Advanced topics in machine intelligence

**MSc Mechatronics and Intelligent Machines**

**Srikar Chavali**

**G20844756**

**School of Engineering, University of Central Lancashire**

**January 05, 2021**

Table of Contents

[1 Introduction 4](#_Toc60770088)

[1.1 Structure of the report 4](#_Toc60770089)

[2 Theoretical Background 4](#_Toc60770090)

[2.1 Machine learning models for fault classification 5](#_Toc60770091)

[2.1.1 K Nearest Neighbours: 5](#_Toc60770092)

[2.1.2 Decision trees (DT): 5](#_Toc60770093)

[2.1.3 Naïve Bayes (Bayesian Learner): 6](#_Toc60770094)

[2.2 Data optimization by Feature selection or Dimensionality reduction 6](#_Toc60770095)

[2.2.1 Minimum Redundancy Maximum Relevance (MRMR) Algorithm 6](#_Toc60770096)

[2.2.2 Neighbourhood component analysis (NCA): 6](#_Toc60770097)

[2.2.3 Principal component analysis (PCA): 6](#_Toc60770098)

[3 Methodology 7](#_Toc60770099)

[4 Code development 7](#_Toc60770100)

[5 Performance metrics 8](#_Toc60770101)

[6 Results and discussion 8](#_Toc60770102)

[7 Conclusions 11](#_Toc60770103)

[8 Appendix 12](#_Toc60770104)

[9 Referencs 13](#_Toc60770105)

# Introduction

This project will develop a machine learning algorithm with the necessary feature selection or dimensionality reduction techniques to predict the faults present in a transmission system

Computer revolution with the birth of machine learning grabbed the attention of scientists and they researched ways to implement these techniques to monitor the machine performance and safeguard their components. Machine learning models are used in the place of humans to remotely monitor the systems (LEI, 2015).

The machine learning algorithm in this study uses the fault data generated by simulating a transmission system modelled in Simulink to predict the presence of defects in the transmission. This study is important because, if successful the machine learning algorithm can be used for condition monitoring and predictive maintenance. In general, the data collected through sensors can be used to study the operating condition of the system through a machine learning algorithm and it is even possible to predict and classify the defects in the system to avoid more serious losses (Sternhagen et al, 2002) (Davis, 1987).

## Structure of the report

This report is organized as follows: Section 2 introduces the idea of machine learning and the machine learning algorithms used in this study. This section also explores some feature selection and dimensionality reduction techniques. Section 3 proposes the methodology followed in the study. Section 4 explains the developed code. Section 5 describes the considered performance metrics. Section 6 evaluates the performance of classification algorithms by comparing their performance metrics. Finally, Section 7 concludes the study and summarizes all the findings of the study.

# Theoretical Background

Machine learning is associated with “teaching” computers which action to perform without being programmed explicitly for all the possible situations (Singh et al, Mar 2016).

Machine learning algorithms are of 2 types

**Supervised learning:**

Supervised learning trains a model using known inputs and labelled outputs so that it can predict new data. These algorithms predict output based on previous learnings amid uncertainty. The training process is carried out until a satisfactory accuracy is reached.

**Unsupervised learning:**

Unsupervised learning works by finding patterns in input data. The input data is not labelled and there is no known output. Clustering is the most commonly used unsupervised machine learning model.

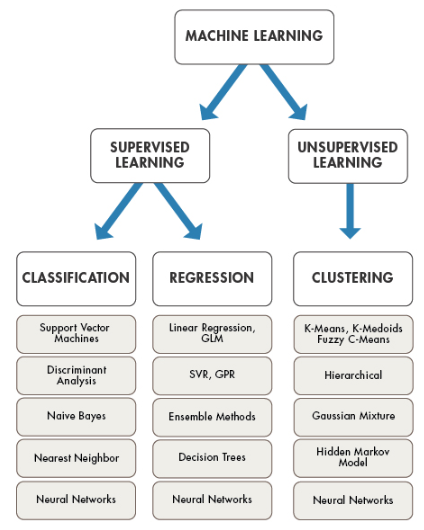


Figure 1 Machine learning models (The MathWorks)

## Machine learning models for fault classification

Fault classification is an important step in the process of prediction. The following classifiers are used in this study to predict the fault code of the transmission system.

