

# Building Intelligent Conveyor System using classification techniques in a logistics Industry



Poornachand P, Vijayaramaraju Poosapati, Abinash Tripathy, Vedavathi Katneni

**Abstract:** Machine learning techniques plays an important role in knowledge discovery and assists humans in decision making. They help to detect patterns and predict the actions/outcome. In a complex industrial environment mode of operations of a machine depends on various internal and external parameters which are often done using expert judgement method which is not accurate and results in machine breakdown thereby resulting in unplanned outage. In this paper, we discussed and demonstrated how machine learning algorithms can help to handle conveyor systems autonomously in an optimum way without any human intervention. A conveyor belt system operational data is used to select the appropriate classification technique for the selected dataset. The details of the dataset collected, algorithms used and the test results are discussed in this paper.

**Keywords:** Classification, Machine Learning Algorithms, Industrial automation, Autonomous conveyor systems

## I. INTRODUCTION

Over last few years, with the widespread usage of industry 4.0 technologies the demand for 100% availability of industrial systems have increased steadily. High productivity and low cost of operations are the key focus areas of all the industrial systems and they are the key differentiators to stay ahead in the competitive global markets. The availability of a machine is highly depended on the expertise of the person handling it and due to lack of experience and inappropriate use of the machine could result in unplanned outage [1]. The need to automate the decision making process led to automation of technical processes of an industry which resulted in more complex automation systems in industry [2]. Preparing a machine to perform a specific task autonomously based on the input data is the primary function of Machine learning algorithms. There are many number of machine learning techniques that are developed to learn various characteristics of a machine based on the past data and help to predict the possible best action that can be taken in a given scenario. The dataset collected in the industrial environment from multiple sensors or industrial ERP, SCADA, other data

gathering and processing systems contains a various features and after processing the huge amount of data [4] the output is grouping based on the combination of multiple features available.

The data collected is divided into two sets, one for training the machine learning algorithm to identify the features and build a model with its combinations, and the other to test the built model to assess the performance of the developed model. The two sets are referred as training dataset and test data set. There are multiple tasks involved in a machine learning like classification, regressions and clustering and the kind of data collected is classified as supervised, un supervised and semi supervised data. The machines that are considered for automation are critical to industrial operations and availability of these machines plays a vital role in improving the productivity of the industry. The objective of the automation systems is to ensure that the machine is operated effectively and results in zero unplanned downtimes [3].

In this paper, the machine learning classification algorithms namely Logistic regression, K nearest Neighbors, Random Forest, Support vector machine, Decision Tree, Linear discriminant analysis and Gaussian Naïve bayes are executed on conveyor belt system data to identify the best classification algorithm to build a model for incorporating intelligence into a conveyor system.

## II. RELEATED WORK

Classification technique is a widely used to train the machines to classify the data into possible action depending upon the patterns or the features detected in the data, over last few years many researchers have discussed the concept of classification and discussed unique advantages of each of the techniques on a given dataset. The performance of algorithms varies depending upon the data set and hence the algorithms has to be chosen based on the performance and not by just the name. N. S. Ketkar et. al. used logistic regression, C4.5 decision tree, K nearest neighbour and Naïve Bayes in both standard and boosted forms to predict the class members for an online community dataset[5]. R. Dixit et. al. compared different classification techniques on credit scoring data sets, they have used classification techniques such as logistic regression, neural networks, decision trees, support vector machines, gradient boosting and random forests to predict the loan defaulters using credit dataset[6]. P. Delimate et. al. [7] used classification in a bio-informatics dataset and demonstrated the importance of rule based decision trees for classification of data.

Revised Manuscript Received on October 30, 2019.

\* Correspondence Author

**Poornachand P**, Computer Science Department, Raghu Engineering College, Visakhapatnam, India. Email: poornachand046@gmail.com

**Vijayaramaraju Poosapati\***, Computer Science Department, GITAM University, Visakhapatnam, India, Email: Vijayaramaraju.poosapati@gmail.com

**Abinash Tripathy**, Professor, Raghu Engineering College, Visakhapatnam, India. Email: abi.tripathy@gmail.com

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](https://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

