

Fault Prediction for Network Devices Using Service Outage Prediction Model

Sunita A Yadwad and Valli Kumari Vatsavayi

Department of CS & SE, Andhra University College of Engineering, AP, India

Email: sunitayadwad@gmail.com; vallikumari@gmail.com

Abstract—Minimization of network downtime is the biggest challenge for service providers and one of its prime causes is equipment failure. On-time prediction and rectification of faults can reduce downtimes. Dynamic and very adaptive algorithms are required for processing huge torrents of data and the generation of predictions based on patterns and trends in the data obtained from trouble tickets and system logs. A novel strategy for fault detection based on the data accumulated has to be applied where the equipment behavior is monitored closely to prevent its failure and further prevent a network failure or downtime. Paper proposes Service Outage Prediction (SOP) that uses hidden Markov models (HMMs) which have a successful record in tasks related to pattern recognition and have been successfully used in the prediction of failures. The features of the aggregated fault data are subject to the supervised learning algorithm, in the initial phase of training. The samples are traced at different stages, and the failures are detected through high priority in tickets. Among the many solutions possible one of the best solutions being the approach of combining the Hidden Markov model and Bayesian Network. The results indicate the strengths of Hidden Markov Models as the probabilistic approach increases the accuracy of the prediction when compared to the other prediction algorithms. The likelihood of a customer raising a trouble ticket with high priority is predicted by the SOP model proposed.

Index Terms—HMM hidden Markov model, Bayesian networks, Viterbi, Baum Welsch

I. INTRODUCTION

Any interruption or discontinuation of internet service is not well received by the customers and is not healthy for the growth of the internet service providers. The root cause of any outage is a delayed prediction of failure in the system. The early prediction of equipment failure not only guards the outlook of a company in terms of service and reputation but also decreases the occurrence of Internet outages and dissatisfaction among customers. The major challenges faced by telecommunication systems are manipulating volumes of data coming from several sources, storing, further analyzing the cause of equipment failure and predicting the upcoming failures if any based on the trend studied. When outages happen a customer with trouble using the services will get in touch with the customer care service of the provider for issuing a customer trouble ticket. This information about the service affected and the symptoms are brought to notice by the customer are identified and the customer service

agents guide the user through the problem. Trivial network faults can be corrected with the assistance of the agents. But the major ones need the immediate attention of the resolution team working for the network providers. Faults need to be addressed at the earliest in an effective manner. Service providers need to assign their workforce for investigation and for resolving of the trouble tickets issued by the customers. The ability with which network fault occurrences are predicted will allow a provider for workforce allocation and optimization. So far lots of research work has been conducted focusing on the fault prediction in the network using the volume of trouble tickets created and time series model to predict the volume of network fault and (ii) using system logs that are generated by the specific network components for prediction of possibility of the components to be faulty. The archive obtained from the customer trouble tickets has been used for labeling the aggregated data as failure or no failure. Classification models are implemented to deduce the features contributing to the making of high-priority trouble tickets. The selected features are beneficial to providers in taking preventive measures. The evaluation of the prediction model is done with the test data set aside while training of the model. The model has to predict the possibility of the customer coming up with a ticket of high priority related to faults in the window of prediction.

Markov models are considered to be the most powerful tools for modeling the time series data. They are both stochastic processes and random processes changing through time with a probability of upcoming states of the process depending only upon the existing state, not on the previous sequence of states like the other time series analysis methods. They analyze a generative observable sequence through an unobservable sequence. The Hidden Markov Model (HMM) is devised from the basic statistical Markov Model with hidden states. The HMM is a very popular method in modeling the time series data owing to its rich mathematical structure and the ease of available practical algorithms. Several papers published about HMM and its applications in speech, the stock market, and biology exist in the literature. However, there is less amount of work done in predicting the failures using HMM.

Bayesian Networks are probabilistic-based models that help visualize the relationships existing between random variables and the causal probabilities for trends observed.

The model is about the study of subjective probabilities that is belief in an outcome not based on past occurrences of events.

The proposed model uses the capabilities of the Bayesian Network and HMM for predicting the failures and rectifying the network before an outage occurs. The warning leading to an equipment failure and the grievances of the customers and the technical staff are studied in detail for building the framework which is designed and tested on the historical data. High priority tickets should be predicted at the earliest as once these tickets surface the equipment or disruption cannot be handled. Lower priority tickets do not cause major problems and can be rectified. The classifiers are used for predicting high-priority tickets which cause major disruption in network devices leading to a standstill or outage. Prediction results are compared by checking the accuracy of all classifiers like the random forest, decision tree, Support Vector Machines etc. Further, all these accuracies are compared with the Bayesian Network and HMM model to conclude that the latter performs well in terms of accuracy for the prediction of high priority tickets indicating major network disruption.

Section II discusses related literature about HMM and its efficacy in solving varied applications, Section III and V discusses the Service Outage Prediction model for fault prediction, Section IV discusses the HMM and Bayesian network architecture and algorithms. Finally Section VI discusses the results and is followed by the conclusion.

II. RELATED WORK

Many machine learning techniques are used in automated anomaly detection to identify failures in the network. These techniques are categorized into classification techniques and statistical techniques which are called supervised learning techniques as they need to be labeled data to supervise the learning process. Correlations existing between the system's behavior and already known problems could be determined but cannot be applied to detect unknown problems. Hence Bayesian networks in combination with HMM are used for classification techniques. An extensive survey is done about how the above techniques are used to improve fault detection.

The paper [1] used rule-based analysis and linear regression for consistent prediction of equipment failure. It was achieved by observing the pattern of the cumulative total of failure warnings and setting a certain range of the warning to effectively predict the equipment failure the next day. Most of the research on service outages and failure prediction focuses on the forecasting of the volume of customer trouble tickets obtained via historical datasets. Paper [2] augmented the customer trouble tickets with the internet's usage data and the other signal measurement metrics for improving network fault prediction. The predictive model was derived from the C5.0 Decision Tree and the Random Forest algorithms.

The Experiment results revealed that the RF algorithm showed a greater AUC value in comparison to the C5.0 Decision Tree algorithm. RF helped in feature importance identification and C5.0. Decision Tree was able to illustrate the decision rules that described the relation among features selected. The paper [3], [4] introduced a new approach towards online failure prediction by forecasting the happening of failures by observing failure-oriented pattern recognition of error-producing events. The model was built on Hidden Semi-Markov Models. The authors used a model operating on the occurrence of failures known as reliability model-based failure prediction. Model, when applied to the data of commercially working telecommunication system for comparison in terms of recall, precision, F-measure, and false-positive rate, showed HSMM was the best. [5]

The study [6] examined the prediction of a number of faults analyzed by the telecommunication services using the Hidden Markov Model (HMM) and the Kalman filter model. The entire assessment was based on the accuracy exhibited in terms of the results obtained through modeling both the data. The Hidden Markov model has been experimented in the anomaly detection of intrusion (IDS) in computer systems [7], [8], for the most accurate fault diagnosis. In [7], IDS alongside HMM technique is used for the detection of varied attack types. The system performance is validated on CICIDS dataset. The proposed technique provided a good defense to DDOS attacks and a better intrusion detection rate as per the authors. On the same note [8] presented a comparative study of SVM and HMM for anomaly detection and identifying distinguishable TCP services in intrusion data.

