Classification Of Dementia In Adults

João Nunes[1], Luís Braga[2], Luís Martins[3], Luís Filipe[4], Cristiana Neto[5], and José Machado[6]

[1]*Departamento de Informática, Universidade do Minho. Email: a82300@alunos.uminho.pt*
[2]*Departamento de Informática, Universidade do Minho. Email: a82088@alunos.uminho.pt*
[3]*Departamento de Informática, Universidade do Minho. Email: a82298@alunos.uminho.pt*
[4]*Departamento de Informática, Universidade do Minho. Email: a83099@alunos.uminho.pt*
[5]*Centro de Investigação Algoritmi, Departamento de Informática, Universidade do Minho. Email: cristiana.neto@algoritmi.uminho.pt*
[6]*Centro de Investigação Algoritmi, Departamento de Informática, Universidade do Minho. Email: jmac@di.uminho.pt*

Abstract

Dementia is a broad term for a large number of diseases and conditions, it's often associated with Alzheimer's. A reliable diagnostic, especially in the early stages, may prove to be valuable so much so, that it may even prevent further complications if it can be correctly evaluated early on. As such, *Machine Learning algorithms* can be applied in order to validate and correctly classify cases of dementia or non dementia in adults. Using a dataset containing magnetic resonance imaging comparisons of demented/non demented adults, it was possible to apply machine learning models, more specifically Support Vector Machines, Decision Trees, Logistic Regression, Neural Networks, Naive Bayes and Random Forest in order to classify instances of dementia. The best classifier proved to be either the *Logistic Regression* or the *SVM* classifier with an overall *accuracy* of 94.49%.

Keywords: Dementia, Machine Learning, Support Vector Machine, Decision Tree, Logistic Regression, Neural Network, Alzheimer.

1 Introduction

Dementia, as stated in the *Abstract*, is a broad term for a large number of conditions and diseases, and it often occurs older adults. According to studies done by the *World Health Organization (WHO)* [6], around 50 million people worldwide suffer from dementia, other studies also show that the number is only going to grow to about 63 million people in the year 2030 [1]. Dementia can put a strain on a family due to the hardships that often come with it, due to it's degenerative nature both physical and mental on the patient, as well as the financial hardship that comes with it. As such, an early intervention and classification can often save and delay it's manifestation. The data that is going to be used, is related to *MRI* scans, in both demented and non demented patients.

Accurate classifications of dementia are crucial, as it's a time sensitive operation [5] both for the medical staff and the family of the patient. In clinical diagnosis of dementia, it often requires multiple pieces of information such as a neuropsychological test score, laboratory studies and so on. And while, it may not be preventable, an early diagnostic can prepare the family for what's to come.

The aims of the paper are to (1) develop, test and benchmark *machine learning algorithms* in order to (2) find the best *algorithm* that correctly classifies dementia in patients, all while following the *CRISP DM* pipeline. *Rapidminer* was the program that was chosen to visualize, prepare, and apply a machine learning model on the data.

2 Related Work

Other works have been done, and published, on this field, where in one case [2] *Deep Learning* was implemented in order to correctly diagnose the *Alzheimer's Disease*, as well as the *SVM* classifier, which was also used in our case. Data from the *J-ADNI* [7] database was extracted and used in order to construct the aforementioned study.

Another study [3] also uses data extracted from *MRI* scans and a *Neural Network* classifier is then trained and consequently tested on this data, in order to detect the stage of the *Alzheimer's Disease*. The approach of that study is to detect the disease as early as possible with the extracted features from the hippocampus region of *MRI* Scans, because this region is the first region that will be affected if the patient has the *Alzheimer's Disease*.

An additional paper [4], also used *classifiers* similar to the ones that were used on this study, such as *Naive Bayes* and *Decision Trees*, which showed good albeit not ideal results on that dataset. The evaluation of the classifiers on that paper was also made using a cross validation with 10 folds, which was also one of the scenarios that was used in order to evaluate the classifiers on this paper.

3 CRISP DM

The following study will follow the standard *CRISP DM* or *Machine Learning* pipeline. Where, after understanding the need to develop an accurate classifier to predict dementia in adults, which is the *Business Understanding* stage, that *phase* is followed by the *Data Understanding* and *Data Preparation* stages, which are going to be to talked about in the next section, where it's important to understand the attributes of the dataset and their relation with the *target*, and after that the data is prepared for the classifiers. Modeling and the evaluation phases are the ones that follow, where it's important to carefully think about the classifiers that make sense, given the problem in question, and applying the said classifiers to the prepared data. The last step in the *pipeline* corresponds to the *deployment* phase, where the model that was constructed, given that it's good enough, may be applied in a real world situation.