T. R. Patil et. al. [8] compared polynomial time complexity and inhibitory decision rules on five different data sets and experiments are performed to evaluate the classifiers performance on the basis of different parameters. M. Esmaili et. al. [9] compared hypothetical target classification analysis methods to demonstrate how pre-processing the data can help to protect the confidence of the result in a multi quantization, Boolean and fuzzy techniques. D. D. Arifin et. al. [10] used naïve bayes classifier to detect spam in the SMS received on mobiles using the data set collected from multiple mobiles. Another similar study on spam filtering is carried out by X. Zhuang et. al. [11] by using a unified score propagation model for detecting the spam, Further O. F. Arar et. al. [12] performed spam detection using Naive Bayes approach to predict the software defects. L. Jiang et. al. [13] used naïve bayes classification technique for text classification. Y. An et. al. [14] studied the correlation of emotion in music with lyrics using naïve bayes classification techniques. H. Lad et. al. [15] studied feature based classification of digital images using tagging techniques. H. Zhang et. al [16] by using phenology features tried to classify the crop data for estimation of acreage .Mucherino, A., et. al. [17] used k- nearest neighbor classification techniques to classify the agricultural data using euclidean distance hamming distance and manhattan distance methods. Unda-Trillas, E. et, al [18] described the methods to build classification and regression trees(CART) which uses Gini Index as impurity index that is a generalization of binomial variance. Berk, R.A et. al [19] used regression techniques along with classification to identify the patterns using statistical learning. Wang, Z., et. al. [20] used multi class super vector machines to build a model for machine applications The discussion of various studies above reveals that no single algorithm is best chosen for all kinds of datasets and hence the intelligent model has to be built based on the accuracy, precision recall and F1 score values given by the classification algorithms.

III. PROPOSED MODEL

The model developed includes all processes from data collection to insights prediction. It is detailed in four steps. 1. Data collection, 2. Data processing 3. Classification and 4. Visualization & Evaluation. The algorithm which gives high accuracy and with good learning improvement rate would be selected for building an autonomous machine. Figure 1 demonstrates the model proposed in this paper.

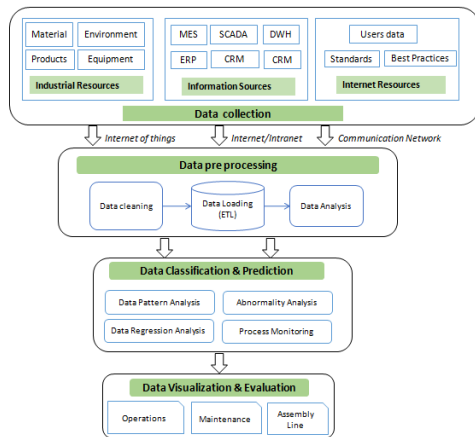


Fig. 1: Proposed model to build Intelligent Conveyor Systems

A. Data collection

Data is collected from industrial information systems using sensors, IoT devices & communication channels which gather huge volume of data, the amount of accurate data collected helps to improve the prediction accuracy of classification algorithms. However collecting huge volume of data to identify the most accurate classification model is very complex. The dataset can be prepared in multiple ways either by stimulation or by collecting the data from test bed. In this paper, to test the accuracy of classification algorithms, we have taken the data from a logistics industry which collected data from conveyor systems to build intelligence.

In the current data set, various parameters related to conveyor belt system and internal industrial business logics are captured, there are 52 columns which gives information about the condition of the conveyor belt such as vibration, temperature, sound and pressure at different locations of the belt and spindle are captured. The corresponding actions taken by operator for the specific data set at a given time are also captured as the end result. There are around 19853 records in the collected dataset which is used for training and testing the performance of various classification algorithms on the data.

To represent the data, a data frame is created using the pandas library and the result is represented as

```
In [4]: df1.head()
Out[4]:
```

	LocAve	LocBve	LocCve	LocDve	LocEve	LocFve	LocAac	LocBac	LocCac	LocDac	...	SpnVib	SpnCv	SpnDv	SpnEv	SpnPv
0	-0.000055	-0.000045	0.000314	7.570000e-07	0.000023	-0.000402	0.024683	0.024728	0.024414	0.021968	...	-1.5037	-1.5036	-1.4988	-1.4987	-1.4959
1	-0.000002	0.000005	-0.000026	-5.980000e-06	-0.000045	-0.000359	0.014573	0.014567	0.014603	-0.026396	...	-1.4978	-1.4979	-1.4970	-1.4970	-1.4958
2	-0.000005	0.000004	-0.000011	2.770000e-07	0.000004	0.000017	-0.030446	-0.030449	-0.030439	-0.017251	...	-1.4994	-1.4995	-1.4984	-1.4984	-1.4955
3	-0.000005	0.000034	-0.000340	1.980000e-06	0.000087	-0.000056	0.013359	0.013365	0.013705	-0.036518	...	-1.4976	-1.4975	-1.5000	-1.5000	-1.5003
4	-0.000002	0.000024	-0.000157	3.050000e-06	0.000011	0.000068	0.030986	0.030972	0.031020	0.091760	...	-1.5067	-1.5064	-1.4986	-1.4987	-1.4958