Using HMM and Markov hypothesis makes it a more reliable method for traffic flow prediction. Use of HMM in stock prediction shows HMM is better than ANN in terms of Bayesian information and Akaike's Criteria (AIC, BIC) and mean absolute percentage Error, in [9], HMM is also used for the detection of machine failure for a process control problem. HMM was tested on four different stocks over a varied period of time with independent modeling for each stock. Using the Mean Absolute Error (MAE) results it was proven that HMM outperformed ARIMA and ANN despite being the simplest method [10].

A novel HMM based on Gaines algorithm and MML estimator for failure prediction is presented by the authors using PFA for optimizing the number of states [11]. In [12] authors introduced an approach based on machine learning-based approach for prediction of the time of occurrence of very rare events with the aid of Markov mixed membership models (MMMM) Authors [13] describe the use of HMM models for learning complicated stochastic degradation patterns of network assets from the data. The failure state of the asset was represented by the terminal states and other healthy states of the assets were represented by interior non-terminal states.

This work [14], [15] proposes a prediction model which is Markov based, which checks multistage entropy for mobile networks. The transition probability matrix here is trained by HMM in a mobile network. HMM uses hybrid time series classifiers which work on the learning by mistake principle for better accuracy. Confusion matrices train the HMM model by learning from misclassified samples and deep understanding about the relationship of patterns in data and not just finding single individual patterns [16].

HMMs are able to detect latent structures in longitudinal settings. The mixture HMMs allow clustering data into homogenous subsets. The seqhmm package in the library of R is fabricated for categorical time-series data and efficient modeling of one or more sequences. Better graphical presentation of data and models are supported by R packages. The paper [17] presented an R package comprised in Comprehensive R Archive Network (CRAN) known as hmmhdd which is developed for both functional and multivariate using HMM has the ability to deal with High dimensional dataset having zero knowledge about the data and seqhmm package for visualizing and plotting parallel sequence data.

The work [18] highlighted the potential in the use of multi-class HMM for IDS in telemetry applications. PCA along with SVM was used in the reduction of dimension for the TCP data post feature selection and feature creation. A vector quantization method that reduces data i.e. K-means clustering marked the cluster labels and fed the sequence of observations to HMM model.

In the paper [19], the authors presented an overall analysis of various research ideas about RF and HMM methods for Internet traffic classification. The methods were tested against accuracy, features overhead, speed, complexity, memory utilization to conclude that RF was better than HMM. However, RF needed additional improvements for dealing with large amounts of data and making memory consumption fit for an online environment. HMM algorithm suffered from complexity problems and computation costs. In paper [20] HMM with its Markovian hypotheses and independence hypotheses for observations was used to provide efficient modeling for traffic flow prediction

The paper [21] proposes a framework for the creation of an appropriate probabilistic model and Bayesian network for the Bosch dataset by a selection of variables that have statistical importance. The network helps in answering probabilistic queries and classification in the production process. The automatic attribute selection in the training and test data can aid in developing the best predictive model. The objectives of any feature selection are improving the prediction performance and providing a better, faster, and very cost-effective model. Paper [22] proposed CSFS algorithm for the best feature selection and hybridization of HMM and SVM classifier for the classification process. They attained 93.4% accuracy. The experimental results were evaluated and compared with the existing classifiers such as SVM, HMM and ANN.

The results of several experimental studies using the IDS dataset show that the Hidden Markov Model combined with various feature selection and discretizing methods exhibits good results be it detection accuracy, misclassifying cost or error rate when compared to SVM or Naïve Bayes methods [23], [24]. The HMM results are better detectors of denial of service attacks. HMM is a powerful Computer Security approach for dynamic classifying normal and attack traffic in Intrusion [25].

The dataset that consists of already categorized tickets can benefit training classification algorithms. The BOW or bag of words approaches extracts features vectors. The paper [26] implemented a plethora of classification algorithms with that exhibited varying behavior when subjected to different datasets and weighting methods. The "No Free Lunch" theorem states no single algorithm for any learning domain can produce the most accurate model. Therefore, the major objective of any learning approach must be the proper utilization of strengths of one technique to accompany the weakness of another [27]. That has been the motto of the proposed work in the paper.

III. SERVICE OUTAGE PREDICTION MODEL

Database comprising of historic incidents, stored as tickets and their corresponding resolutions and actions are maintained by management systems of network providers. Historical data is extracted from the grievances of customers as Customer Trouble tickets. The data is also collected from customer's internet usage and the signal processing data at the provider's site. The search for the correlation between tickets for finding the common solution for an issue is a cost involving expensive process in terms of both manual labor and productivity. All the above information goes into the IDM dataset [28], [29].

The IDM dataset under study contains 46K tickets covering the span of two years 2014-2017. It contains a number of fields, with 25 attributes associated such as category, subcategory, open time, closed time, urgency, impact, priority and description, type, reassignment count, knowledge base etc. The tickets are a combination of both numerical and categorical values. From the categorical ones, some are redundant and remain unused in the process.

In the proposed work, the tickets data over a network are discretized into a number of time slices which helps in categorizing the data. The data is bifurcated into testing, training data on the basis of time boundary slots. The data consisting of feature vectors recorded in the time interval $[0, m]$ is the training dataset and the feature vectors corresponding to time $[m, t]$ represent the testing dataset. The Service Outage Prediction model is better explained with the following steps as shown in Fig. 1.

a. Preprocessing of the dataset: The data to be trained is normalized by means of the normalization method. Preprocessing involves feature selection to estimate the variables that are important by assigning feature

importance to the variables. Machine learning models can rarely use all the variables in the dataset to build a model. Redundant variables reduce the accuracy and irrelevant variables do not contribute to the performance. As per Occam Razor’s Law of Parsimony the best explanation to any given problem is the one involving fewer possible assumptions. Thus this makes feature selection important for finding the best set of features as part of preprocessing to increase efficiency. After identifying the features that are important as shown in Fig. 2, the features listed are sent for further analysis.

