* **Decision Tree and Algorithms**

The routine for selecting the appropriate machine learning method will initially be composed of a Decision Tree, which will identify the best machine learning approach based on features of a given dataset. The short-term goal is to classify dataset into three broad categories: Supervised Learning, Unsupervised Learning and Semi-supervised Learning according to whether there is output data in the dataset. The team decided to provide 18 potential machine learning algorithms:

• Multiclass Neural Network

• Linear Regression

• Random Forest Regression

• Sum Regression

• Logistic Regression

• Multi Layer Perceptron Neural Network

• Naive Bayesian Network

• Support Vector Machine

• Multiclass Logistic Regression

• Self-Organizing Map

• K-means Clustering

• Principle Component Analysis • Forced Clustering

• Self-Training

• Deep Learning

• Recurrent Neural Network

• Time Delay Neural Network

•Feature Selection Principal Component Analysis

The most suitable tool will be suggested from this list of machine learning methods by using decision algorithm outlined below.

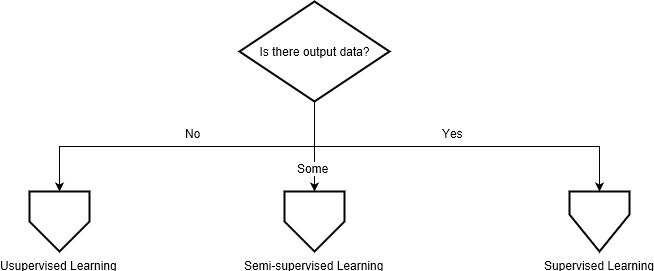
First of all, the algorithm will determine whether the dataset is supervised; according to whether there is output data in the dataset, the dataset will be divided into three categories: Unsupervised Learning, Semi-supervised Learning and Supervised Learning, as can be seen in ﬁgure 1.

Figure 1: Deciding the class of machine learning to use

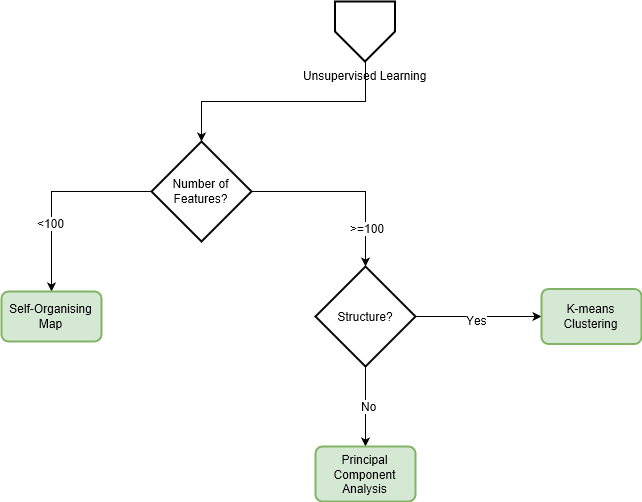
If the dataset has no output data, it is classed as an Unsupervised Learning problem. Within Unsupervised Learning, if there are lots of features in the dataset, a Self-Organizing Map is optimal. Otherwise, if the datasets are structured, suggest K-means Clustering to analyze. If they are not, Principal Component Analysis will be suggested. This decision is modelled in ﬁgure 2.

Figure 2: Deciding which unsupervised learning technique to suggest

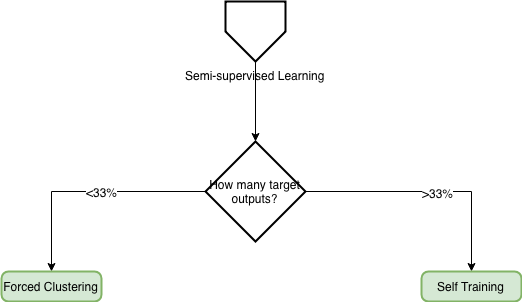
Semi-Supervised Learning is a class of machine learning tasks and techniques that also makes use of unlabeled data for training – typically a small amount of labeled data with a large amount of unlabeled data. Self-Training will be suggested if only 33% or fewer of the entries in the dataset are labelled. Otherwise Forced Clustering is suggested. The semi-supervised decision is shown in ﬁgure 3.

