Process-Monitoring-for-Quality — A Model Selection Criterion for Shallow Neural Networks

Carlos A. Escobar¹, Ruben Morales-Menendez²

1 Global Research & Development, General Motors, Warren, MI, USA
carlos.1.escobar@gm.com

2 Tecnológico de Monterrey, Monterrey NL, México
rmm@tec.mx

ABSTRACT

Since most manufacturing systems generate only a few defects per million of opportunities, rare quality event detection is one of the main applications of the Process Monitoring for Quality philosophy. Single-hidden-layer feed-forward neural networks have been successfully applied to perform this task. However, since the best network structure is not known in advance, many models need to be learned and tested to select a final model with the right number of hidden neurons. A new three-dimensional model selection criterion (3D−NN) is introduced for the application of shallow neural networks to highly/ultra unbalanced binary data structures. Proposed criterion combines three of the most important attributes – prediction, fit, complexity – of a network structure and map them into a three dimensional space to select the best one. It is simple, intuitive and more stable than widely used criteria – Akaike information criterion, Bayesian information criterion and validation cross-entropy error – when dealing with these data structures.

1. INTRODUCTION

Process Monitoring for Quality (PMQ) is a big data-driven quality philosophy aimed at defect detection through binary classification (Abell et al., 2017). It is founded on Big Models (BM), a modeling paradigm based on optimization, machine learning and statistics, Fig. 1. It includes a learning component that requires many models to be created to find the final model (classifier) (Escobar, Abell, Hernández-de Menéndez, & Morales-Menéndez, 2018). Since many Candidates Models (CM) are created, Model Selection (MS) is one of the main challenges. The concept of using three attributes to evaluate the fitness of a CM was initially introduced in (Escobar, Weggner, Gaur, & Morales-Menéndez, 2019) for genetic programming, extended to the support vector machine (Escobar & Morales-Menéndez, 2019b) and logistic regression (Escobar & Morales-Menéndez, 2019a). However, these criteria cannot be directly applied to the Artificial Neural Network (ANN) algorithm.

Feed-Forward Neural Network(s) (FFNN) are universal approximators (Hornik, Stinchcombe, & White, 1989). Their widespread popularity in many domains is mainly due to their ability to approximate complex nonlinear patterns directly from the input samples, without assuming any parametric form for distinguishing between classes. In manufacturing, shallow ANN have been successfully applied for rare quality event detection (Escobar et al., 2018), as depicted in Fig 3.
In designing a functional ANN structure, determining the number of hidden neurons \((H_n)\) is one of the most important challenges because there is no analytical method to define the optimal structure in advance. In the context of generalization (prediction on unseen data), if the ANN is too large (too small), it could potentially overfit (underfit) the data. In either of these cases, the network would not generalize well. Existing methods (Sheela & Deepa, 2013) to determine \(H_n\) are heuristic in nature and/or trial-and-error dependent – training and estimating the generalization capacity. Since many ANN structures – different numbers of \(H_n\) – need to be learned and tested to find the best one (final model), MS is one of the most important steps in the development of a functional structure.

In today’s high conformance manufacturing environment, data sets for binary classification of quality (good,bad) tend to be highly/ultra (bad class count < 1%) unbalanced. Therefore, detecting these few Defects Per Million of Opportunities (DPMO) is one of the main challenges addressed by PMQ. Therefore, in manufacturing modeling, functional refers to a parsimonious (Burnham & Anderson, 2003) classifier with high detection ability.

A new Three Dimensional (3D) MS criterion \((3D−NN)\) for single-hidden-layer FFNN is presented. It is based on three of the most relevant attributes of an ANN structure: (1) prediction (generalization), (2) fit (robustness of predictions), and (3) complexity \((H_n)\). Criterion enables the development of a structure with high defect detection ability while avoiding overcomplexity; extra neurons with negligible contribution to the first two attributes.

The rest of the paper is organized as follows. Acronyms in Table 1, a brief theoretical background is in section 2. Section 3 describes the MS criterion. To evaluate the performance of the criterion, a comparative analysis using many public data sets is presented in section 4. Section 5 shows the conclusions and future research.

2. Artificial Neural Networks

Although ANN approach can be applied for different purposes: function approximation, probability estimation, pattern recognition, clustering, and prediction (Demuth, Beale, De Jess, & Hagan, 2014), the type of Neural Network (NN) for which the \(3D−NN\) MS is developed is very specific, and widely used in manufacturing. In this section, a brief overview of the ANN architecture of interest is provided.