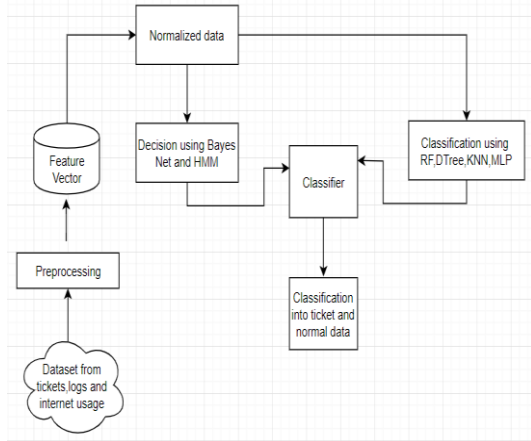


Fig. 1. Service outage prediction model

b. Training the dataset: The training of normalized data is done through various prediction algorithms like: a) Random Forest, b) Artificial Neural Network, c) Support vector Machines, d) XGBoost e) Decision Tree to predict the high priority tickets.

- Feature: CI_Name, Score: 0.01222
- Feature: CI_Cat, Score: 0.00272
- Feature: CI_Subcat, Score: 0.00988
- Feature: Incident_ID, Score: 0.00027
- Feature: Status, Score: 0.00000
- Feature: Impact, Score: 0.31163
- Feature: Urgency, Score: 0.25593
- Feature: Priority, Score: 0.35427
- Feature: number_cnt, Score: 0.00018
- Feature: Category, Score: 0.04279
- Feature: KB_number, Score: 0.00958
- Feature: No_of_Reassignments, Score: 0.00053

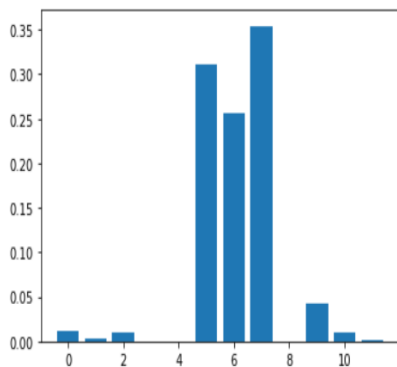


Fig. 2. Feature importance in IDM dataset

c. Prediction of High Priority Tickets: The data tested is flagged as either high priority or low priority by the

Classification decision function. The accuracies obtained by the various classifiers are recorded for the prediction of high-priority tickets [30]. The classification of the tickets to predict high-priority tickets is done based on the decision function using all the steps from pre-processing and the detection phase comprising of HMM and Bayes model used at the decision function unit. Since all the classification methods enlisted are supervised learning methods and deal with static feature vectors the proposed model uses a dynamically adaptive HMM algorithm with Bayesian networks.

IV. HMM AND BAYESIAN NETWORKS

In the proposed work, the HMM is applied for the prediction of failure detection through the prediction of high-priority tickets for overcoming service outages. The transition and emission probabilities are based on hidden states plus joint distribution and conditional distribution plus the present states.

A. Bayesian Network Model

Bayesian Networks are belief network models representing the acyclic directed graph representing vertices and edges that represent the random variables and the table of a probability distribution. The probabilities of a variable value are largely dependent on the probable combinations of the parental values. The Bayesian network is represented as $G(V, E)$ with V representing vertices and E representing the edges. The random variables are associated with each node and the directed connections between the nodes are the conditional probabilities. Another important point to be noted is that there is no possibility to navigate the graph in a cyclic manner making it impossible to have a loop in them. For the IDM dataset variables get finalized post feature importance and correlation computation. The important features which are correlated are fed to the Bayesian network (BN). BN exhibits the probabilistic relationships between the variables category(X_1), impact(X_2), urgency(X_3), number_count (X_4) and no_of_reassignments(X_5). The table corresponding to each variable is called the Conditional Probability Table, which encodes the CPD of every variable given the value of its parents.

i. Bayes Net Algorithm: For building a model with Bayesian Net, the following algorithm is used:

1. Choose relevant variables from the training dataset and make them available as the state variables for the HMM model. In the paper context, the variables chosen are category, impact, urgency, no_of_reassignment and number_count.

2. The variables are systematized as $X_1, X_2, X_3, \dots, X_n$. Here X_1 is the first variable, X_2 second and so on in the ordering. Here $X_1 = \text{category}$

3. For each $i = 1$ to 5

- a) Firstly the node X_i gets added in the network.

- b) The Parent (X_i) is a minimal subset of $\{X_1, X_{i-1}\}$ all the dependences and the independence for (X_i) and other members is obtained through the parent.
- c) Define $P(X_i = m \mid \text{allocation of Parents } (X_i))$, the Probability table given Parent (X_i).

B. HMM Modeling and Its Architecture

Hidden Markov model, a sequence classifier represents a set of finite states where each state has a probability distribution corresponding to it. The state transitional probabilities represent the possible transitions among states. The outcome or observation for every state is evident to an observer despite having hidden states. This gives the name to the model as a hidden Markov model. Like any other machine learning algorithm it can be trained, with knowledge of a labeled sequence of observations, and the learned parameters it can assign a sequence of labels to the sequence of observations given .

The HMM has a framework containing the following components:

1. Firstly the number of hidden state variables N which I the context is 5 based on the number of chosen variables.: $T=t_1, t_2, \dots, t_N$.
2. The numbers of symbols of observations representing the various values the variables take. For example, the priority takes three values high, low, med. Similarly, all the other variables combined there are 15 distinct symbols. $W=w_1, w_2, \dots, w_N$.
3. $A = \{a_{ij}\}$, a set of transition probabilities of state.

The HMM's probability distribution is denoted by the triple notation $\lambda = (B, A, \pi)$ where probability matrices B is emission, A is transition. The transition probabilities go from one state to another, Emission probability has the probabilities of an observation that are generated from a state and the initial probabilities π . The general design of an HMM also known as the trellis diagram is depicted in Fig. 3.

An HMM of First-order have the following assumptions:

1. The probability of any particular state is dependent only on the probabilities of its previous state. Formally: $P(t_i | t_1, \dots, t_{i-1}) = P(t_i | t_{i-1})$. This is a Markovian assumption.
2. Output independence indicates that the output observation probability w_i depends only on the observation t_i producing state, not any other observations or states. Formally:

$$P(w_i | t_1 \dots q_i, \dots, q_T, o_1, \dots, o_i, \dots, o_T) = P(o_i | q_i)$$

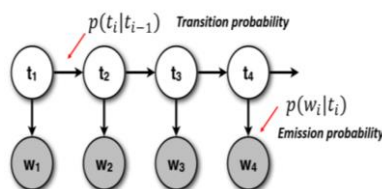


Fig. 3. The HMM model architecture

Discrete values get generated from the categorical distribution whereas the continuous values are generated from the Gaussian distribution. The parameters of HMM are grouped as two types: the transition probabilities and output probabilities known as emission probabilities. Given the values of the hidden state at $t-1$ time, the values of the hidden state at t time is chosen and managed by transition probabilities. The Bayesian HMM model which is based on Bayesian Network is constructed using the historical tickets and internet data. The information from the network helps in the calculation of state probabilities and emission probability. These parameters are the basis of the HMM constraints. The predictive model of HMM helps in distinguishing between the normal data and the anomalous data where a fault has been detected.