### K Nearest Neighbours:

KNN uses the concept of feature similarity to make predictions. The class of the data point is the one which is most common among its k nearest neighbours (Tutorials Point,).

The Euclidean distance between any two points ‘p’ and ‘q’ can be calculated by using the formula

|  |  |
| --- | --- |
|  | (1) |

### Decision trees (DT):

The Decision tree works by growing iteratively by creating new nodes by branching the dataset into subsets based on the value of classification purity. The branching of data is stopped upon making the decision (Shahrtash and Jamehbozorg, Apr 2008). Decision Trees can identify and classify transmission line faults reliably (J et al, 2019).

The Gini index is used to calculate the node impurity. The Gini index describes the best path to split the features in the tree. (P. P. Wasnik et al, 2020).

|  |  |
| --- | --- |
|  | (2) |

### Naïve Bayes (Bayesian Learner):

The Bayesian learner is a probabilistic learning algorithm (Rish, 2001). The model works in two steps. In the training step the model estimates the probability distribution. In the prediction step the model calculates the posterior probabilities and classifies basing on the largest probability. This learning model is suitable for large data sets because of its low training time and high accuracy (Hasan et al, May 2017).

## Data optimization by Feature selection or Dimensionality reduction

A common problem in machine learning is identifying relevant predictors for developing a classification model. The most vital task of feature selection is to identify and eliminate irrelevant or redundant data to improve the accuracy of learning algorithms (Visalakshi and Radha, Dec 2014). It is necessary to reduce the number of predictors to minimize the computational cost and, in few cases, to improve the accuracy of the model by eliminating noise (Jason Brownlee, 2019).

The following Feature selection algorithms were used in this study:

### Minimum Redundancy Maximum Relevance (MRMR) Algorithm

MRMR finds the ideal set of predictors which are mutually and highly exclusive and can represent response variables efficiently (DING and PENG, 2011). MRMR tries to minimize the redundancy and maximize the relevance of the predictors with respect to the feature set.

### Neighbourhood component analysis (NCA):

Functionally, NCA resembles K-NN as it uses the same working principle. The aim of NCA algorithm is to learn a distance metric which maximizes the leave-one-out (LOO) classification performance by finding a linear transformation of input data (Jacob Goldberger et al, 2005)..

### Principal component analysis (PCA):

PCA reduces the dimensionality of the data which has several interrelated features or samples, while retaining the maximum possible variation present in the data. PCA does this by transforming the existing data into a new set of variables known as principal components (PCs) (I.T. Jolliffe, 2002).

# Methodology

MATLAB is used for developing, testing and training the machine learning model.

After the data is imported into the MATLAB workspace, it is split into training data and testing data. Following this, three classification models KNN, Naïve Bayes and Decision trees were trained, and the performance of the models was evaluated by calculating the accuracy.

Next the training data is optimized through feature selection/dimensionality reduction. MRMR, NCA and PCA algorithms were used to modify data.

The above three classification models were again trained and tested using the modified data. The accuracies and other performance metrics of the models are calculated.

Finally, confusion charts of the models were developed and compared to select the best machine learning model for the task.

# Code development

The first step is importing data into MATLAB workspace. This is done by using the “load” command. Next the imported data is shuffled and split into two parts: Training data (70%) and Testing data (30%). All the features are assigned to XTrain0 and XTest0, all the class labels are assigned to YTrain and YTest.

Three supervised learning algorithms KNN, Decision trees and Naïve Bayes are trained on the unaltered training data and accuracies of the models are calculated.

Following this step, the data is standardized. The “Zscore” function is used to standardize the data. The three models are again trained using the scaled data and the accuracies are calculated.

To further improve the data MRMR, NCA and PCA algorithms are applied on the training data. The output of MRMR algorithm is a vector whose elements are the indices of the features sorted in the order of highest significance. From this result vector the first five features are selected.

NCA estimates the feature weights of all the features. Basing on these feature weights the features are sorted in descending order and the top 4 features are selected.