2.1. Single Hidden Layer Feed-Forward

The FFNN was the first and simplest type of ANN developed. It has been successfully applied to solve a wide range of complex-classification problems across domains (Escobar et al., 2018; Boland & Murphy, 2001; Saxena & Saad, 2007). In this network, the information moves in just one direction, forward. There are no cycles or loops from the outputs of the neurons towards the inputs throughout the network (Sazli, 2006; Auer, Burgsteiner, & Maass, 2008). To explain in brief, the information enters the network through the input neurons of the first layer, which then develop a mathematical process (F. Amato nd A. López, na Méndez, Vaêhara, Hampl, & Havel, 2013) by using activation functions (Valente Klaine, Ali Imran, Onireti, & Demo Souza, 2017) and finally is transferred to the neurons of the following layer. Each neuron is connected to each neuron of the forward layer by a weighted relation, which indicates the strength of the link. Finally, the neurons of the last layer provide the outcome.

Figure 2 shows the ANN structure of interest, which is suitable for binary classification problems. A single hidden layer FFNN with sigmoid transfer and activation functions, the classification threshold, \(\gamma\), is tuned with respect to the Maximum Probability of Correct Decision (MPCD) following the OCTM algorithm (Escobar & Morales-Menendez, 2017). For the purposes of this paper, a network with only a single hidden layer is called shallow.

The process of assigning the predicted label \((\hat{y})\) to a manufactured item is defined as follows:

\[
\hat{y}_i = \begin{cases} 
1 & \text{if } O(y = 1|x; \theta) \geq \gamma \Rightarrow \text{ith predicted bad (+)} \\
0 & \text{if } O(y = 1|x; \theta) < \gamma \Rightarrow \text{ith predicted good (–)}.
\end{cases}
\]

Since the number of hidden layers is constant and the transfer/activation functions are the same, complexity (Kon & Plaskota, 2006) can be efficiently defined in function of number of parameters \((N_p)\).

\[
N_p = (m \times n) + (n \times 1)
\]

2.2. Maximum Probability of Correct Decision

In the field of machine learning, specifically applied to classification problems, a confusion matrix (Fawcett, 2006) is a technique for summarizing the performance of a classifier. It is a table with two rows and two columns that contrasts predictions with real-world values Table 2.

<table>
<thead>
<tr>
<th></th>
<th>Predicted good</th>
<th>Predicted bad</th>
</tr>
</thead>
<tbody>
<tr>
<td>Real-world good</td>
<td>True Negative (TN)</td>
<td>False Positive (FP)</td>
</tr>
<tr>
<td>Real-world bad</td>
<td>False Negative (FN)</td>
<td>True Positive (TP)</td>
</tr>
</tbody>
</table>

A type-I error \((\alpha)\) may be compared with a FP prediction; a type-II \((\beta)\) error may be compared with a false FN (Devore, 2015):

\[
\alpha = \frac{FP}{FP + TN}, \quad \beta = \frac{FN}{FN + TP}
\]
Table 1. Acronyms Definition

<table>
<thead>
<tr>
<th>Acronyms</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANN</td>
<td>Artificial Neural Network</td>
</tr>
<tr>
<td>AIC</td>
<td>Akaike Information Criterion</td>
</tr>
<tr>
<td>BIC</td>
<td>Bayesian Information Criterion</td>
</tr>
<tr>
<td>BM</td>
<td>Big Models</td>
</tr>
<tr>
<td>CEE</td>
<td>Cross-Entropy Error</td>
</tr>
<tr>
<td>CM</td>
<td>Candidate Model</td>
</tr>
<tr>
<td>DPMO</td>
<td>Defect Per Million of Opportunities</td>
</tr>
<tr>
<td>FFNN</td>
<td>Feed-Forward Neural Network(s)</td>
</tr>
<tr>
<td>LSW</td>
<td>Laser Spot Welding</td>
</tr>
<tr>
<td>MPCD</td>
<td>Maximum Probability of Correct Decision</td>
</tr>
<tr>
<td>MS</td>
<td>Model Selection</td>
</tr>
<tr>
<td>NN</td>
<td>Neural Network(s)</td>
</tr>
<tr>
<td>OCTM</td>
<td>Optimal Classification Threshold with respect to MPCD</td>
</tr>
<tr>
<td>3D</td>
<td>Three Dimension</td>
</tr>
<tr>
<td>UMW</td>
<td>Ultrasonic Metal Welding</td>
</tr>
</tbody>
</table>

Figure 2. Single-hidden-layer FFNN (fully connected) structure with sigmoid transfer function and sigmoid activation function.