Fig. 2: Sample Data Representations

The data contains the data like temperature, pressure, vibration, accelerometer and RPM captured by various sensors from 6 different locations of the conveyor belt systems and different spindles attached to them. The condition of the belt can be determined based on these parameters and this data is combined with industrial data that comes from ERP's to capture the information of fleet movement, distance to source and various other information related to vendor and the handling capacity of the fleet. The actions taken by human at a given scenario from multiple such types of conveyor systems in the industry are taken to train the model to build an intelligent conveyor system.

The collected dataset is further divided into train and test data in a ratio of 66:18, where as the other 15% of data is considered for evaluating the relearning rate of the model in a conveyor system. The actions of the dataset are classified into 3 different classes as 1,2,3 which are the level of speed in which the belt is operated by the human at a point when the action was taken.

B. Pre processing

The data collected requires preprocessing to eliminate the junk values or to remove null records or to scale the data to match the other data items in the dataset. This process is done in preprocessing step where all the non-numeric data is eliminated. In doing preprocessing the importance of the data to predict the outcome is considered and based on that it is either eliminated or transformed using scaling techniques.



The preprocessing step helps to reduce the time required for the model to get trained and also improves the accuracy of the algorithm. For the current dataset, by using Seaborn library's heat map, we check the data to ensure that there are no null values in our dataset. The heat map is used to look at the visual representation of data. Post removing the null values, the dataset is scaled by using the Standard Scaler function of sklearn.preprocessing.

```
from sklearn.preprocessing import StandardScaler
scaler_train = StandardScaler()
scaler_train.fit(df_train.drop("activity", axis=1))
scaled_features_train=
scaler_train.transform(df_train.drop("Activity", axis=1))
df_feat_train = pd.DataFrame(scaled_features_train,
columns=df_train.columns[:-1])
```

This step will ensure that all data is in the same scale for plotting and comparisons.

C. Data classification & Prediction

A classification algorithm helps to detect patterns in the data without much knowledge on the underlying complexity in the algorithms. Classification is a supervised learning method which helps to predict the category to which the possible output would belong to. Following four classification techniques are considered to identify the best algorithm to build the model and then learning rate of the model is determined.

I. K nearest Neighbors

K nearest Neighbors - KNN is an instance-based learning algorithm that stores all available data points and classifies the new data points based on similarity measure such as distance. 13237 records are used to train the model and the model is tested based on 3500 records. Using the confusion\_matrix function of the sklearn.metrics, we can print the confusion\_matrix on the results of the model run on our test dataset. The results of the K nearest Neighbors are represented as below. The accuracy, precision, Recall and F1 Score are as follows.

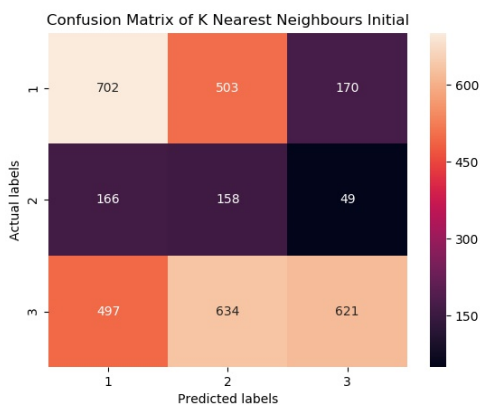
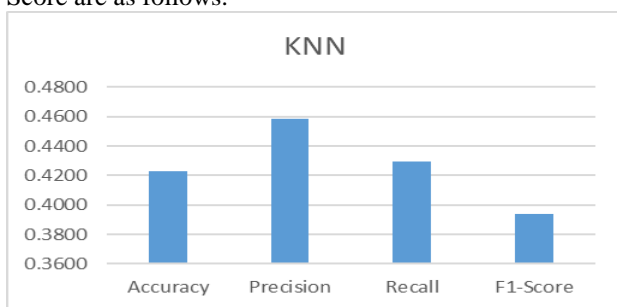


Fig. 3: K-Nearest neighbors results and confusion matrix on dataset

II. Logistic regression

It is a predictive analysis algorithm and based on the concept of probability, which uses a complex cost function called a 'Sigmoid function' instead of a linear function. The dataset prepared, in ran using a scikit lean Logistic regression model. The same dataset and test set used for K nearest neighbor algorithm are used for this model too.

