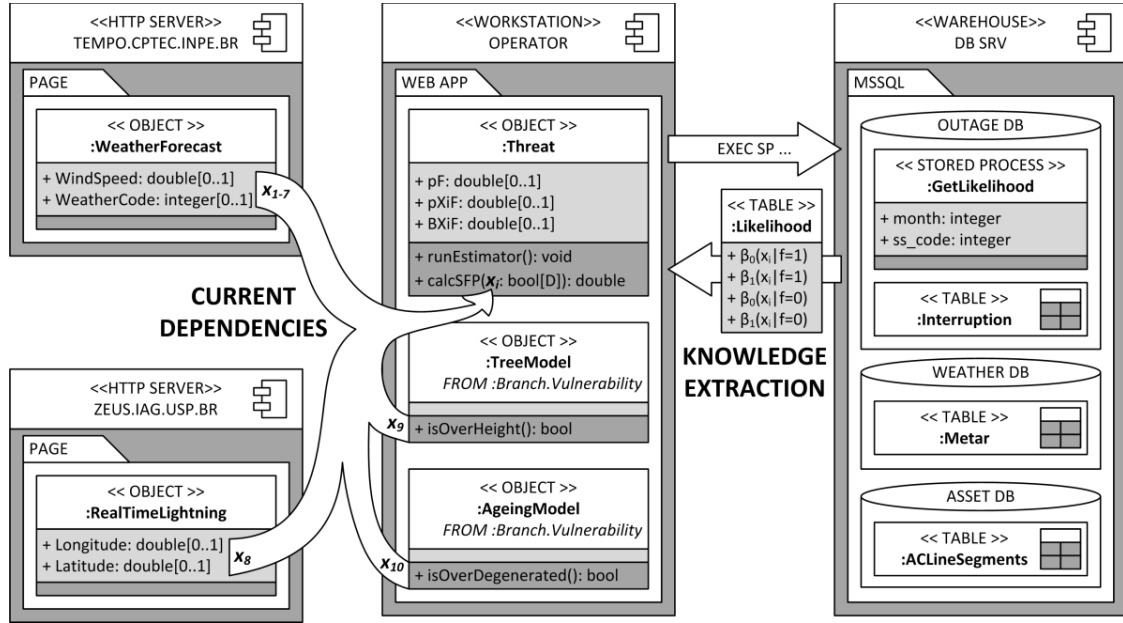


Fig. 3. Information flow involving the Naive Bayes probability estimation.

information in the computation of probabilities for features of interest. However, data collected from different sources typically have large volume with unnecessary information and, sometimes, the data is inconsistent. Data mining or knowledge discovery from data (KDD) comprises a sequential group of functions for dealing with these difficulties from massive database [21]. In Fig. 4, the table diagram illustrates attributes of the useful information after the data processing using the KDD functions such as data cleaning, integration and selection.

Once the database is pre-processed, data transactions allow for obtaining probabilities of interest. Fig. 4 highlights one intersection point between weather and power outage information that is a *datetime* variable representing the hour of weather data collection (*date_hour*) and time when the interruption event happens (*start_date*). A data transaction using this relationship should access other weather attributes like the integer variable coded in *present_weather* attribute that classifies the collected weather information according to [22] for calculating, for instance, the probability of thunderstorm during a failure event, as follows:

```

select  $N_{f=1} = \text{count}(1)$ ,
 $N_{x_{7=1}, f=1} = \text{count}(\text{case when } W.\text{present\_weather} = C,$ 
                                then 1 else null end)
from weather W, interruption I
where W.date_hour.hour = I.start_date.hour,
      W.date_hour.day = I.start_date.day,
      W.date_hour.month =  $t_m$ , W.date_hour.year =  $t_y$ , ...
set  $p(x_7 = 1 | f = 1) = N_{x_{7=1}, f=1} / N_{f=1}$ 
    
```

The presented pseudocode is the application of maximum likelihood learning in structured query language (SQL) where

the set C_f comprises the coded integer values for classifying thunderstorm and t_m and t_y are respectively the month and the year of the analysis. Despite the existence of complex relationships among the useful information, this basic pseudocode is sufficient for obtaining other probabilities of interest since it undergoes some needed adjustments.