The PCA algorithm transforms the training data into corresponding principal components. The first 9 principal components which represent 95% of variation present in training data are selected.

Finally, confusion charts are developed for the model to compare their performance.

# Performance metrics

To measure the performance of the classifiers four common metrics accuracy, precision, recall, F-measure are calculated. These performance metrics are calculated as follows:

|  |  |
| --- | --- |
|  | (3) |
|  |  |
|  | (4) |
|  |  |
|  | (5) |
|  |  |
|  |  |
|  | (6) |
|  |  |

Where TP = true positive; FP = false positive; TN = true negative; FN = false negative

# Results and discussion

In this study the accuracy is considered as the primary performance metric. Precision, recall, and F-measure are secondary performance measures. All the performance values are tabulated in table 1.

Table 1 Model performance comparison with original data

|  |  |  |  |
| --- | --- | --- | --- |
| Performance metric | Classifier | | |
| **KNN** | **Decision Trees** | **Naïve Bayes** |
| **Accuracy** | 0.35 | 0.94 | 0.64 |
| **Precision** | 0.93 | 0.97 | 0.97 |
| **Recall** | 0.91 | 0.99 | 0.98 |
| **F-Measure** | 0.92 | 0.98 | 0.97 |

From the initial analysis it can be observed that decision trees have the highest classification accuracy of the 3 learning models. To prevent the classifiers for being biased the training data was standardized. It can be observed form table 2 that even after standardizing the data decision trees has the highest accuracy. Several values for ‘K’ (number of nearest neighbours) were tried but the improvement in the accuracy of the KNN model was not significant.

Table 2 Model performance comparison with standardized data

|  |  |  |  |
| --- | --- | --- | --- |
| Performance metric | Classifier | | |
| **KNN** | **Decision Trees** | **Naïve Bayes** |
| **Accuracy** | 0.57 | 0.65 | 0.5 |
| **Precision** | 0.95 | 0.97 | 0.94 |
| **Recall** | 1 | 0.97 | 1 |
| **F-Measure** | 0.97 | 0.97 | 0.97 |

Based on the results from Table 1 and 2 the decision tree classifier was selected. and attempts were made to further improve the model. MRMR algorithm was used on the training data and the result suggested that five features had higher significance compared to others. From figure 3 it can be observed that the drop in the predictor score of first five predictors is large which suggests that the MRMR algorithm is confident about the significance of these features. So, these features were selected to train the model. However, no considerable improvement in accuracy was observed in the model. Several iterations of the same experiment were run and in most of the cases the decision tree classifier with and out MRMR had a similar accuracy. This suggests that only the first five significant features can be selected to predict the results instead of using all 18 features.

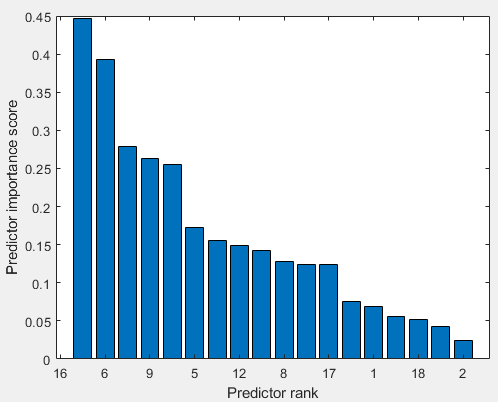


Figure 3 MRMR predictor ranking

Another feature selection algorithm NCA was used with the training data. NCA predicted the feature weights and based on their weights (see figure 4) top 4 features (11,14,15,18) were selected and the model was trained. NCA also had no considerable effect on the performance of the model.

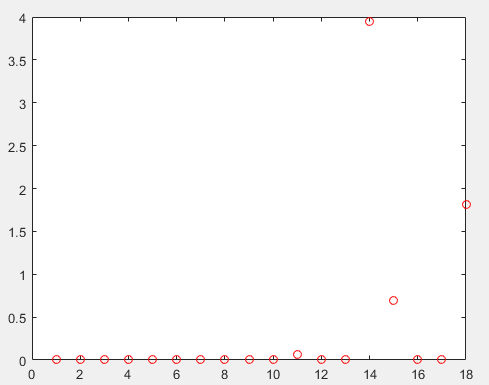


Figure 4 NCA feature weight prediction

Finally, PCA was used for feature reduction of the data. Upon training the model with PCs it was observed that the accuracy of the model decreased. This indicates that data transformation is not suitable for the given data.