The results of the logistic regression are represented as below. The accuracy, precision, Recall and F1 Score are as follows.

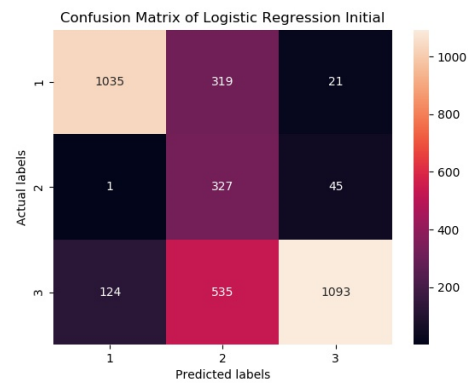
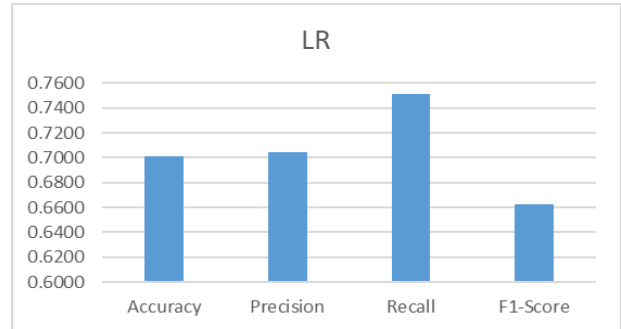
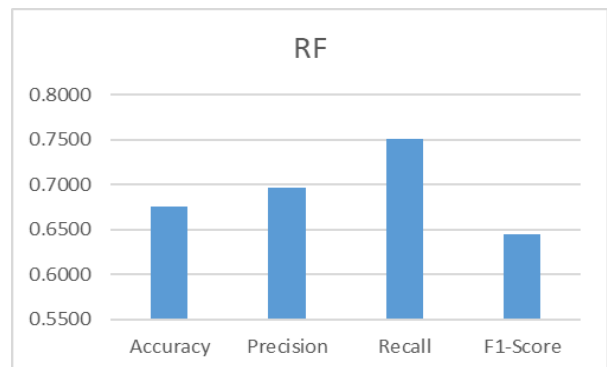


Fig. 4: Logistic regression results and confusion matrix on dataset

III. Random Forest

Random Forest - Random forest algorithm works by constructing multiple decision trees on various sub-samples of the datasets and output the class that appear most often or mean predictions of the decision trees. Using the same dataset and test set a confusion matrix is generated to predict the prediction accuracy of the Random Forest algorithm. Below is the result



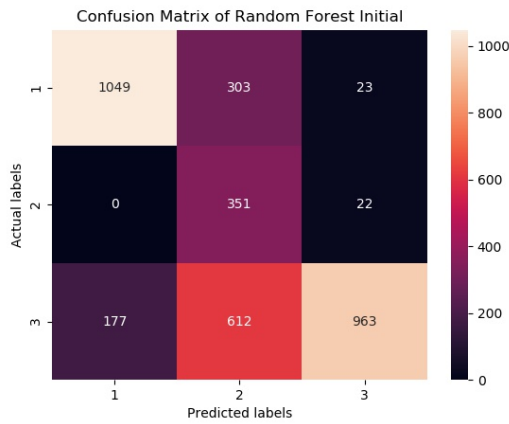


Fig. 5: Random Forest results and confusion matrix on dataset

IV. Support vector machine

Support vector machine - SVM discriminates a set of high-dimension features using a set of hyperplanes that gives the largest minimum distance to separates all data points among classes.