B. Current Dependencies from Vulnerability Models

In accordance with Fig. 3, the set of current dependencies, from x_1 to x_8 , comes from external sources that employ atmospheric general circulation model [23] and lightning detection network [24] for climate impact diagnosis. The vegetation and ageing features are thus observed using internal vulnerability models. In power distribution systems, vulnerability models derive from reliability analysis where an assumption is taken that repair actions bring the system to a state as good as new and the mean time between failures is exponentially distributed [25]. In the vegetation model, the time to fault is based on tree trimming schedule while the ageing model uses the equipment lifetime.

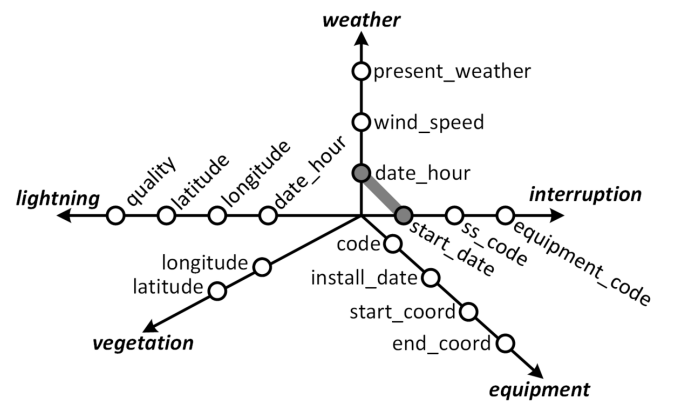


Fig. 4. Table diagram with dimensions of the useful information.

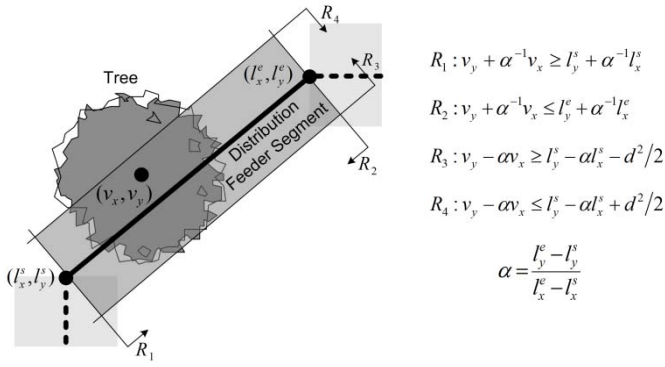


Fig. 5. Top view of a distribution feeder segment cover area.

Distribution networks are mainly overhead where the electrical conductors of distribution feeders are vulnerable to environmental threats like vegetation contact. The contact of tree branches with distribution feeder conductors frequently causes power flow interruption reducing the reliability indices. In this way, the localizations of trees and growth monitoring are worthy to guarantee the energy delivery with high quality. The vegetation localization can be performed using any type of remote sensing technology in association with GIS [26]. Once the trees are localized, the vegetation and equipment information must be correlated for identifying distribution feeders that are vulnerable to tree growth. Fig. 5 illustrates the top view of one distribution feeder segment overhanging a tree. The intersection area among hyperplanes, from R_1 to R_4 , defines the distribution feeder security area where any enclosed tree center point represents a potential threat; thus, distribution feeder segments with free security area do not incorporate the vulnerability model given by (6).

$$H_i = A_i \left(1 - P e^{-g_i(w_{temp}, w_{prec})(t_{curr}^i - t_{trim}^i)} \right) \quad (6)$$

A vegetation growth model predicts the height of trees as a function of time or age [27]-[28], in (6), the growth time comprises the time span between the current date, t_{curr}^i , and last trimming date, t_{trim}^i . Additionally, the growth rate, $g_i(w_{temp}, w_{prec})$, depends upon weather parameters such as monthly average values of temperature and precipitation. In the nature, A_i is the asymptotic tree height and $A_i(1-P)$ is the trimmed tree height that follows the security standards established by the utility. Hence, the vegetation related current dependence, x_9 , goes to one whenever the computed height is larger than a maximum height H_{MAX} , indicating the vegetation is over allowable height.