Figure 3: Deciding which semi-supervised learning technique to suggest

If the dataset is completely labelled with output data, divide it into Supervised Learning. If the outputs are categories, the data represents a Classiﬁcation problem. Otherwise, if the outputs are continuous values, the dataset is best modelled by Regression.

For classiﬁcation, the optimal machine learning method depends on the number of categories present, the simplicity and size of the dataset, and the complexity of relations. Machine learning algorithms for regression depend on the size and simplicity of the dataset.

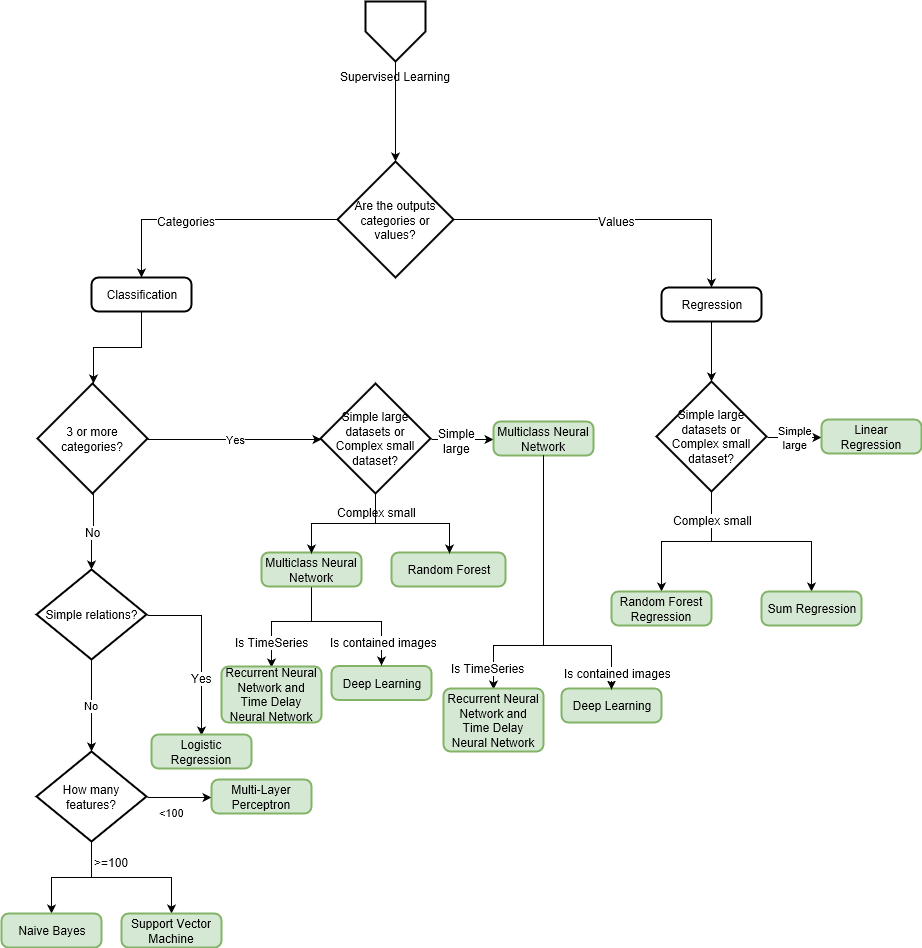
Figure 4 visualizes this set of decisions.

Figure 4: Deciding which supervised learning technique to suggest

* **IMPLEMENTATION OF DECISION TREE:**

The main concept to implement the decision tree is using the expert system. In the code level, the implementation is in a JavaScript file named Analyze.

In the code level, the team use plenty “if…then”, “else…then” statements according to the expert system to implement the decision tree. For instance: the codes below is the implementation of the Unsupervised Learning.