For the data set, we run the SVM classifier by using the SVC function of sklearn.svm using the default parameters

The precision and the recall values, confusion matrix for the dataset using confusion matrix for the dataset used in previous algorithm is shown below

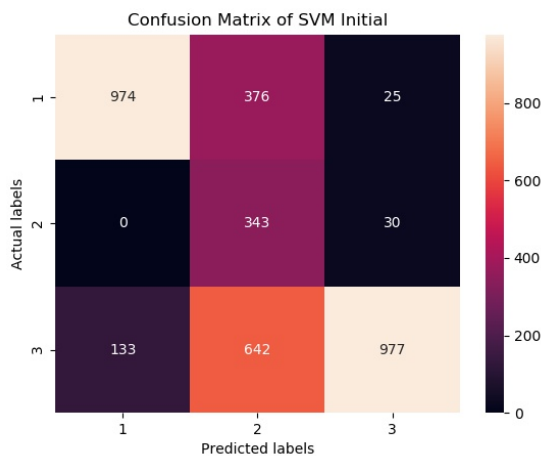
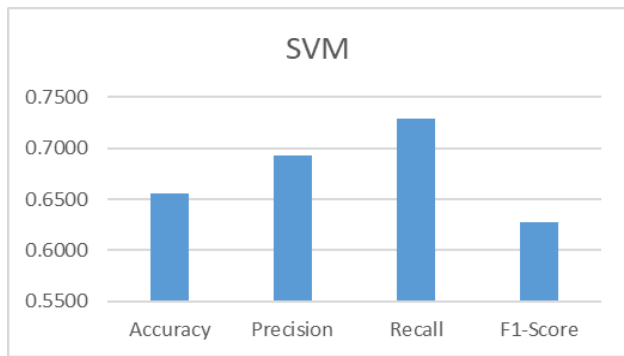


Fig. 6: Support Vector Machine results and confusion matrix on dataset

D. Data Visualization and Evaluation

Once the model is built, evaluating the best model to suite the kind of data is key for successful implementation. The data model is accurate if all the true positives are classified as class Y and all the true negatives as class N, any other combination results in an incorrect result which will impact the accuracy of the data model.

The below fig. represents the predicted class vs actual class for each of the class available. In the current dataset, 3 classes are represented as below.

True positives are normally denoted by ‘TP’, True negatives denoted by ‘TN’. False positives denoted by ‘FP’ and False negatives are denoted by ‘FN’. As we have 3 classes in the given dataset, the class representation would vary from class to

Table 1: Data Visualization for Class 1

	Class 1	Class 2	Class 3
Class 1	True Positive	False Negative	False Negative
Class 2	False Positive	True Negative	True Negative
Class 3	False Positive	True Negative	True Negative

Table 2: Data Visualization for Class 2

	Class 1	Class 2	Class 3
Class 1	True Negative	False Positive	True Negative
Class 2	False Negative	True Positive	False Negative
Class 3	True Negative	False Positive	True Negative

Table 3: Data Visualization for Class 3

	Class 1	Class 2	Class 3
Class 1	True Negative	True Negative	False Positive
Class 2	True Negative	True Negative	False Positive
Class 3	False Negative	False Negative	True Positive

The objective of any model is to reduce the number of false negatives and false positives so that the overall accuracy of the model increases.

With these parameters we can calculate the accuracy, precision, recall and f1 score of a given model using the following formulas.

Accuracy – It is a ratio of correctly predicted observations vs the total observations. Accuracy is the most intuitive performance measure and it is simply a ratio of correctly predicted observation to the total observations.

$$\text{Accuracy (A)} = (TP+TN) / (TP+FP+FN+TN)$$

Precision – it is the ratio of correctly predicted positive observations out of the total predicted positive observations. High precision value means that the model is working pretty well.

$$\text{Precision (P)} = TP/TP+FP$$

Recall or often referred as Sensitivity is the ratio of correctly predicted positive observations to the total number of observations of a class Y.

$$\text{Recall (R)} = TP/TP+FN$$

F1 score is another measure which is the weighted average of Precision and Recall. It is calculated by

$$\text{F1 Score} = (2 * R * P) / (R + P)$$



	Accuracy	Precision	Recall	F1-Score
KNN	0.4231	0.4585	0.4295	0.3937
LR	0.7014	0.7041	0.7511	0.6628
RF	0.6751	0.6961	0.7512	0.6442
SVM	0.6554	0.6929	0.7285	0.6274

Table 4: Test dataset results on 13237 records

#### IV. MODEL EVALUATION & RESULTS

The precision, recall and f1score comparisons of various classification algorithms run on a data set with around 19K+ rows of dataset, taking 66% as train data and 18% as test data are shown above. The learning rate improvement of the model is calculated using the 15% test data set. The logistic regression model is used to build the model to automate the operation of the conveyor system as it predicted the results with 70% accuracy. Accuracy, Precision, Recall and F1-score with training dataset of 16737 is given below. The data is tested with 3116 records