The last feature of interest is determined by the equipment lifetime and, as the vegetation vulnerability model, is given by an exponential function that expresses the equipment ageing degradation under electrical, mechanical and thermal stresses [29]. The ageing vulnerability takes the repair cycle assumption for correlating equipment status and power interruption information and, differently from vegetation model, it is applied for every devices in a distribution network using (7).

$$P_L^i = 1 - e^{-\left(\frac{t_{curr}^i - t_{inst}^i}{L_0^i}\right)^{\delta_i}} \quad (7)$$

Any electrical equipment may fail whether caused by electrical or mechanical stress as well as due to the degradation process by temperature or time. Equation (7) determines the likelihood of failure, P_L^i , comparing the expected lifetime, L_0^i , with operating time, *i.e.* deviation between the current, t_{curr}^i , and installation date, t_{inst}^i . This comparison is shaped using a scale-parameter, δ_i , that represents the dispersion in time-to-failure for different types of electrical devices such as circuit-breaker, conductor and distribution transformer. Thus, an electrical equipment is over degraded whenever it reaches at least 63% of possibility to fail.

IV. FAILURE PROBABILITY CLASSIFICATION AND EVALUATION

The proposed data mining approach determines the failure probability metric for each section in a distribution network, or feeder. A distribution network section comprises several electrical components and it is limited by two or more switches that must open to isolate the faulted section during repair actions. Since calculations are performed for each equipment, the section failure probability is equal to the maximum value obtained among electrical components comprising the section.

The test distribution network is a real feeder with nine sections limited by ten sectionalizing switches that have its data available in [30]. The learning information comprises seven years, from 2009 to 2015, of historical databases from two weather stations and one lightning detection network provided by external sources, as shown in Fig. 3. In regard to vulnerability models, parameters of the tree growth model are adjusted for a trimming schedule of one year while parameters of the equipment degradation model to different devices can be acquired using the method discussed in [29]. Calculations are performed using a programming language of general purpose (C++) and integrated with a distribution network simulation platform [20] that provides support for the use of georeferenced data.

The failure probability value alone does not have much meaning in itself, so a more semantic description of obtained values becomes necessary. One way to facilitate the understanding resides in the classification of failure probability. Jenks' optimization, or Jenks natural breaks classification [31], identifies break points boundary to natural grouping inherent in the data that maximize the difference between classes. Natural breaks classification is ideal for mapping values that are not uniformly distributed [32]. In the classification process, the histogram was built using failure probabilities obtained by permuting the observing vector for each month after the learning.

Table II presents six likelihood classes that are categories for classifying the achieved failure probability. Each class covers a probability interval with inferior and superior limits. Thus, the Likely class covers the shortest interval, from 0.91 to

TABLE II. FAILURE PROBABILITY CLASSIFICATION.

Class	Color	Interval Limits		Sum of Squared Deviations
		Inferior	Superior	
Likely	Red	0.91	1.00	0.1511
Unlikely	Orange	0.75	0.91	0.1483
Very Unlikely	Yellow	0.54	0.75	0.1659
Doubtful	Light Green	0.31	0.54	0.1118
Highly Unlikely	Cyan	0.10	0.31	0.1941
Extremely Unlikely	Blue	0.00	0.10	0.0407

1.00, while the longest interval, from 0.31 to 0.54, is covered by the Doubtful class. These classes has variable intervals because the failure probability is not uniformly distributed. The quality of achieved intervals can be checked by the sum of squared deviations for each class. In the Jenks optimization, the squared deviations are minimized reflecting in calculation of the goodness of variance fit (GVF) that, in this case, is equal to

0.98704 wherein $GVF=1$ is perfect fit.

After the failure probability classification, the proposed methodology is evaluated under a real distribution network conditions. Fig. 6 presents the characterization of distribution network sections according to failure probability metrics for a sequence of weather changes. The observed weather event starts on January, 14th at 20:00 when the weather stations detects a thunderstorm event with medium wind speed. The observed vector is equal to $X = \{0100001x_8x_9x_{10}\}$ for every distribution network section that has its intrinsic features represented by the last three binary variables. This climate combination is sufficiently dangerous and results in completely alarming the distribution network in red color, as shown in Fig. 6 (a), indicating the likely occurrence of a failure event.