The \(MPCD\) is a probabilistic-based measure of classification performance – driven by detection – that is highly sensitive to \(FN\) in highly/ultra unbalanced classes. The \(\alpha\), and beta \(\beta\) errors are combined to estimate its score:

\[
MPCD = (1 - \alpha)(1 - \beta)
\]  
(4)

where higher score \((0 \leq MPCD \leq 1)\) indicates better classification performance.

Figure 3 shows how \(BM\) is applied to process data to monitor and detect those very few \(DPMO\) that are generated by the manufacturing process. Where the predominant goal for a classifier is detection \((\beta = 0)\) with a small as possible false alarm rate \(-FP\ (\alpha)\).

2.3. Cross-Entropy Error

The probability distribution of the class label \(y\), given a feature vector \(x\) is modeled by (S. Lee, Lee, Abbeel, & Ng, 2006):

\[
O(y = 1|x; \theta) = \sigma(\theta^T x) = \frac{1}{1 + \exp(-\theta^T x)}
\]  
(5)

where \(\theta \in \mathbb{R}^N\) are the parameters of the model and \(\sigma(\cdot)\) is the sigmoid function that maps values from \((-\infty, \infty)\) to \([0, 1]\). The Cross-Entropy Error (CEE) of the NN is defined by (Murphy, 2012):

\[
\sum_{i=1}^{M} [y^{(i)} \log O^{(i)} + (1 - y^{(i)}) \log(1 - O^{(i)})]
\]  
(6)

3. The Criterion

A 3D MS criterion \((3D - NN)\) is presented which is aimed at analyzing highly/ultra unbalanced data structures. Due to the importance of detecting rare quality events generated by manufacturing systems, proposed criterion is mainly driven by detection ability. It uses three of the most important attributes of an ANN structure – prediction (rewarding attribute), fit (re-
warding attribute), complexity (penalizing attribute) – of each CM to map them into a three dimensional space to select the best one. Criterion avoids overcomplexity; extra neurons with negligible contribution to the rewarding attributes.

Once a set of Candidate Models, \(CM_j^m\), \(j = 1, \ldots, m\), – where \(m\) is the number of models – are trained by exploring different numbers of \(H_m\), the three attribute values are computed. To match the MPCD scale, fit and complexity attributes are rescaled to \([0, 1]\) using the \(\text{Min} – \text{Max}\) (Mohamad & Usman, 2013) normalization method.

1. Prediction (p)
   A rewarding attribute based on validation MPCD. In highly unbalanced data structures, this measure of classification performance tends to reflect the model’s capacity in detecting the minority class.
   \[
   p_j = 1 – \text{MPCD}_j \quad (7)
   \]

2. Fit (f)
   A rewarding attribute based on the validation CEE. It is a relative measure that describes how well a candidate model fits the validation data, where smaller values describe more robust predictions.
   \[
   f_j = \text{CEE}_j \quad (8)
   \]
   \[
   \text{MM}(f_j) = \frac{f_j – \min(f)}{\max(f) – \min(f)} \quad (9)
   \]

3. Complexity (c)
   A penalizing attribute based on \(N_p\).
   \[
   c_j = N_{pj} \quad (10)
   \]
   \[
   \text{MM}(c_j) = \frac{c_j – \min(c)}{\max(c) – \min(c)} \quad (11)
   \]

For each \(CM_j\) the tree associated attribute values are mapped into a three-dimensional space and the weighted Euclidean (\(E_{w_j}\)) distance (Deza & Deza, 2009) to the utopian point \((0, 0, 0)\) is computed, eqn 12. Then, the closest model \((3D – NN^*)\) to the utopian point is selected, Eq. 13. The utopian point, is an ideal model that optimizes the three attribute-functions simultaneously; however, most of the times, a model cannot be improved in any of the attributes without degrading at least one of the others. An overview of the MS process is illustrated in Fig. 4.

\[
E_{w_j} = \sqrt{w_p(p_j – 0)^2 + w_f(f_j – 0)^2 + w_c(c_j – 0)^2} \quad (12)
\]

where \(w_p = 1, w_f = 1\) and \(w_c = 0.01\)

\[
3D – NN^* = \min(E_{w_j})^m. \quad (13)
\]

3.1. Discussion

The fundamental principle of BM learning paradigm, is that none of the models developed using process (empirical) data is the true model that generates the observed data. Based on this premise, proposed criterion’s objective is not to search for the true model, but to efficiently solve the posed tradeoff between these three competing attributes.