V. FRAMEWORK FOR THE SERVICE OUTAGE PREDICTION MODEL

A Service Outage Prediction model for failure detection based on HMM and Bayes network is presented in figure 5. This failure detection framework comprises of the following levels: Read the dataset, pre-processing of data, Bayes net, HMM parameters initialization, generation of states and sequence, estimation of state transition estimation, matrix of emission probability, and the model evaluation. The levels are described below

A. Reading of the Historical Data

The model based on fault detection is trained with the IDM dataset and then tested. Once the dataset is chosen, it is pre-processed. The sample data from the dataset is enlisted in Fig. 4.

The data that is pre-processed is used for guidance and also testing. The incessant variables are discretized and the discrete values obtained are expressed using symbols.

CI_Name	CI_Cat	CI_Subcat	Incident_ID	Status	Impact	Urgency	Priority	number_cnt	Category	KB_number	No_of_Reassignments	
0	SU/8000508	subapplication	Web Based Application	IM0000004	Closed	4	4	4	0.601292	incident	KM0000553	26.0
1	WBA000124	application	Web Based Application	IM0000005	Closed	3	3	3	0.415050	incident	KM0000611	33.0
2	DTA000024	application	Desktop Application	IM0000006	Closed	0	3	0	0.517551	request for information	KM0000339	3.0
3	WBA000124	application	Web Based Application	IM0000011	Closed	4	4	4	0.642927	incident	KM0000611	13.0
4	WBA000124	application	Web Based Application	IM0000012	Closed	4	4	4	0.345258	incident	KM0000611	2.0
5	WBA000124	application	Web Based Application	IM0000013	Closed	4	4	4	0.006678	incident	KM0000611	4.0
6	WBA000082	application	Web Based Application	IM0000014	Closed	4	4	4	0.152907	incident	KM0000401	2.0
7	WBA000124	application	Web Based Application	IM0000015	Closed	4	4	4	0.996468	incident	KM0000611	5.0
8	WBA000124	application	Web Based Application	IM0000017	Closed	0	3	0	0.269574	incident	KM0000611	2.0
9	WBA000082	application	Web Based Application	IM0000018	Closed	4	4	4	0.306670	incident	KM0000401	2.0
10	WBA000124	application	Web Based Application	IM0000019	Closed	4	4	4	0.439909	incident	KM0000611	6.0
11	WBA000124	application	Web Based Application	IM0000020	Closed	4	4	4	0.291928	incident	KM0000611	8.0
12	WBA000082	application	Web Based Application	IM0000021	Closed	4	4	4	0.776496	incident	KM0000401	5.0
13	WBA000124	application	Web Based Application	IM0000022	Closed	4	4	4	0.592713	incident	KM0000611	17.0
14	SBA000698	application	Server Based Application	IM0000023	Closed	3	3	3	0.025310	incident	KM0000070	1.0

Fig. 4. The samples of data set

The data is in the text format and consists of 25 dimensions with 4,60,000 records of which 46000 is considered as 10% of the total size. It has categorical, continuous and ordinal types of data type attributes. The data format is CSV (i.e. Comma Separated Values),

which is easily readable and analyzed. The samples of about 46000 records with 5 attributes have been considered. The attributes with higher correlation are considered. The sample data are shown in Fig. 4.

Fig. 5 shows the framework for SOP model.

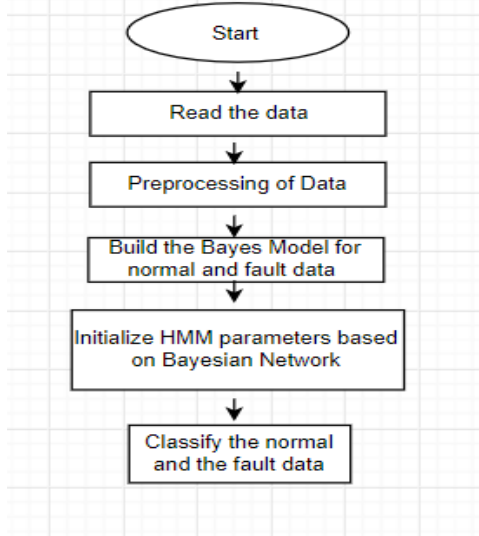


Fig. 5. The framework for SOP model

B. Build the Bayes Network

The best way one can represent a Bayesian network is using HMM. Bayesian networks are dynamic in nature and they sequence the order of the state variables. The state variables in the network are represented graphically with nodes and the edges representing relationships and dependency among the nodes. This dependency has an associated table called the conditional probability table. Bayesian network for data with high priority tickets is as shown in Fig. 6.

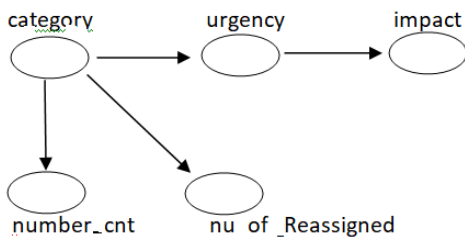


Fig. 6. State transition diagram-based Bayesian network

C. HMM Parameters

A subset of 5 features out of 25 features are chosen as hidden variables, category, impact, urgency, No_of_reassignments and, number_count based on the feature importance achieved among the attributes. These hidden variables chosen are the hidden state variables and every variable has a set of distinct values emitting a symbol. Every state variable has R distinct values say, V₁; V₂; . . . V_R. They form the observations X(t). The transition state diagram of HMM for ticket type records is shown in figure below.

- The state hidden variables N have a count of 5 on basis of the volume of chosen variables.

Therefore 5X5 is the size of the state transition matrix.

- The unique emission symbols count is 15 so M is equal to 15. Therefore 5x15 is the size of the matrix called emission transition probability.
- The probability distribution initially is considered as $\pi = \{0.000518, 0.261092, 0.08893, 0.372858, 0.275188\}$;The matrix of state transition A initialized along with the random variables based on the Bayesian model is as shown in Fig. 7. The Emission probability B is initialized based on the state variable of emission probability.