Two hours later, *i.e.* at 22:00, the observed weather pattern changed, the tempest ceased and the weather becomes rainy with low speed wind where its vector representation is given by $X = \{1000010x_8x_9x_{10}\}$. This observed weather event is less hazardous than anterior event but failure probabilities for some sections still remain high as is shown in Fig. 6 (b). The weather normalizes the next morning, *i.e.* on January, 15th at 08:00, when it does not rain anymore and the observed weather is good with low wind speed that is represented by the vector $X = \{1001000x_8x_9x_{10}\}$. In calm weather, probabilities indicate that the occurrence of a failure event in any distribution network section is highly and extremely unlikely, Fig. 6 (c). Sections with extremely unlikely failure probability are more reliable because they do not have any vulnerability exposure while other sections with highly unlikely failure probability may have over grown vegetation and/or over degraded components.

The exposed behavior is better understood by comparing the probabilities obtained in the learning stage. Fig 7 shows a 100% stacked column chart where the first column represents the failure probability, $p(f)$, in January where it is around 20%. When the weather is calm, $x_1=1$ and $x_4=1$, the probability for a failure occurring is about 40% while in 80% of the time that the weather is stormy, $x_2=1$ and $x_7=1$, at least a failure happed in the distribution network. Since the thunderstorms are not so frequent, then most likely a failure event will occur as was effectively characterized in the Fig. 6 (a).

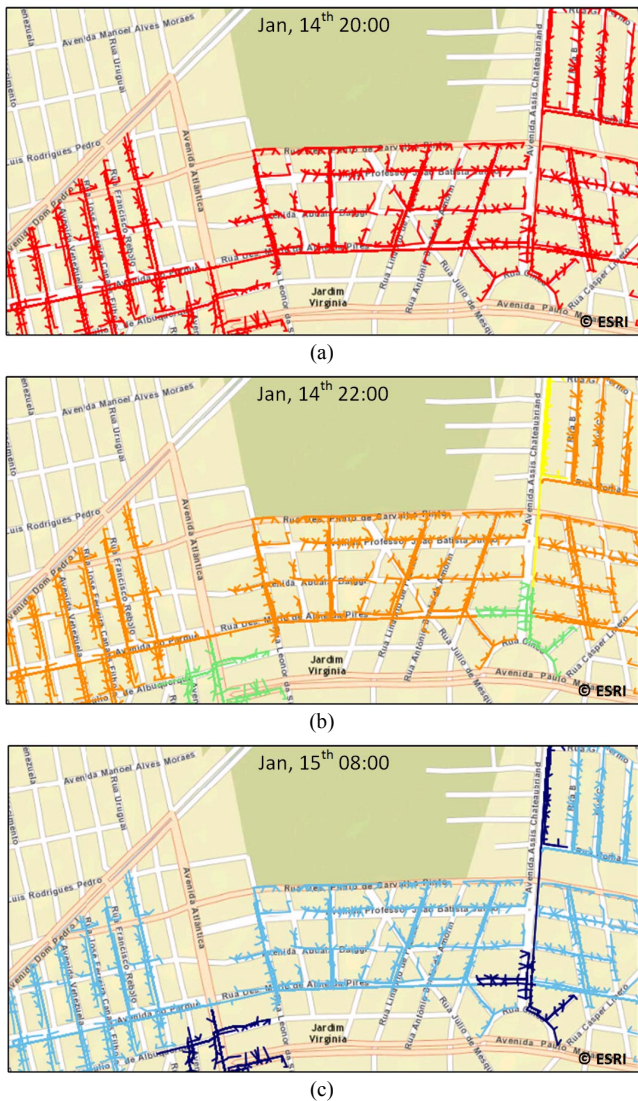


Fig. 6. Classification of distribution network section according with failure probability metric.

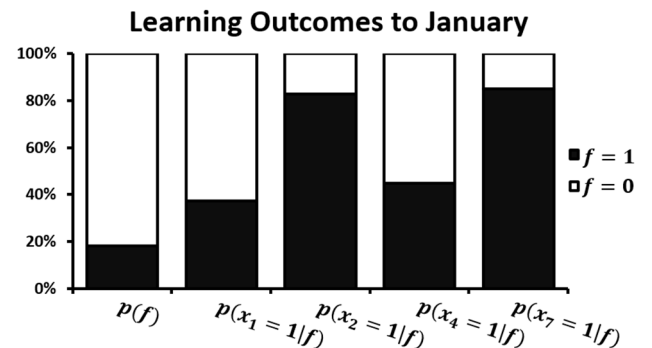


Fig. 7. Comparison of probabilities using 100% stacked column chart.