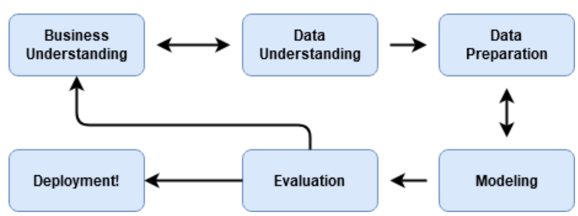


Figure 1. Crisp DM process diagram.

4 Data Visualization

Visualizing the data is one of the most important steps in the pipeline, as there often are hidden inferences between attributes in the dataset. It's also important in the sense that it can provide a better understanding of the attributes and their distributions, which can lead to removing useless attributes, and so on. The dataset that was used, contains 373 instances and 15 attributes, the classification target is the attribute Group. It also contains missing data, and *outliers* in most attributes.

Table 1. Attributes in the dataset.

Name	Description	Type
Subject ID	Identification of the patient	Polynomial
MRI ID	MRI Exam identification	Polynomial
Group	Mental state of the patient (Demented, Non demented, converted)	Polynomial
Visit	Visit order	Integer
MR Delay	MRI Delay time	Integer
M/F	Gender of the patient	Polynomial
Hand	Dominant hand of the patient	Polynomial
Age	Age in years	Integer
EDUC	Years of education	Integer
SES	Social Economic Status (0 is the lowest and 5 the highest status)	Integer
MMSE	Mental state examination (30=best and 0=worst)	Integer
CDR	Clincial dementia rating (0=no dementia and 2=moderate AD)	Real
eTIV	Estimated intracranial volume	Integer
nWBV	Normalized whole brain volume	Real
ASF	Atlas scaling factor	Real

The age of the patients that were used in order to construct the aforementioned dataset vary between 60 and 98 years old, and it also follows the following distribution.

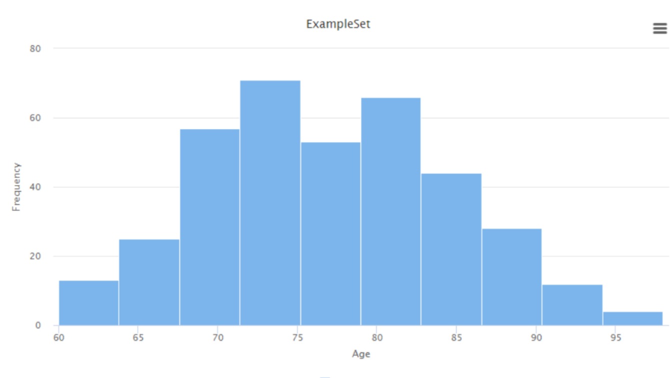
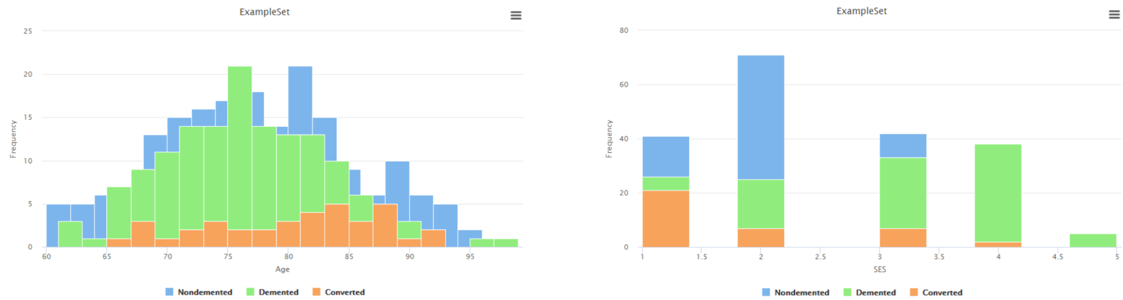


Figure 2. Age distribution.

However, it's more relevant to analyze the relation that the target attribute has with the rest of the attributes. For this reason, and since dementia is often associated with age, an histogram was made comparing the target attribute with the age of the adults in the dataset, also, another histogram was made to analyze the relation between the social economic status of the person in regards to the patient being demented or not.