Category	Urgency	Impact	number_cnt	No_of_Reassignments
Category	0	0.33	0	0.33
Urgency	0	0	0	0
Impact	0	1	0	0
number_cnt	0	0	0	0
No_of_Reassignments	0	0	0	0

Fig. 7. Matrix of state transition A for the dataset

D. Baum-Welch (Forward-backward) Algorithm

The Baum Welch training algorithm [31], is used in the estimation of HMM parameters from the Bayesian Network. For every iteration, probable state paths of type high priority or low priority are considered in the model for each and every known observation sequence O₁,O₂,O₃....O₁₅ of both types of tickets in terms of updating the calculated number of counts for all emission and transition. HMM parameters are modified for getting a new set of parameters up to a point where the sample likelihood is locally maximal. The equations for the above are done using equations 1 to 3:

Re-estimating initial state probability

$$\pi_i = Y_1^{(i)} \tag{1}$$

Re-estimating state transition Probabilities

$$\overline{a_{ij}} = \frac{\sum_{t=1}^{T-1} \phi_t(i,j)}{\sum_{t=1}^{T-1} \gamma_t(i,j)} \tag{2}$$

Re-estimating observation symbol Probabilities

$$\overline{B_j(V_k)} = \frac{\sum_{t=1}^T \text{where } O_t=V_k \gamma_t^{(j)}}{\sum_{t=1}^T \gamma_t^{(j)}} \tag{3}$$

E. Viterbi Training Algorithm

Viterbi algorithm [31] is an alternative method for estimating model parameters. For every unknown observation sequence O₁,O₂,O₃....O₁₅ most probable state path q* of high priority and low priority tickets for the test dataset is formulated by the algorithm. For estimation of the number of transaction count and emission symbol count, the path is used. The parameters use equations 4, 5 for calculation:

Re-estimating initial state probability

$$\bar{a}_{1j} = \frac{\sum_{t=1}^T x_{t(i)} x_{t+1(j)}}{\sum_{t=1}^T x_{t(i)}} \quad (4)$$

Re-estimating state transition Probabilities

$$\bar{b}_j(V_k)_{\square} = \frac{\sum_{t=1}^T \mathcal{O}_{t=V_k} x_{t(j)}}{\sum_{t=1}^T x_{t(j)}} \quad (5)$$

F. Evaluation Procedure

The unknown observation sequence $P = (O|\lambda)$ probability is calculated. The model for observation sequence $O_1, O_2, O_3, \dots, O_{15}$ of high priority and low priority tickets are solved by the use of the Forward-Backward algorithm. The evaluation process is explained in the following sections.

1) Forward algorithm

The variable $\alpha_t(i)$ illustrates the probability of the part unknown observation sequence $O_1, O_2, O_3, \dots, O_{15}$, and state i at the time t , for the given λ model. The major stages involved for Forward algorithm is illustrated using equations 6-8:

Initialization

$$\alpha_t(i) = \pi b_i(O_1) \text{ where } 1 \leq i \leq N \quad (6)$$

Induction

$$\alpha_{t+1}(j) = [\sum_{i=1}^N \alpha_t(i) a_{ij}] b_j(O_{t+1}) \quad (7)$$

where $1 \leq t \leq T - 1$; $1 \leq i \leq N$

Termination

$$P[(O|\lambda)] = \sum_{i=1}^N \alpha_T(i) \quad (8)$$

2) Backward algorithm

The variable $\beta_t(i)$ which is a backward variable can express part sequence's conditional probability from O_{t+1} till the last, given state i at t time. The major stages that of the backward algorithm are illustrated using 9-10 equations.

Initialization

$$B_T(i) = 1 \text{ where } 1 \leq i \leq N \quad (9)$$

Induction

$$B_t(i) = \sum_{j=1}^N a_{ij} b_j(O_{t+1}) \beta_{t+1}(j) \quad (10)$$

where $T - 1, T - 2, \dots, 1$ and $1 \leq i \leq N$

The whole procedure followed in the paper is described deliberately in the steps given below:

1) Select a few variables out of a total of 25 features from the IDM dataset as variables in question for building the prediction algorithms.

2) The whole IDM dataset is divided into test and train set by random sampling.

3) Use the normalization method to train and test data and perform discretization of data.

4) Use the feature selection method to determine feature importance in the dataset and plot the same.

5) Use the subset variables and determine the correlation matrix to find the dependency among the variables.

6) Find the Bayesian network structure for the IDM training dataset, and estimate the conditional probability for every variable through Bayesian Estimation.

7) Apply the SVM, ANN, RF, DTree, XGBoost and KNN classifier for the prediction dataset; calculate the classification metrics like accuracy, recall, precision, F1-score and the Area Under ROC (AUC) for prediction of high priority tickets.

8) Initialize the HMM parameters based on the Bayesian Network. Apply Service Outage Prediction model to the IDM dataset for prediction of high priority tickets. The results for each classifier are compared to determine the best performer.

VI RESULTS

In the study proposed, IDM dataset is classified into low priority ticket and high priority ticket data. A service outage Prediction model based on supervised prediction learning algorithms and HMM is proposed. Dataset comprising of previously categorized tickets are used in training the model. The feature vectors are accessed after preprocessing. The data is nominal, ordinal and categorical as well. For the purpose of prediction of the priority tickets, supervised classification algorithms used are Decision Tree, Support Vector Machine, Random Forest, K-Nearest Neighbor, XGBoost Classifier and Artificial Neural Network. SVM classifier is the most popular discriminate-based algorithm. The classifier is bothered about examples close to the discriminator and not of instances that are far [30]. The complexity of the SVM classifier does not depend on the data size but depends on merely the count of support vectors, not the dataset size. KNearest neighbor algorithm uses distance measurement to the closest predefined instance whenever a new instance emerges [32]. This algorithm is commonly preferred when the distribution of data is not well known. It is a classification algorithm based on instances. A decision tree is a widely used classification technique mostly popular owing to its simplicity. The classifier is comprised of a sequence of test questions in a tree structure [33]. The greedy algorithm is known for their top to bottom tree construction. At every node, the best split is intended for the remaining data. Biological neural network-inspired ANN has a great ability in determining rules and the meaning of complicated data [34], [35]. Random Forest, which is an assembly of trees, applies to bootstrap for multiple samples extraction in random to overcome Dtree weaknesses. RF model does not over fit and has higher predictive accuracy as the majority of votes in the decision tree are considered for the final result [36]. XGBoost has gained huge popularity and attention as a machine learning algorithm. It is used as a distributed gradient boosting algorithm. XGBoost facilitates efficient optimization by choosing a weaker

classifier. The L2 regularization term is added for the weights of the leaf to gain lower variance. All these factors add up in improving the predictive accuracy [37], [38].