V. CONCLUSION

The study has made the following contributions:

- In the light of the failure unpredictability, the proposed use of data mining present a way to classify distribution network sections in accordance with their failure probability values estimated using a machine learning technique.
- The joint use of Naive Bayes learning and natural breaking classifier embedded in a GIS platform provides a tool that is capable of effectively merging a massive variety of information into just one semantic characteristic.
- The proposed method for determining the failure probability can be added to the previously proposed interruption cost method [12], and turned into an online risk assessment framework to identify distribution network regions with high risk aiding the risk mitigation.
- Presented study allows the incorporation of weather parameters, like average temperature and lighting incidence, straightforwardly associated to thermal and electrical stresses.

REFERENCES

- [1] S. Borlase, "Smart Grid Technologies," in *Smart Grids: Infrastructure, Technology and Solutions*, vol. 1, 1st ed. Boca Raton, FL: CRC Press, 2013, pp. 67-494.
- [2] B. M. Ayyub, *Risk Analysis in Engineering and Economics*, Boca Raton, FL: Chapman & Hall/CRC, 2003.
- [3] T. Dokic, P. Dehghanian, P. Chen, M. Kezunovic, Z. Medina-Centina, J. Stojanovic and Z. Obradovic, "Risk Assessment of a Transmission Line Insulation Breakdown due to Lightning and Severe Weather," in *Proceeding IEEE CSHawaii International Conference on System Sciences (HICSS)*, Kauai, Hawaii, 2016, pp. 2488-2497.
- [4] S. Nattian and M. Kezunovic, "A Risk-Based Approach for Maintenance Scheduling Strategies for Transmission System Equipment," in *Proceeding 10th International Conference on Probabilistic Methods Applied to Power Systems (PMAPS 2008)*, Rincón, Puerto Rico, 2008, pp. 1-6.
- [5] R. Shah, R. Yan and T. K. Saha, "Chronological Risk Assessment Approach of Distribution System with Concentrated Solar Power Plant," *IET Renew. Power Gener.*, vol. 9, no. 6, pp. 629-637, Aug. 2015.
- [6] C. J. Wallnerstron and P. Hilber, "Vulnerability Analysis of Power Distribution Systems for Cost-Effective Resource Allocation," *IEEE Trans. Power Syst.*, vol. 27, no. 1, pp. 224-232, Feb. 2012.
- [7] T. V. Garces and A. T. de Almeida, "Multidimensional Risk Assessment of Manhole events as a Decision Tool for Ranking the Vaults of an Underground Electricity Distribution System," *IEEE Trans. Power Deliv.*, vol. 29, no. 2, pp. 624-632, Apr. 2014.
- [8] G. Li, P. Zhang, P. B. Luh, W. Li, Z. Bie, C. Serna and Z. Zhao, "Risk Analysis for Distribution Systems in the Northeast U.S. under Wind Storms," *IEEE Trans. Power Syst.*, vol. 29, no. 2, Mar. 2014.
- [9] P.-C. Chen and M. Kezunovic, "Fuzzy Logic Approach to Predictive Risk Analysis in Distribution Outage Management," *IEEE Trans. SmartGrid*, Early Access.
- [10] P.-C. Chen, et al., "Predicting Weather-Associated Impacts in Outage Management Utilizing the GIS Framework," in *Proceeding IEEE PES Innovative Smart Grid Technologies Latin America (ISGT LATAM)*, Montevideo, Uruguay, 2015, pp. 417-422.
- [11] Q. Yan, T. Dokic and M. Kezunovic, "GIS-Based Risk Assessment for Electric Power Consumers under Severe Weather Conditions," in *Proceeding 18th Mediterranean Electrotechnical Conference (Melecon 2016)*, Limassol, Cyprus, 2016, pp. 1-6.
- [12] J. B. Leite, et al., "The Impact of Time Series-Based Interruption Cost on Online Risk Assessment in Distribution Networks," in *Proceeding 2016 IEEE PES Trans. and Distrib. Conf. and Exposition Latin America (T&D LA 2016)*, Morelia, Mexico, 2016, pp. 1-6.