// Unsupervised Learning

if (dataset.labelsRatio == 0) {

console.log("Unsupervised Learning");

methodType = "Unsupervised Learning";

if (dataset.numFeatures < 100) {

console.log("Self-Organising Map");

bestMethod = "Self-Organising Map";

} else {

if (dataset.structure < 0) {

console.log("Principle Component Analysis");

bestMethod = "Principle Component Analysis";

} else {

console.log("K-means Clustering");

bestMethod = "K-means Clustering";

}

}

}

The central idea is to put all dataset’ s information which get from the database(by using dataset. Patterns) into the decision tree and store the output(machine learning tools) into a variable BestMethod.

* **Understanding Machine Learning Algorithms**

The team felt it would be beneﬁcial to have a basic understanding of each algorithm, their respective principles, and how and when to use each one.

**Self-Organising Map**

A Self-Organising Map (SOM) converts complex relationships within data of high dimensions into simple relationships in lower dimensions, all while preserving the most important relationships of the starting data [5].

**K-Means Clustering**

K-means clustering classiﬁes a set of data into k clusters (where k > 0) by assigning it to a centroid (central location of a cluster). This is achieved by assigning it to the nearest centroid. When each data point has been assigned, k centroids are recalculated and the datapoints are reassigned. This process of deﬁning and assigning to new centroids is repeated until no data points move [6].

**Principal Component Analysis**

Principal Component Analysis (PCA) ﬁnds a small subset of combinations of correlated parameters which have the most inﬂuence on variation within the dataset [7].

**Linear Regression**

Linear Regression is a form of predictive analysis where the values of (an) independent variable(s) are used to predict future values of a dependent variable by ﬁtting a line through the values the variables [8].

**Multi Layer Perceptron Neural Network**

Multi Layer Perceptron Neural Networks are networks of processing nodes (perceptrons), each connected via a weight. The weighted sum of all input values into a node is ran through an activation function. If the function outputs a number higher than a threshold, its weighted sum is propagated into the input of all nodes connected to its output [9].

**Multiclass Neural Network**

A Multiclass Neural Network extends on a neural network by providing one output neuron for each possible output class of a given problem [10].

**Random Forest Regression**

Random Forests select a random subset of features, observe their outputs, and generate a decision tree based on the features selected. This process is repeates lots of times (hundreds to thousands), and each tree is averaged into the ﬁnal decision tree. [11]

**Logistic Regression**

Logistic Regression is a method for binary prediction when the outputs are dichotomous (true or false). It is used to ﬁnd the relationship of one dichotomous variable, and one or more other independent variable. [12]

**Multiclass Logistic Regression**

A generalisation of Logistic Regression, which can identify more than two possible outcomes [13].

**Na¨ıve Bayesian Network**

A Na¨ıve Bayesian Network calculates the conditional probability of each attribute of a class given its label. By applying Bayes rule, labels are classiﬁed based on the instances of each attribute. Its na¨ıvety stems from the fact that it assumes that all attributes are conditionally independent on the class [14].

**Support Vector Machine**

Support Vector Machines take an input vector and map it, non-linearly, into a higher dimension feature space, in which a linear, highly generalisable, decision surface is created [15].

**Self-Training**

Self training trains a model with labeled data, which is used to classify unlabeled data. The unlabeled data that the model classiﬁed with the highest conﬁdence are added back to the set of labeled data [16].

**Semi Supervised K-means**

An extension to K-means, where the initial centroids are chosen based on data that is already classiﬁed [17]

**Deep Learning**

Enter an object, perform multi-dimensional operations layer by layer, and finally output the result. Deep learning is actually a relative expert system. The previous judgment is a top-down idea defined by a large number of "If - Then" rules. Deep learning is a model of artificial neural network bottom-up thinking, trying to imitate the patterns of communication between neurons in the brain and processing information[19].

**Recurrent Neural Network**

The idea behind circular neural networks is to use sequence information. Cyclic neural networks are called loops because they perform the same tasks for each element in the sequence, and the output depends on the previous calculations[20].