All the above-mentioned algorithms are implemented in python to evaluate performances comparatively. Consequently, the proposed approach reduces not only human errors but also the efforts to improve customer satisfaction and better service. The analysis of the dataset is done in the following steps.

- i. Firstly by Predicting the Priorities. The predictor variables Category, CI Category, CI Subcategory WBS, etc are used to predict the high priority tickets initially as these tickets have an intense effect on the functioning of equipment. The tickets with priority 1, 2 disrupt the services and once they are surfaced make it difficult to be handled.
- ii. The request for change also needs to be predicted using variables Category, CI Subcategory, WBS, Priority, Number of Related Interaction, and Number of Related Incidents for possible misconfigurations of IDM assets.
- iii. Forecasting the tickets Volume with the predictor variable Open_Time helps in quarterly and annual preparation for the network providers to be prepared with the resources and technology planning. Finding out the factors that affect high priority and training a model which predicts the same.

The IDM data gets imported with all the necessary libraries and the missing values are taken care of. Selection of features for analysis, label encoding of ordinal column splitting of train and test data is followed. The model with the highest accuracy is considered the best algorithm for the ticket priority prediction. Thus the Machine learning looks at perspective for improving the process of fault management for providers through prediction and automation.

The supervised machine learning techniques i.e. K-Nearest Neighbor, Decision Tree, Support Vector Machine, K-Nearest Neighbor, Random Forest, XGBoost Classifier and Artificial Neural Network (ANN) which allow non-linear models are implemented in python. The Comparative performances of the metrics of classifiers are illustrated in Fig. 7.

For the results in the study, measures like the False Alarm Rate, Recall, Precision, F-Value and accuracy are useful to validate the performance. The tabular representation of the predicted versus actual classification is known as the confusion matrix.

Here Recall or detection rate is the proportion of the correctly classified test results by the model [39]. Whereas, the precision is the probability of correct classification. Accuracy is the ratio of how many instances are correctly classified by sum total of the instances. The SOP model uses all 25 variables. If accuracy alone is used as a measure for the performance

test of the system, the system performance could be biased. Therefore all metrics are used. They are computed using the confusion matrix. The accuracy for the datasets: accuracy of SVM is 73%, XGB, RF, KNN and Decision tree is 80%; and accuracy of ANN is 75%.

The performance metrics for comparison of the prediction algorithms for high priority tickets are done for SVM, RF, D-Tree, KNN, ANN, XGB and the results are illustrated in the Table I below.

TABLE I: COMPARISON OF ACCURACY OF PREDICTION ALGORITHMS

	Classification Accuracy					
	SVM	XGB	Decision Tree	RF	KNN	ANN
Accuracy	0.7302	0.8078	0.8077	0.8086	0.8049	0.7511
Precision	0.72	0.79	0.79	0.79	0.78	0.74
Recall	0.73	0.81	0.81	0.81	0.80	0.75
F1 -Score	0.71	0.80	0.80	0.80	0.79	0.73

The receiver operating characteristic is also known as ROC. ROC curve is a graphical representation that plots the classification model's performance at varied classification thresholds. The curve plots the true positive and false positive parameters. ROC- AUC score is the area under the ROC curve, and hence, the name Area under the ROC-Curve (AUC). **ROC-AUC** score is considered to be one of the major metrics for assessing the performance of classification models. For a better interpretation of the performance of the algorithms in predicting the high priority tickets the scores of all the classification algorithms are plotted as shown in Fig. 8.

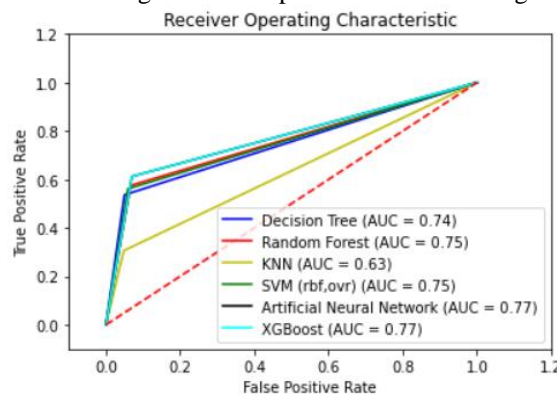


Fig. 8. ROC curve for prediction of high priority tickets

Correlation is a statistical measure that gives an indication of the extent of fluctuation of variables when they are together. It is generally performed between two or more variables and can be found using metrics like Kendall rank, Spearman rank, Pearson, etc. Having fewer and relevant features is always ideal as more features in any model could lead to a downfall in accuracy. The correlation metric could be a value either positive or negative or zero. In the proposed study the attributes selected after feature importance namely category, impact, urgency, nu_of_reassignment and num_count are subjected to correlation. The correlated attributes help us in the establishment of dependencies for the Bayesian

Network. The metric gives a glimpse of the relationship and dependency between the concerned variables. The correlation matrix with the values is depicted in Fig. 9 and Fig. 10 .

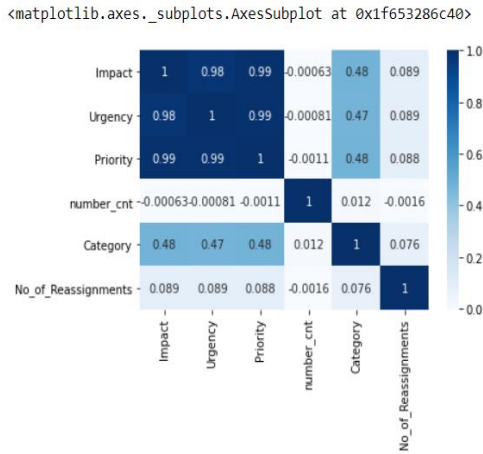


Fig. 9. Correlation Matrix for selected attributes

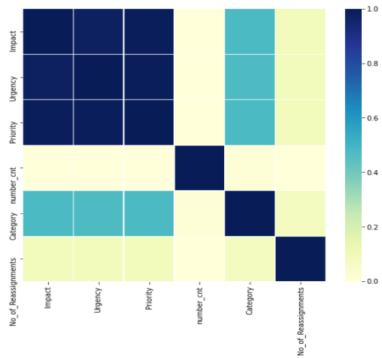


Fig. 10. Correlation matrix for selected attributes

An alternative method of prediction proposed in the work is without having to use the supervised classification algorithms is the Service Outage Prediction using Hidden Markov Model (HMM) algorithm and the Bayesian Network. The Bayesian output is fed to the HMM as a parameter. The HMM when applied as a predictor gives better accuracy and exceeds the performance of the traditional classifiers like SVM, Decision Tree, XGB, KNN, ANN and RF. HMM also reduces the cost of computation.