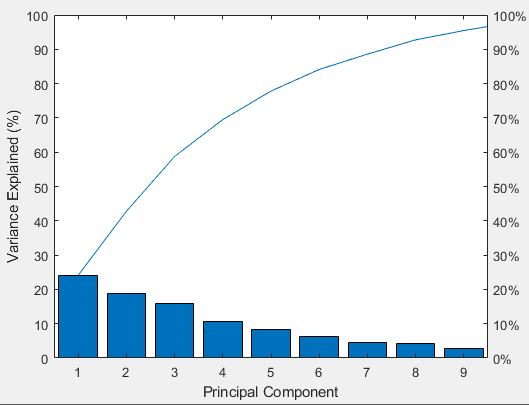


Figure 5 PCA scree plot

Table 3 performance comparison of Decision tree classifier

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Performance metric | **Original data** | **standardized data** | **MRMR 5 features** | **NCA 4 features** | **PCA 9 features** |
| **Accuracy** | 0.94 | 0.65 | 0.94 | 0.93 | 0.64 |
| **Precision** | 0.97 | 0.97 | 0.98 | 0.98 | 0.96 |
| **Recall** | 0.99 | 0.97 | 0.99 | 0.98 | 0.97 |
| **F-Measure** | 0.98 | 0.97 | 0.99 | 0.96 | 0.96 |

From Table 3 it can be observed that the accuracy of decision tree model with and without feature selection is almost same. So, to reduce the training and computation time decision tree classifier was trained on features selected using MRMR algorithm to predict results for actual test data.

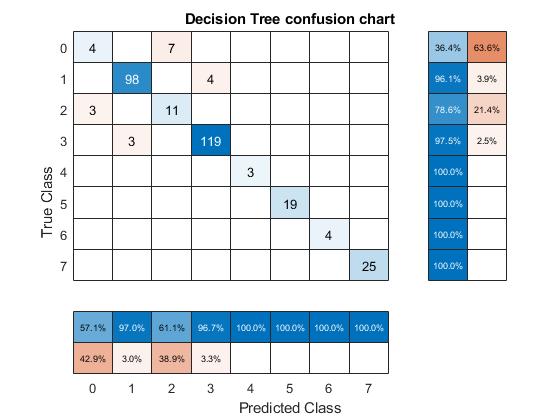


Figure 6 Confusion chart for Decision tree classifier using MRMR algorithm

# Conclusions

From the given data features having the highest significance for classifying faults were identified using MRMR and NCA techniques. MRMR predicted that 5 features played a key role for classification. Based on the feature weights NCA suggested 4 features were important for making predictions. The dimensionality of the data set was reduced by using PCA algorithm, this resulted in a decrease in accuracy of the models indicating that the models are not strongly correlated. Since MRMR had the highest accuracy in most of the trials it was chosen for data optimization.

Three machine learning classifiers KNN, Decision trees and Naïve Bayes were trained and tested on the data. The performance of the machine learning models was evaluated by comparing their performance metrics and by generating confusion charts. The experimental results suggest that Decision trees had the highest classification accuracy and precision among the three classifiers. Hence, decision tree classifier with MRMR algorithm was used make predictions for test data.

# Appendix

After running the code 18 figures are generated. The description of each figure are as follows