(a) Histogram Group/Age

(b) Histogram Group/SES

Figure 3. Histograms comparing the target attribute with others attributes in the dataset.

Thus, it's possible to check that the large portion of the demented patients ages vary from 70 to 85 years old, and that the oldest demented patient is 98 years old. In regards to the social economic status *SES*, knowing that a *SES* of 5 is related with worse economic conditions, it might be possible to say that dementia often is associated with poor social economic backgrounds.

It was also possible to check that the dataset contained some redundancy, for example the attribute *Hand* only had one type of value which was *R*, which means that every single patient that was analyzed had the right hand as the dominant hand, therefore the attribute is not needed. This situation will be dealt with accordingly in the next section.

5 Data Preparation

In this stage of the *CRISP DM pipeline* it was necessary, before applying any *ML* models, to first prepare the data in order to remove any inconsistencies. Firstly, and since the *target attribute Group* contained the value *Converted*, which indicated that the first diagnostic of the adult showed that he didn't have dementia, and that in a later diagnostic signs of dementia were discovered. Thus, this type of value was converted simply to *Dementia* which actually balanced the *target* attribute, this attribute was converted at a later stage to *binominal*, where *true* = *Demented* and *false* = *Non demented*.

Figure 4. Balanced distribution in the target attribute group.

After balancing the target attribute, and since the dataset contained missing values, this missing attributes were dealt with by replacing them with the average of the attribute in each instance that contained missing values. After that, in the first scenario that was prepared, the correlated attributes that had a correlation superior than 80% were also removed, because higher correlation means that they are similar in nature therefore keeping this attributes in the dataset usually leads to redundancy. Outliers can also be seen as *data anomalies*, therefore they are not useful and quite harmful (in large quantities) in the prediction made by the model, however since the dataset is not big, it only contains 373 instances, only 10 outliers were specified to be found and consequently removed from it, which reduced the size to only 363 instances.

The last step in the data preparation phase was to normalize the numeric attributes to values between 0 and 1, in order to have a common scale between all attributes.

All these options that were taken led to a *dataset* with 363 instances, and 8 attributes.

Table 2. Attributes that remained in the dataset and their data types, with the aforementioned data preparation.

Name	Type
Group	Binominal
Age	Real
EDUC	Real
SES	Real
MMSE	Real
CDR	Real
eTIV	Real
nWBV	Real

Another scenario was prepared, where instead of using an operator in *Rapidminer* to remove the attributes that had an higher correlation than 80%, the attribute selection was made manually, where the only attributes that were left out were the ones that didn't make sense, such as the id's, the attribute *hand* that had the same value in all instances of the dataset, and also the attribute *Visit*. All the other steps that were taken in the data preparation of this new scenario, were the same as the ones that were described before.

Table 3. Attributes that remained in the dataset and their data types, with a manual feature selection.

Name	Type
Group	Binominal
MR Delay	Real
Age	Real
EDUC	Real
SES	Real
MMSE	Real
CDR	Real
eTIV	Real
nWBV	Real
ASF	Real

6 Modelling

In this phase of the *pipeline*, it's important to consider what *Machine Learning* models to use, given the prepared data. Due to the nature of the problem, which is classification, 6 models were prepared. The first model to be used was *Decision Tree*, which uses a tree like graph to model decisions and their possible outcomes, where in each node a test is made on the attribute, and through the outcome of this test a new path is made. A variant of the decision tree algorithm that was also used was the *Random Forest* where a large number of individual decision trees are used in order do predict the outcome, or the classification, on the test data, the predicted value can be found by using the most common outcome predicted by all the trees or by using the outcome that has the highest accumulated confidence.

Logistic regression is a simple approach to take in this situation. It's a technique often used in binary classification problems, which is this case, where the *Group attribute* value is either true or false. The input attributes are combined linearly in order to predict the output value, which is true/false. Support vector machines, also uses the same principle as the regression algorithms, however it's more indicated with classification problems in the sense that the optimal outcome is to find an hyperplane that distinctly classifies the *target* attribute.

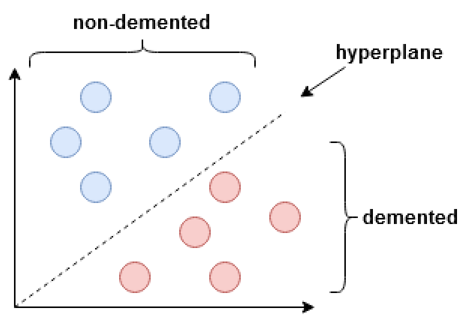


Figure 5. SVM feature distinction.