HMM state transition matrix from predicted labels of the unlabeled dataset is estimated. Estimation of class probability distributions for test dataset is done using trained random forest, KNN, XGB, Decision tree, ANN and SVM. Prediction of the most likely sequence of states for each session in the test dataset is done using HMM algorithm. HMM as the name suggests is defined by hidden states, state transition probabilities, possible observations and their emission probabilities. In the problem at hand, the HMM parameters are Hidden states drawn from the categorical distribution of the classification class labels, State transition probabilities represented by the state transition matrix, Possible observations drawn from the categorical distribution of

the classification class labels, Emission probabilities estimated by the prediction probability estimates. The HMM revises the predictions accordingly to their uncertainty and the state transition matrix estimated from unlabeled data using the Viterbi algorithm. Accuracy is used as the evaluation metric.

The HMM provided an increase in accuracy of 7.56 % (absolute percentage). By varying the properties of the dataset one can obtain an accuracy increase that ranges between zero and 10%, depending on how much data one is using for training the classifier and the transition matrix and the noise standard deviation. In general, the HMM is rather robust: either it provides a better accuracy or accuracy remains the same.

Forecasting the high priority tickets volume quarterly and annually helps in being better prepared with resources and technology planning for the service providers. The tickets from the IDM dataset are analyzed. The tickets range between the years 2014-2017. The predictor variable Open_Time is used with the count of tickets per date to forecast the values for the years 2020. Predicting the future values and the confidence interval between the Time ('2017-01-01') and ('2017-12-31') based on the historical dataset is as shown in Fig. 11 as one step ahead forecasting. Further, the larger steps ahead forecast is depicted in Fig. 12. up to the year 2025. The forecasting of tickets helps in identifying the monthly trends, find a realistic estimate of the number of tickets that will be generated, figuring out the hiring of human resources and procurement of equipment and reduce the aftermath of faults. Overall the service outages are well estimated and can be prevented.

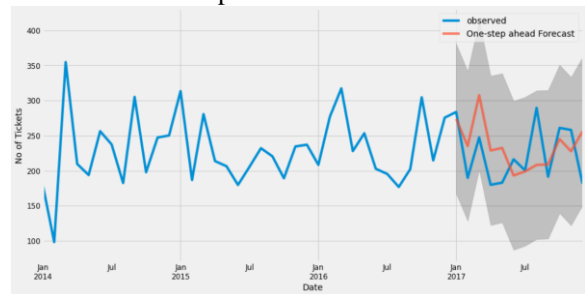


Fig. 11. Forecast of the ticket volume for next 6 months with one step ahead and next steps

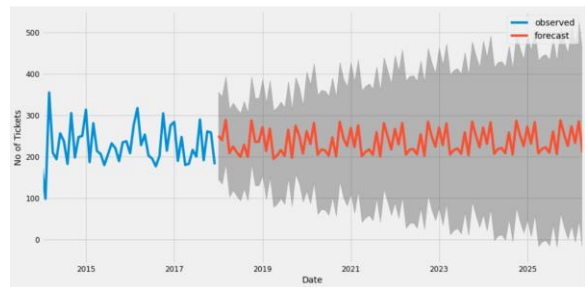


Fig. 12. Forecast of the ticket volume for next 5 years.

Inferences:

Firstly the prediction algorithms are applied to the dataset to predict the high priority tickets to prevent

further faults and avoid a service outage. The accuracies of various algorithms are compared. The dataset is used to build a Bayesian Network which provides the state transition matrices for HMM computation. The HMM modeling is used which improves the prediction accuracy between 7-10 percent. Time series forecasting is used to forecast the volume of tickets for future consecutive years for good preparedness of the service providers.

VII. CONCLUSION

Fault detection is carried out by using Bayesian Network and HMM to build a Service Outage Prediction model on the customer tickets and log data. The SOP model based on Bayesian Network is constructed using the training data from the IDM ticket dataset. The predicted Conditional Probabilities and the variable's dependencies are identified by the Bayesian Network. Based on the information obtained from the network, the matrices of state alteration probabilities and emission probability are calculated and initialized, where the parameters are considered as HMM constraints for building the model. Fault detection initially involves using several classifiers like RF, XGBoost, KNN, SVM, ANN and Decision Tree and which analyze the IDM tickets dataset and predict high priority tickets that cause maximum disruption to the networks. Further, the dataset is subjected to a Bayesian Network-based HMM classifier which on being applied to the ticket dataset exhibits better performance w.r.t accuracy and has a higher capability of predicting high priority tickets. Thus the prediction of high priority tickets helps the service providers from detecting faults at the earliest to prevent a service outage. The work can be expanded by adding additional warning types in the Syslogs and by establishing significant relationship among the different types of equipment failure. This includes association between various equipments failure and real time failure detection to ease the service outage management. With the Internet of Things (IoT) becoming a part of daily life and our environment, rapid growth in the number of connected devices is expected. With this growth, fog computing, along with its related edge computing paradigms is seen as promising solutions for handling the large volume of security-critical and time-sensitive data that is being produced by the IoT. Equipment failure detection can be extended to the IoT devices to support fog computing analytics.

CONFLICT OF INTEREST

The authors declare no conflict of interest

AUTHOR CONTRIBUTIONS

The second author is my guide and research supervisor. She has equal contribution in my thesis and paper work.