There is no universal best Euclidean weight-combination, instead they are hyper-parameters that can be adjusted based on the goals of a particular project. Since complexity increases very rapidly, its influence is kept low, otherwise it would become a dominating attribute. Proposed weights maintain prediction \((w_p = 1)\) and fit \((w_f = 1)\) as the main drivers. However, the light penalization for complexity \((w_c = 0.01)\) prevents the selection of over-complex structures.

3.2. The MS Process

To illustrate the MS process, a case study is derived from Ultrasonic Metal Welding (UMW) (Shao et al., 2013) of bat-
tery tabs for the Chevrolet Volt (Abell et al., 2017), an extended range electric vehicle. A very stable process, that only generates a few defective welds per million of opportunities. The data set has 54 features and a binary outcome (good/bad). It is highly unbalanced, since it contains only 29 bad batteries out of 30,731 examples (0.0944%). Because manufacturing systems tend to be time-dependent, the data set is partitioned following the time-ordered hold-out validation scheme (Escobar et al., 2018): training set (18,495 - including 20 bafs), validation set (12,236 - 9 fabs). Following a widely used rule-of-thumb “the number of hidden neurons should be less than twice the size of the input layer” (Heaton, 2008), 108 candidate models, $CM_j$, are generated using early stopping (Demuth et al., 2014) with the scaled conjugate gradient algorithm (Møller, 1993).

Candidate model information is summarized in Fig. 5, and the MS process is illustrated in Fig. 6. First, the three attributes of each $CM_j$ are mapped into a three dimensional space, Fig. 6(a), then, the weighted Euclidean distance of each $CM_j$ is computed, and the one closest to the utopian point is selected, Fig. 6(b). According to the MS criterion, $CM_7$ should be selected ($E_w = 0.0258$). This model has the highest prediction, relatively low CEE (good fit) and low complexity.

4. Comparative Analysis

To verify the performance of proposed criterion, the final model selected by $3D - NN$ is compared against the final model selected by each of the three criteria/method, which have been widely used to search for the most efficient NN structure: (1) Akaike information criterion (AIC) (Panchal, Ganatra, Kosta, & Panchal, 2010), (2) Bayesian information criterion (BIC) (H. K. Lee, 2001), (3) validation CEE (Demuth et al., 2014), and (4) validation $MPCD$ (Escobar & Morales-Menendez, 2017). To perform this analysis, 10 highly/ultra unbalanced data sets (8 publicly available) are used, Table 3. The associated attribute values of each final model by data set are summarized in Table 4.

$3D - NN$ vs:

- **AIC**, the final models selected by the $3D - NN$ criterion show significantly better generalization ability and fitting properties. This gain is obtained at the expense of a slightly higher complexity than the structures selected by $AIC$.

- **BIC**, the $3D - NN$ criterion show significantly better generalization ability and fitting properties. This gain is obtained at the expense of a slightly higher complexity.

- Validation CEE and $MPCD$, the $3D - NN$ criterion show competitive generalization ability. However, both approaches show competitive generalization at the expense of overcomplexity.

The proposed criterion shows competitive performance at solving the posed tradeoff between the three competing attributes. The $AIC$, $BIC$ and validation CEE raise a red flag when dealing with highly/ultra unbalanced data structures, as they selected myope structures – a solution that fails to capture the pattern (e.g., $MPCD = 0.3078$, $MPCD = 0.3338$, $MPCD = 0.4000$ respectively).

Whereas measures of classification performance (e.g., $MPCD$) can be used as a $MS$ criterion, there is a risk associated – as shown in by data set 8 – since the final model may be an overcomplex $NN$ structure with virtually the same prediction ability of a simpler one – e.g., $MPCD = 0.8621$ with $N_p = 30$ Vs $MPCD = 0.8678$ with $N_p = 1650$.

5. Conclusions

A new model selection criterion for a single-hidden-layer FFNN structures was developed. It maps three of the most important attributes of an ANN structure – prediction, fit, complexity – into a three dimensional space and uses the weighted Euclidean distance to the utopian point – where the three attributes are optimized simultaneously.