Another algorithm to be used, that is also suitable in classification problems, is the *Naive Bayes classifier*. This algorithm is based on the *Bayes theorem* which is a probabilistic theorem, which is used to find out the probability of something happening (A) by knowing that that (B) has occurred.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Therefore, and given the problem at hand, the A variable can be seen as the target feature, the attribute *Group* which is the attribute that needs to be classified. The B variable can be extended and seen as the rest of the features of the dataset.

$$B = (Age, Educ, SES, ...)$$

$$A = Group$$

$$P(A|B) = \frac{P(Age|A)P(Educ|A)P(SES|A)(...)P(A)}{P(Age)P(Educ)P(SES)(...)}$$

The last *ML* model that was considered to be applied to the data was the *Neural Network* algorithm. Neural Networks as the name might suggest were modeled loosely after the human brain, in the sense that it tries the replicate the connection and inner workings of the neurons, creating an artificial neuron capable of processing data. Each neuron in the network combines the input data with a weight that has been pre-determined, and in the end the final result is combined with a bias value. This final value is then processed through an activation function, and that value is then carried on to the next set of neurons in the next layer.

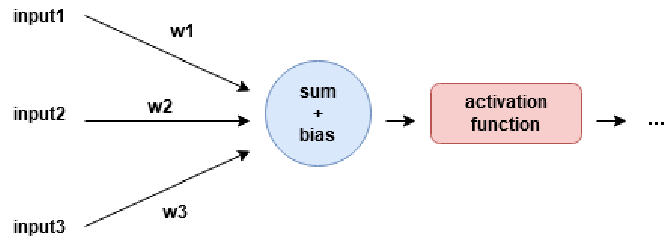


Figure 6. Neural Network Diagram.

7 Evaluation

After carefully thinking about the problem and the various Machine Learning algorithms that could be used to tackle it, the stage was finally set to finally apply the algorithms mentioned in the last section to the prepared data. Each algorithm was tested two times, the first time the algorithm was executed on the data a simple *train/test split* was used to split the data into a training test and a test set. Cross validation was also set up in order to get a more complete understanding of the algorithm and the results it provides, since it splits the dataset on multiple groups, where one *fold* is used to test, and the rest are used to train the algorithm. A 10 *fold* cross validation was used, the rest of the *ML algorithms* were used with its *default settings*. The first results that are going to be presented in the form of a table are the *accuracy* values of each algorithm with a standard train/test split.

The metrics that were used to evaluate the performance of the classifiers were, firstly accuracy where the number of correct predictions is divided by the number of wrong predictions. Precision corresponds to the number of accurate predictions that are actually accurate.

$$Precision = \frac{TP}{TP + FP}$$

Sensitivity in it's own turn, will measure the proportion of people suffering from dementia who got predicted correctly as the ones suffering from non dementia.

$$Sensitivity = \frac{TP}{TP + FN}$$

The last metric that was used to evaluate the *algorithms* was the Specificity, where it's measured the proportion of people who don't have dementia and actually got predicted correctly as not having dementia.

$$Specificity = \frac{TN}{TN + FP}$$

Apart from applying each classifier in both situations, train/test split and cross validation, each sampling technique is also applied in either the dataset with a manual feature selection and the dataset with an automatic feature selection based on correlation.

Table 4. Holdout Sampling Technique, with a 70% split for training and 30% split for testing, using the dataset with a correlation feature selection.

Algorithm	Accuracy	Precision	Sensitivity	Specificity
Decision Tree	91.74%	91.30%	89.36%	93.55%
Random Forest (100 trees)	93.58%	95.45%	89.36%	96.77%
Logistic Regression	93.58%	95.45%	89.36%	96.77%
Support Vector Machine	93.58%	95.45%	89.36%	96.77%
Naive Bayes	94.50%	95.56%	91.49%	96.77%
Neural network	93.58%	95.45%	89.36%	96.77%

Table 5. Holdout Sampling Technique, with a 70% split for training and 30% split for testing, using the dataset with a manual feature selection.