**Time Delay Neural Network**

The time delay neural network (TDNN) is the first model to use multiple CNN layers for ASR. This model applies convolution operations on both the time and frequency axes. Time-Delay allows it to learn the timing dependencies between features[21].

**Feature Selection Principal Component Analysis**

PCA is used to find out which feature is the best to describe the variance in a dataset. It is mostly used to reduce the dimensionality of a dataset which contained a great amount of data. Therefore applying machine learning becomes more practical when the original data has an inherently high dimension[22].

[5] T. Kohonen, The self-organizing map (som), Mar. 2005. [Online]. Available: http://www.cis.hut.fi/projects/somtoolbox/theory/somalgorithm. shtml.

[6] Clustering - k-means. [Online]. Available: https://home.deib.polimi. it/matteucc/Clustering/tutorial\_html/kmeans.html.

[7] M. Einasto, L. J. Liivam¨agi, E. Saar, J. Einasto, E. Tempel, E. Tago, and V. J. Mart´ınez, “Sdss dr7 superclusters,” Astronomy & Astrophysics, vol. 535, A36, 2011. doi: 10.1051/0004-6361/201117529.

[8] A. C. Davison, Statistical models. Cambridge University Press, 2003, pp. 18– 21.

[9] L. Hardesty, Explained: Neural networks, Apr. 2017. [Online]. Available: <http://news.mit.edu/2017/explained-neural-networks-deeplearning-0414>.

[10] Multi-class neural networks: One vs. all — machine learning crash course — google developers, Oct. 2018. [Online]. Available: https://developers. google.com/machine-learning/crash-course/multi-class-neuralnetworks/one-vs-all.

[11] T. Hastie, R. Tibshirani, and J. H. Friedman, The elements of statistical learning: data mining, inference, and prediction. Springer, 2017.

[12] What is logistic regression? [Online]. Available: https://www.statisticssolutions. com/what-is-logistic-regression/.

[13] W. H. Greene, Econometric analysis. Pearson, 2012, pp. 803–806.

[14] N. Friedman, D. Geiger, and M. Goldszmidt, “Bayesian network classiﬁers,” Machine learning, vol. 29, no. 2-3, pp. 131–163, 1997.

[15] C. Cortes and V. Vapnik, “Support-vector networks,” Machine learning, vol. 20, no. 3, pp. 273–297, 1995.

[16] N. Fazakis, S. Karlos, S. Kotsiantis, and K. Sgarbas, “Self-trained lmt for semisupervised learning,” Computational Intelligence and Neuroscience, vol. 2016, p. 2, 2016. doi: 10.1155/2016/3057481.

[17] J. Yoder and C. E. Priebe, “Semi-supervised k-means++,” arXiv preprint arXiv:1602.00360, 2016.

[18] C. Roadknight, D. Suryanarayanan, U. Aickelin, J. Scholeﬁeld, and L. Durrant, “An ensemble of machine learning and anti-learning methods for predicting tumour patient survival rates,” in 2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA), Oct. 2015, pp. 1–8. doi: 10.1109/DSAA.2015.7344863.

[19] Sunzhijun, Xuelei, Xuyangming, Wangzheng. Review of deep learning research[J]. Computer Applied Research, 2012, 29(8): 2806-2810. 《Artificial Intelligence and Robotics Research》, Vol.5 No.4, 2016-11-25

[20] Denny Britz, Recurrent Neural Networks Tutorial, Part 1 – Introduction to RNNs. September 17, 2015. Available:

http://www.wildml.com/2015/09/recurrent-neural-networks-tutorial-part-1-introduction-to-rnns

[21] Alexander Waibel, etc. Phoneme Recognition Using Time-Delay Neural Network, VOL. 37. NO. 3.MARCH 1989. Available:

https://note.youdao.com/ynoteshare1/index.html?id=5ef2bf22a4398d2a85d9ac61275d3d9b&type=note

[22]

Roger Rowland, Apr 27 '13, Difference between PCA (Principal Component Analysis) and Feature Selection. Available:

https://stackoverflow.com/questions/16249625/difference-between-pca-principal-component-analysis-and-feature-selection