- [13] P. Chen, T. Dokic and M. Kezunovic, "The Use of Big Data for Outage Management in Distribution System," in *Proceeding International Conference on Electricity Distribution (CIRED)*, Rome, Italy, 2014, pp. 1-5.
- [14] M. Kezunovic, L. Xie and S. Grijalva, "The Role of Big Data in Improving Power System Operation and Protection," in *Proceeding Bulk Power System Dynamics and Control Symposium*, Rethymnon, Greece, 2013, pp. 1-9.
- [15] D. Barber, "Naive Bayes," in *Bayesian Reasoning and Machine Learning*, vol. 1, 1st ed. New York, NY, Cambridge University Press, 2012, pp. 243-255.
- [16] D. Lowd and P. Domingos, "Naive Bayes Models for Probability Estimation," in *Proceeding 22nd International Conference on Machine Learning*, Bonn, Germany, 2005, pp. 1-8.
- [17] D. J. Hand and K. Yu, "Idiot's Bayes – Not So Stupid After All?," *International Statistical Review*, vol. 69, no. 3, pp. 385-398, Mar. 2001.
- [18] R. Vieira, "Stored Procedures," in *Beginning SQL Server 2005 Programming*, vol. 1, 1st ed. Indianapolis, IN, Wiley Publishing Inc., 2006, pp. 341-408.
- [19] J. B. Leite and J. R. S. Mantovani, "Development of a Smart Grid Simulation Environment, Part I: Project of the Electrical Devices Simulator," *J. Control Autom. Electr. Syst.*, vol. 26, no. 1, pp. 80-95, Feb. 2015.
- [20] J. B. Leite and J. R. S. Mantovani, "Development of a Smart Grid Simulation Environment, Part II: Implementation of the Advanced Distribution Management System," *J. Control Autom. Electr. Syst.*, vol. 26, no. 1, pp. 96-104, Feb. 2015.
- [21] J. Han, M. Kamber and J. Pei, "Data Mining: Concepts and Techniques," vol. 1, 3rd ed. Waltham, MA, Elsevier Inc., 2012, p. 740.
- [22] Code table 20003, accessed: Oct. 06, 2016. [Online]. Available: http://bancodados.cptec.inpe.br/~rbanco/bdm/tabela_bufr_20003.pdf.
- [23] I. F. A. Cavalcanti, et al., "Global Climatological Features in a Simulation using the CPTec-COLA AGCM," *J. Climate*, vol. 15, no. 21, pp. 2965-2988, Nov. 2002.
- [24] C. A. Morales, et al., "Sferics Timing And Ranging NETwork – STARNET: 8 Years of Measurements in South of America," *Proceedings XV Int. Conf. on Atmospheric Electricity*, 2014, pp. 1-8.
- [25] N. Balijepalli, S. S. Venkata and R. D. Christie, "Modeling and Anlysis of Distribution Reliability Indices," *IEEE Trans. Power Deliv.*, vol. 19, no. 4, pp. 1950-1955, Oct. 2004.
- [26] W. Luo, et al., "Analysis of College Urban Forest Structure using RS and GIS Technology," *Proceedings of the 18th international Conference on Geoinformatics*, Beijing, China, 2010, pp. 1-5.
- [27] D. T. Radmer, et al., "Predicting Vegetation-Related Failure Rates for Overhead Distribution Feeders," *IEEE Trans. Power Deliv.*, vol. 17, no. 4, pp. 1170-1175, Oct. 2002.
- [28] F. B. Martins, C. P. B. Soares and G. F. da Silva, "Individual Tree Growth Models for Eucalyptus in Northern Brazil," *Sci. Agric.*, vol. 71, no. 3, pp. 212-225, May 2014.
- [29] X. Zhang and E. Gockenbach, "Component Reliability Modeling of Distribution Systems based on Evaluation of Failure Statistics," *IEEE Trans. Dielec. And Elect. Insul.*, vol. 14, no. 5, pp. 1183-1191, Oct. 2007.
- [30] Distribution testing system of 1807 lines, accessed: Oct. 12, 2016. [Online]. Available: http://www.feis.unesp.br/Home/departamentos/engenhariaeletrica/lapsee807/home/distribution_network_1806_lines.rar
- [31] G. F. Jenks, "The Data Model Concept in Statistical Mappng," in *International Yearbook of Cartography*, 7, 1967, pp. 186-190.
- [32] T. Slocum, R. McMaster, F. Kessler and H. Howard, "Thematic Cartography and Geovisualization," vol. 1, 3rd ed. Upper Saddle River, NJ: Pearson Prentice Hall, 2009, p. 561.