REFERENCES

- [1] L. H. Shuan, G. Xiaoning, T. Y. Fei, S. W. King, and L. Z. Mein, "Network equipment failure prediction with big data analytics," *International Journal of Advances in Soft Computing & Its Applications*, vol. 8, no. 3, 2016.
- [2] J. S. Tan, A. H. L. Lim, and C. K. Ho, "Predicting network faults using random forest and C5.0," *International. J. Eng. Technol.*, vol. 7, no. 2, pp. 93–96, 2018.
- [3] P. Hofmann and Z. Tashman, "Hidden markov models and their application for predicting failure events," In *Computational Science – ICCS 2020. ICCS 2020. Lecture Notes in Computer Science*, V. V. Krzhizhanovskaya, et al., Eds., vol. 12139. Springer, Cham, 2020.
- [4] F. Salfner and M. Malek, "Using hidden semi-markov models for effective online failure prediction," in *Proc. 26th IEEE International Symposium on Reliable Distributed Systems (SRDS 2007)*, Beijing, China, 2007, pp. 161-174.
- [5] A. Martino, G. Guatteri, and A. M. Paganoni, "Hidden markov models for multivariate functional data," *Statistics & Probability Letters*, vol. 167, 2020.
- [6] Z. Deljac and M. Kunstic, "A comparison of methods for fault prediction in the broadband networks," in *Proc. International Conference on Software, Telecommunications and Computer Networks*, 2010.
- [7] J. Shrijit and P. Vir, "Investigating hidden Markov models capabilities in anomaly detection," in *Proc. Annual Southeast Conference*, 2005, pp. 98-103.
- [8] J. Ruchi and A. Nasser, "A comparative study of hidden markov model and support vector machine in anomaly intrusion detection," *Journal of Internet Technology and Secured Transaction*, pp. 176-184, 2013.
- [9] T. Allen, C. Wai-Ki, and L. Y. Chan, "Detection of machine failure: Hidden markov model approach," *Computers & Industrial Engineering*, pp. 608-619, 2009.
- [10] N. Nguyet, "An analysis and implementation of the hidden markov model to technology stock prediction," *Risks*, 2017.
- [11] N. Muthumani and T. Antony, "Optimizing hidden markov model for failure prediction—comparison of Gaine's optimization and Minimum message length estimator," *International Journal on Computer Science and Engineering*, 2011.
- [12] A. Martino, *Giuseppina Guatteri and Anna Maria Paganoni*, Hmhd Package: Hidden Markov Model for High Dimensional Data
- [13] S. Felix, Predicting Failures with Hidden Markov Models, 2005.
- [14] W. Liang, J. Long, Z. Chen, X. Yan, Y. Li, Q. Zhang, K. C. Li, "A security situation prediction algorithm based on hmm in mobile network," *Wireless Communications and Mobile Computing*, vol. 2018, p. 11, 2018.
- [15] Martino, Guatteri, Paganoni, Hmhd Package: Hidden Markov Model for High Dimensional Data.
- [16] B. Esmael, A. Arnaout, R. K. Fruhwirth, and G. Thonhauser, "Improving time series classification using hidden markov models," in *Proc. 12th International Conference on Hybrid Intelligent Systems*, Pune, India, 2012, pp. 502-507.

- [17] H. Satu and H. Jouni, Using the seqHMM Package for Mixture Hidden Markov Models, 2015.
- [18] Z. Y. Liu and H. Qiao, "Hidden markov model based intrusion detection," in *Intelligence and Security Informatics. WISI 2006. Lecture Notes in Computer Science*, H. Chen, F. Y. Wang, C. C. Yang, D. Zeng, M. Chau, K. Chang, Eds., vol. 3917. Springer, Berlin, Heidelberg, 2006.
- [19] M. Alhamza, O. R. Razif, A. Ali, and A. Mohammed, A Performance Study of Hidden Markov Model and Random Forest in Internet Traffic Classification, 2016.
- [20] Z. Shu-xu, W. Hong-wei, and L. Chang-rong, "Traffic flow prediction based on optimized hidden Markov model," *Journal of Physics: Conference Series*, 2019.
- [21] C. M. Carbery, R. Woods, and A. H. Marshall, "A bayesian network based learning system for modelling faults in large-scale manufacturing," in *Proc. IEEE International Conference on Industrial Technology*, Lyon, France, 2018, pp. 1357-1362.
- [22] K. Nancy and V. Kansal, "An efficient data mining approach to improve students' employability prediction," *International Journal of Computer Applications*, vol. 178, pp. 29-35, 2019.
- [23] M. V. Kotpalliwar and R. Wajgi, "Classification of attacks using Support Vector Machine (SVM) on KDDCUP'99 IDS database," in *Proc. Fifth International Conference on Communication Systems and Network Technologies*, Gwalior, India, 2015, pp. 987-990.
- [24] S. Taruna and S. Hiranwal, "Enhanced Naïve Bayes Algorithm for Intrusion Detection in Data Mining, 2013.
- [25] D. Nagaraju, *et al.*, "Intrusion detection system using bayesian network and hidden markov model," *Procedia Technology*, vol. 4, pp. 506-514, 2012.
- [26] M. Altintas and A. C. Tantug, "Machine Learning Based Ticket Classification in Issue Tracking Systems, 2014.
- [27] R. Polikar, "Ensemble based systems in decision making," *IEEE Circuits and Systems Magazine*, Third Quarter 2006.
- [28] Data Science Live Projects, Sep 24 2018, DataMites ITSM Machine Learning ITSM=ML-PR0012.
- [29] ITSM incidents Prediction. [Online]. Available: <https://www.kaggle.com/asvvisb/itsm-incidents-pred>
- [30] R. Aleksandra, B. Krisztian, and M. Vera, "IT ticket classification: The simpler, the better," *IEEE Access*, 2020.
- [31] L. R. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, no. 2, pp. 257-285, Feb. 1989.
- [32] P. Cunningham and S. J. Delany, "k-Nearest neighbour classifiers," *Multiple Classifier Systems*, vol. 34, pp. 1-17, 2007.
- [33] S. Tsang, B. Kao, K. Y. Yip, W. S. Ho, S. D. Lee, "Decision trees for uncertain data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 23, pp. 64-78, 2009.
- [34] V. G. Marcel and B. Sander, "Editorial: Artificial neural networks as models of neural information processing," *Front Computational Neurosci.*, 2017.
- [35] H. White, "Learning in artificial neural networks: a statistical perspective," *Neural Comput.*, vol. 1, no. 4, pp. 425-464, 2014.
- [36] L. Breiman, "Random forests," *Mach Learn*, vol. 45, no. 1, pp. 5-32, 2001.
- [37] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD International Conference on Knowledge*, 2001.
- [38] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Ann Stat.*, vol. 29, no. 5, pp. 1189-232, 2001.
- [39] D. S. Batista, *Hidden Markov Model and Naïve Bayes Relationship*, Blog Published, 2017.

Copyright © 2022 by the authors. This is an open access article distributed under the Creative Commons Attribution License ([CC BY-NC-ND 4.0](https://creativecommons.org/licenses/by-nc-nd/4.0/)), which permits use, distribution and reproduction in any medium, provided that the article is properly cited, the use is non-commercial and no modifications or adaptations are made.



Ms. Sunita A Yadwad is a Research Scholar at Andhra University with over 22 years of teaching experience. An MTech in Computer Science and Engineering and a BE in CSE her research interests mainly focus on Data Analytics and Machine learning.



Dr Valli Kumari Vatsavayi, a Professor at Andhra University with over 27 years of experience in teaching and research, holds a PhD degree with Gold medal for best research. An MTech in Computer Science and Technology and a BE in ECE she researches on Block Chain, Computer Vision, Data Analytics, Security, Privacy, Forensics and Threat Intelligence. She is the Principal Co- Investigator of the Cyber Security and Data Analytics Centre sponsored by ISEA Project, MEITY, GOI.