Figure 1: Confusion chart of KNN classifier with all data

Figure 2: Confusion chart of Decision tree classifier with all data

Figure 3: Confusion chart of Naïve Bayes classifier with all data

Figure 4: Confusion chart of KNN classifier with standardized data

Figure 5: Confusion chart of Decision tree classifier with standardized data

Figure 6: Confusion chart of Naive Bayes classifier with standardized data

Figure 7: MRMR predictor ranking

Figure 8: Confusion chart of KNN classifier with MRMR 5 features

Figure 9: Confusion chart of Decision tree classifier with MRMR 5 features

Figure 10: Confusion chart of Naïve Bayes classifier with MRMR 5 features

Figure 11: PCA scree plot

Figure 12: Confusion chart of KNN classifier with PCA 9 features

Figure 13: Confusion chart of Decision tree classifier with PCA 9 features

Figure 14: Confusion chart of Naïve Bayes classifier with PCA 9 features

Figure 15: NCA predictor weights plot

Figure 16: Confusion chart of KNN classifier with NCA 4 features

Figure 17: Confusion chart of Decision tree classifier with NCA 4 features

Figure 18: Confusion chart of Naïve Bayes classifier with NCA 4 features

# Referencs

Davis, J. H. (1987). 'Calibration program for the 16-foot antenna’, 15.

DeepAI. 'What is the curse of dimensionality?'. Available at: <https://deepai.org/machine-learning-glossary-and-terms/curse-of-dimensionality>. Accessed: 24/12/.

DING, C. and PENG, H. (2011). 'MINIMUM REDUNDANCY FEATURE SELECTION FROM MICROARRAY GENE EXPRESSION DATA', *Journal of Bioinformatics and Computational Biology,* **3**(2), pp. 185-205.

Hasan, A.N., Eboule, P.S.P. and Twala, B. (May 2017). 'The use of machine learning techniques to classify power transmission line fault types and locations'.

I.T. Jolliffe (2002). 'Principal Component  
Analysis,  
Second Edition'*, Principal Component Analysis,* pp. 1-3, Springer.

J, C. R. K., D, V. K., D, B., B, M. A., R, J. and Majid, M. A. (2019). 'VLSI design and implementation of High-performance Binary-weighted convolutional artificial neural networks for embedded vision based Internet of Things (IoT)', *Procedia computer science,* **163**, pp. 639-647.

Jacob Goldberger, Sam Roweis, Geoff Hinton and Ruslan Salakhutdinov. (2005). 'Neighbourhood Components Analysis'. 2005.

Jason Brownlee. (2019). 'How to Choose a Feature Selection Method For Machine Learning'. Available at: <https://machinelearningmastery.com/feature-selection-with-real-and-categorical-data/>. Accessed: 24/12/.

LEI, Y. (2015). 'A Deep Learning-based Method for Machinery Health Monitoring with Big Data', *Ji xie gong cheng xue bao,* **51**(21), pp. 49.

P. P. Wasnik, Dr. N. J. Phadkule and Dr. K. D. Thakur (2020). 'Fault Detection and Classification in Transmission Line by using KNN  
and DT Technique', *International Research Journal of Engineering and Technology (IRJET),* **07**(04).

Rish, I. (2001). 'An Empirical Study of the Naïve Bayes Classifier', *IJCAI 2001 Work Empir Methods Artif Intell,* **3**.

Shahrtash, S.M. and Jamehbozorg, A. (Apr 2008). 'A decision tree based method for fault classification in transmission lines'.

Singh, A., Thakur, N. and Sharma, A. (Mar 2016). 'A review of supervised machine learning algorithms'.

Sternhagen, J. D., Mitzner, K., Berkenpas, E., Karlgaard, M., Wold, C. E. and Galipeau, D. W. (2002). 'A direct digital synthesis system for acoustic wave sensors', *IEEE sensors journal,* **2**(4), pp. 288-293.

The MathWorks, I. (a). 'Decision Trees'. Available at: <https://www.mathworks.com/help/releases/R2020b/stats/decision-trees.html>. Accessed: 24/12/.

The MathWorks, I. (b). 'Machine Learning in MATLAB'. Available at: <https://www.mathworks.com/help/releases/R2020b/stats/machine-learning-in-matlab.html>. Accessed: 24/12/.

Tutorials Point. 'KNN Algorithm - Finding Nearest Neighbors'. Available at: <https://www.tutorialspoint.com/machine_learning_with_python/machine_learning_with_python_knn_algorithm_finding_nearest_neighbors.htm>. Accessed: 24/12/.

Visalakshi, S. and Radha, V. (Dec 2014). 'A literature review of feature selection techniques and applications: Review of feature selection in data mining'.

stylefix