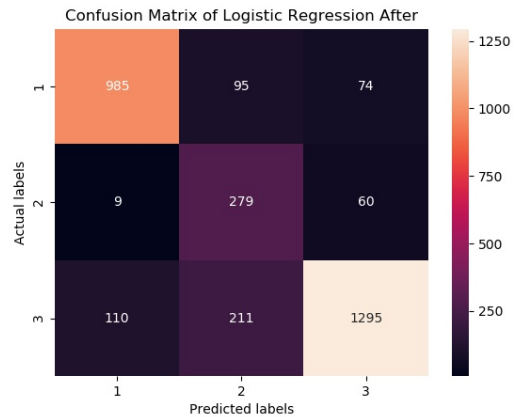


Fig. 7: Logistic Regression learning with 16K+ Train dataset.

	Accuracy	Precision	Recall	F1-Score
Logistic Regression(With 13237 records)	0.7014	0.7041	0.7511	0.6628
Logistic Regression (With 16353)	0.8207	0.7585	0.8189	0.7737

Table 5: Test dataset results on 13K+ records

#### Learning Rate Improvement:

The model relearning rate is improved by 17% for dataset with 3500 more records in training dataset.

Based on the above results, Logistic regression classification is chosen to automate the conveyor system operations in a industrial setup depending on it accuracy and the learning improvements demonstrated.

#### V. CONCLUSION

Human action prediction is an essential component of building an intelligent conveyor system. The data from industrial systems is the key to design a model with high accuracy, in this paper we have built a model using Logistic regression classification on the conveyor system dataset. However with change in industrial systems and parameters the performance of the selected technique would change. Also, limited number of records are considered to train the classification model, however with increase in the records and variety of data, the accuracy of the algorithm might change and required to be revalidated frequently.

#### REFERENCES

- H. Mubarak and P. Göhner, "An agent-oriented approach for self management of industrial automation systems," 8th IEEE International Conference on Industrial Informatics, p. 721, 2010.
- B. Botcher, J. Badinger and O. Niggemann, "Design of industrial automation systems — Formal requirements in the engineering process," 6th IEEE International Conference on Emerging Technologies & Factory Automation , pp. 1-4, 2013.
- J. Laprie, "Dependability – Its Attributes, Impairments and Means," published in Predictably Dependable Computing Systems, p. 4, 1995.
- W. Meeker and Y. Hong, "Reliability Meets Big Data: Opportunities and Challenges," Quality Engineering, vol. 26, pp. 102-116, 2014.
- N. S. Ketkar, L. B. Holder, D. J. Cook, "Empirical comparison of graph classification algorithms", IEEE Symposium on Computational Intelligence and Data Mining, Nashville, USA, pp. 259-266, March 30-April 2, 2009
- P. Delimata, B. Marszał-Paszek, M. Moshkov, P. Paszek, A. Skowron, Z. Suraj, "Comparison of some classification algorithms based on deterministic and nondeterministic decision rules", in: Transactions on Rough Sets XII, Springer, pp. 90-105, 2010
- R. Dixit, H. Singh, "Comparison of detection and classification algorithms using boolean and fuzzy techniques", Advances in Fuzzy Systems, Vol. 2012, Article No. 406204, 2012
- T. R. Patil, V. Thakare, S. Sherekar, "A Combined Naïve Bayes and URL Analysis Based Adaptive Technique for Email Classification", International Journal of Electronics, Communication and Soft Computing Science & Engineering, Special Issue: International Conference on "Advances In Computing, Communication and Intelligence", pp. 88-90, 2014
- M. Esmaceli, A. Arjomandzadeh, R. Shams, M. Zahedi, "An Anti-Spam System using Naive Bayes Method and Feature Selection Methods", International Journal of Computer Applications, Vol. 165, No. 4, pp. 1-5, 2017
- D. D. Arifin, M. A. Bijaksana, "Enhancing spam detection on mobile phone Short Message Service (SMS) performance using FP-growth and Naive Bayes Classifier", IEEE Asia Pacific Conference on Wireless and Mobile, Bandung, Indonesia, pp. 80-84, September 13-15, 2016
- Vijayaramaraju Poosapati, Vijaya Killu Manda, Vedavathi Katneni, "Cognitive Automation Opportunities, Challenges and Applications". International Journal of Computer Engineering and Technology, 9(5), 2018, pp. 89-95