Based on empirical results, proposed criterion shows better performance and stability at solving the posed tradeoff between these three competing attributes than conventional model selection criteria when dealing with highly/ultra unbalanced data structures. As it selected structures with high detection ability, avoided overcomplexity and never selected a myope solution.
Table 3. Data sets information (preprocessing details are provided in Appendix A).

<table>
<thead>
<tr>
<th>Data set</th>
<th>Description</th>
<th>Features</th>
<th>Instances (T/V)</th>
<th>Negative class (T/V)</th>
<th>Overall %</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>UMW</td>
<td>54</td>
<td>18,495/12,236</td>
<td>20/9</td>
<td>0.09††</td>
</tr>
<tr>
<td>2</td>
<td>LSW</td>
<td>240</td>
<td>1,502/760</td>
<td>32/15</td>
<td>2.07†</td>
</tr>
<tr>
<td>3</td>
<td>AID373*</td>
<td>155</td>
<td>47,831/11,957</td>
<td>50/12</td>
<td>0.11††</td>
</tr>
<tr>
<td>4</td>
<td>AID604AID644*</td>
<td>154</td>
<td>47,826/11,956</td>
<td>54/13</td>
<td>0.11††</td>
</tr>
<tr>
<td>5</td>
<td>AID746AID1284*</td>
<td>154</td>
<td>47,828/11,956</td>
<td>46/11</td>
<td>0.09††</td>
</tr>
<tr>
<td>6</td>
<td>Statlog (class 1)</td>
<td>36</td>
<td>4,435/2,000</td>
<td>1,072/461</td>
<td>23.82†</td>
</tr>
<tr>
<td>7</td>
<td>Statlog (class 2)</td>
<td>36</td>
<td>4,435/2,000</td>
<td>479/224</td>
<td>10.92†</td>
</tr>
<tr>
<td>8</td>
<td>Credit Card Fraud</td>
<td>29</td>
<td>200,000/84,807</td>
<td>385/107</td>
<td>0.17††</td>
</tr>
<tr>
<td>9</td>
<td>Occupancy Detection</td>
<td>5</td>
<td>6,587/1,791</td>
<td>173/98</td>
<td>3.23†</td>
</tr>
<tr>
<td>10</td>
<td>HTRU2</td>
<td>8</td>
<td>12,000/5,898</td>
<td>1,484/155</td>
<td>9.15†</td>
</tr>
</tbody>
</table>

T=Training set V=Validation set *Subsets of PubChem Bioassay Data †highly unbalanced ††ultra unbalanced

Table 4. Characteristics of the selected model based on $3D-NN$, AIC, BIC and validation CEE

<table>
<thead>
<tr>
<th>Data Set</th>
<th>3D-NN</th>
<th>AIC</th>
<th>BIC</th>
<th>CEE</th>
<th>MPCD</th>
</tr>
</thead>
<tbody>
<tr>
<td>p</td>
<td>f</td>
<td>c</td>
<td>p</td>
<td>f</td>
<td>c</td>
</tr>
<tr>
<td>1</td>
<td>0.9767</td>
<td>64.45</td>
<td>385</td>
<td>0.7659</td>
<td>65.49</td>
</tr>
<tr>
<td>2</td>
<td>0.8782</td>
<td>46.05</td>
<td>10485</td>
<td>0.6559</td>
<td>60.41</td>
</tr>
<tr>
<td>3</td>
<td>0.7667</td>
<td>223.29</td>
<td>27144</td>
<td>0.4784</td>
<td>247.42</td>
</tr>
<tr>
<td>4</td>
<td>0.7835</td>
<td>336.86</td>
<td>4650</td>
<td>0.3078</td>
<td>420.43</td>
</tr>
<tr>
<td>5</td>
<td>0.8890</td>
<td>213.94</td>
<td>4650</td>
<td>0.6283</td>
<td>329.31</td>
</tr>
<tr>
<td>6</td>
<td>0.9863</td>
<td>48.67</td>
<td>518</td>
<td>0.9694</td>
<td>167.41</td>
</tr>
<tr>
<td>7</td>
<td>0.9755</td>
<td>56.57</td>
<td>185</td>
<td>0.9605</td>
<td>132.13</td>
</tr>
<tr>
<td>8</td>
<td>0.8621</td>
<td>662.8</td>
<td>30</td>
<td>0.8621</td>
<td>662.8</td>
</tr>
<tr>
<td>9</td>
<td>0.9582</td>
<td>185.03</td>
<td>12</td>
<td>0.9675</td>
<td>190.46</td>
</tr>
<tr>
<td>10</td>
<td>0.8891</td>
<td>1118.1</td>
<td>18</td>
<td>0.8891</td>
<td>1118.1</td>
</tr>
</tbody>
</table>

Future research along this path could be the application of the $3D-NN$ criterion to deep neural network structures, since the $N_p$ tend to explode, defining its influence on the criterion seems an interesting research challenge.