12. X. Zhuang, Y. Zhu, C.-C. Chang, Q. Peng, F. Khurshid, "A unified score propagation model for web spam demotion algorithm", *Information Retrieval Journal*, Vol. 20, No. 6, pp. 547-574, 2017
13. O. F. Arar, K. Ayan, "A Feature Dependent Naive Bayes Approach and Its Application to the Software Defect Prediction Problem", *Applied Soft Computing*, Vol. 59, pp. 197-209, 2017
14. L. Jiang, C. Li, S. Wang, L. Zhang, "Deep feature weighting for naive Bayes and its application to text classification", *Engineering Applications of Artificial Intelligence*, Vol. 52, pp. 26-39, 2016
15. Y. An, S. Sun, S. Wang, "Naive Bayes classifiers for music emotion classification based on lyrics", *IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS)*, Wuhan, China, pp. 635-638, May 24-26, 2017
16. H. Lad, M. A. Mehta, "Feature Based Object Mining and Tagging Algorithm for Digital Images", in: *Proceedings of International Conference on Communication and Networks*, Singapore, *Advances in Intelligent Systems and Computing*, Vol. 508, pp. 345-352, 2017
17. H. Zhang, Q. Li, J. Liu, J. Shang, X. Du, L. Zhao, N. Wang, T. Dong, "Crop classification and acreage estimation in North Korea using phenology features", *GI Science & Remote Sensing*, Vol. 54, No. 3, pp. 381-406, 2017
18. Mucherino, A., Papajorgji, P.J., Pardalos, P.M.: k-nearest neighbor classification. In: Mucherino, A., Papajorgji, P.J., Pardalos, P.M. (eds.) *Data Mining in Agriculture*, pp. 83– 106. Springer, New York (2009). [https://doi.org/10.1007/978-0-387-88615-2\\_4](https://doi.org/10.1007/978-0-387-88615-2_4)
19. Unda-Trillas, E., Rivera-Rovelo, J.: A Method to Build Classification and Regression Trees. In: Bayro-Corrochano, E., Hancock, E. (eds.) *CIARP 2014*. LNCS, vol. 8827, pp. 448–453. Springer, Cham (2014). [https://doi.org/10.1007/978-3-319-12568-8\\_55](https://doi.org/10.1007/978-3-319-12568-8_55)
20. Berk, R.A.: *Classification and regression trees (CART)*. In: Berk, R.A. (ed.) *Statistical Learning from a Regression Perspective*, 129–186. Springer, Cham (2016). [https://doi.org/10.1007/978-3-319-44048-4\\_3](https://doi.org/10.1007/978-3-319-44048-4_3)



**Prof. Vedavathi Katneni** is presently working as Head of the Department, Computer Science Department, GITAM, Visakhapatnam. She is a life member of Indian Society of Technical Education, Computer society of India and Indian Science congress association. She completed her Ph.D. from Andhra University, Visakhapatnam. During her current tenure at GITAM, she was awarded with multiple awards from women empowerment cell of GITAM for her contribution towards empowering women and initiatives taken towards empowerment.

### AUTHORS PROFILE



**PoornaChand P** is currently pursuing M Tech in the department of computer science and engineering at Raghu Engineering College, Visakhapatnam. He is passionate to work on latest technologies like AI, IoT and Web application development and has built multiple application to cater the needs of small and middle scale industries. Machine Learning and Deep learning are his areas of interests. He is an active blogger

and mostly writes on advanced technology concepts.



**Vijayaramaraju Poosapati**, a research scholar from GITAM, Visakhapatnam. Prior to that he worked as a Sr. Project manager in Tech Mahindra and was responsible for driving Automation initiatives for one of the top fortune 500 companies. He received many accolades for his automation initiatives and his work was recognized as the best automation initiative of the year. He holds a master's degree from Sikkim

Manipal University. His previous research work involves building smart industrial machines by using cognitive intelligence.



**Dr. Abinash Tripathy** is presently working as associate professor in the department of computer science and engineering at Raghu Engineering College, Visakhapatnam. He has completed his Ph.D. from National Institute of Technology, Rourkela. He had completed his M.Tech. from Kalinga Institute of Industrial Technology, Bhubaneswar. His area of specialization are Sentiment Analysis, Natural Language processing,

Software Engineering.