ACKNOWLEDGMENT

Authors deeply appreciate the feedback provided by Dr. Michael A. Wincek. A special gratitude to Dr. Jeffrey Abell, whose ideas and contributions illuminated this research. And to Dr. Jorge Arinez for supporting the deployment of PMQ across General Motors manufacturing plants. This work was partially supported by CONACYT and Tecnológico de Monterrey.

REFERENCES


Figure 5. Candidate model information.

(a) Prediction attribute.

(b) Fit attribute.

(c) Complexity attribute.

Figure 6. MS process based on the weighted Euclidean distance.


**Biographies**

Carlos A. Escobar Díaz earned an Industrial Engineering degree with concentration in automated manufacturing from the Instituto Tecnológico de Ciudang Juarez in 2003, a masters degree in engineering with Concentration in Quality and Productivity Systems from the Instituto Tecnológico y de Estudios Superiores de Monterrey, Campus Juarez in 2005, and a master of science in Industrial Engineering from New Mexico State University, University of Las Cruces, New Mexico in 2015. He was certified six-sigma black belt from Arizona State University in 2008 and design for six-sigma black belt from University of Michigan in 2012. In 2013 he was inducted into the Alpha Pi Mu Industrial Engineering honor society and into Tau Beta Pi engineering honor society in 2014. In 2017, Carlos was ranked into the top 3% in TEXATA, the Big Data Analytics World Championships. He received a PhD in Engineering Sciences program from the Instituto Tecnológico y de Estudios Superiores de Monterrey, Campus Monterrey in May 2019.

Ruben Morales-Menendez received a Ph.D. degree in Artificial Intelligence. He was a Visiting Scholar with the Lab. of Computational Intelligence, University of British Columbia, Canada. He is currently a consultant specializing in the analysis and design of automatic control systems for continuous processes. He is the Dean of Graduate Studies with the School of Engineering and Sciences at Tecnológico de Monterrey. He is a member of the National System of Researchers (leveI), the Mexican Academic of Sciences and the Mexican Academic of Engineering.

**APPENDIX**

- **Data set 1-2**
  UMW and LSW data sets are privately stored in the Manufacturing Research Lab of General Motors. Because of the active research aimed at understanding and improving these processes, General Motors can not make these data sets publicly available.

- **Data set 3-5**
  Data sets derived from the PubChem Bioassay data set (Dheeru & Karra Taniskidou, 2017). These highly...
imbalanced bioassay data sets are from the differing types of screening that can be performed using HTS technology. These data sets were created from 12 bioassays.

- **Data set 6-7**
  Data sets derived from the Statlog (Landsat Satellite) data set (Dheeru & Karra Taniskidou, 2017). The original data set contains seven classes (with no instances with class 6). In this study, only classes 1 and 2 are considered – class 1 vs. all, class 2 vs. all.

- **Data set 8**
  Credit Card Fraud data set (https://github.com/ellisvalentiner/credit-card-fraud), it contains credit card transactions over a two day collection period in September 2013 by European cardholders. There are a total of 284,807 transactions, of which 492 (0.172%) are fraudulent. First 200,000 are used for training and the last 84,807 for validation.

- **Data set 9**
  Occupancy Detection data set (Candanedo & Feldheim, 2016; Dheeru & Karra Taniskidou, 2017). To create an unbalanced data structure, one out of 10 class 1 are included in the data sets (index 1, 10, 20, etc.) and the remaining nine eliminated, all 0 class are included.

- **Data set 10**
  HTRU2 data set (Lyon, Stappers, Cooper, Brooke, & Knowles, 2016; Dheeru & Karra Taniskidou, 2017). Pulsar candidates collected during the HTRU survey. Pulsars are a type of star, of considerable scientific interest. Candidates must be classified into pulsar and non-pulsar classes to aid discovery. First 12,000 are used for training and the last 5,898 for validation.