Algorithm	Accuracy	Precision	Sensitivity	Specificity
Decision Tree	88.99%	86.27%	89.80%	88.33%
Random Forest (100 trees)	94.50%	100.00%	87.76%	100.00%
Logistic Regression	94.50%	100.00%	87.76%	100.00%
Support Vector Machine	94.50%	100.00%	87.76%	100.00%
Naive Bayes	95.41%	100.00%	89.80%	100.00%
Neural network	94.50%	100.00%	87.76%	100.00%

As it can be observed, overall the results were good in every algorithm that was used, and very similar in both situations, which can be explained by the lack of instances of the dataset, where only 109 instances are used to test the model. The best *ML* model in the dataset with an automatic feature selection based on correlation, was the *Naive Bayes* algorithm with an accuracy of 94.5%, and 95.41% using the dataset with manual feature selection, and a sensitivity of 91.49% and 89.80% correspondingly in each scenario (also the best in all classifiers) and a *specificity* of 96.77% and 100% in a five way tie. Which means that, the algorithm correctly predicts dementia 95.56%/100% of the times, and in 94.5%/95.41% of times it correctly predicts whether the patient has dementia or not. The classifier that had the worst performance overall in both scenarios was the *Decision Tree* classifier, with an accuracy of 91.74%/88.99%, the worst in both scenarios, and the worst precision also in both scenarios.

Table 6. Cross Validation with 10 folds, using an automatic feature selection based on correlation.

Algorithm	Accuracy	Avg Accuracy	Precision	Sensitivity	Specificity
Decision Tree	93.68% +/- 3.66%	93.66%	95.29%	91.53%	95.70%
Random Forest (100 trees)	94.23% +/- 2.98%	94.21%	97.56%	90.40%	97.85%
Logistic Regression	94.50% +/- 2.57%	94.49%	98.76%	89.83%	98.92%
Support Vector Machine	94.50% +/- 2.57%	94.49%	98.76%	89.83%	98.92%
Naive Bayes	94.23% +/- 3.26%	94.21%	97.56%	90.40%	97.85%
Neural network	94.23% +/- 2.72%	94.21%	98.15%	89.83%	98.39%

Table 7. Cross Validation with 10 folds, using a manual feature selection.

Algorithm	Accuracy	Avg Accuracy	Precision	Sensitivity	Specificity
Decision Tree	94.24% +/- 3.46%	94.21%	97.60%	90.56%	97.81%
Random Forest (100 trees)	94.51% +/- 3.86%	94.49%	98.19%	90.56%	98.36%
Logistic Regression	93.69% +/- 3.64%	93.66%	97.01%	90.00%	97.27%
Support Vector Machine	94.23% +/- 2.98%	94.21%	97.56%	90.40%	97.85%
Naive Bayes	94.24% +/- 3.93%	94.21%	97.60%	90.56%	97.81%
Neural network	94.52% +/- 3.83%	94.49%	98.78%	90.00%	98.91%

With a 10 folds cross validation, in general, the average accuracy actually increased in all classifiers, where the best overall accuracy's were obtained using the automated feature selection scenario. Using the previous dataset or the previous scenario, the best classifiers were either the *SVM* classifier or the *Logistic Regression* algorithm, in which both had an average accuracy of 94.49%. In terms of the best precision, once again *SVM* and *Logistic Regression* are the best, which means these two classifiers are the ones that are capable of predicting with more certainty the accurate predictions that are actually accurate *TP*, with 98.76%. In it's own turn, the *Decision Tree* classifier beats all other algorithms when it comes to measuring *Sensitivity*, with an accuracy of 91.53%. Once again, *Logistic Regression* and *SVM* reached an higher standard when it comes to *Specificity*, which measures the proportion of people who don't have dementia and got predicted correctly, with 98.92% in this category.

Using the dataset with an automated feature selection based on correlation, where the results can be found in *Table 7*, in terms of the average accuracy, the best classifiers by a small margin were the *Random Forest* classifier or *Neural Network* with 94.49% average accuracy in both, which is the same average accuracy that was produced by the *SVM* and *Logistic Regression* classifier using a feature selection based on correlation. In terms of the precision metric, the classifier that had the best result in both datasets with cross validation was the *Neural Network* with a precision of 98.78%, and the best sensitivity belongs to either the *Decision Tree* algorithm or *Random Forest*. The best specificity was once again reached using the *Neural Network* algorithm with 98.91%. Therefore, the best overall algorithm using a manual feature selection is the *Neural Network* classifier.

All in all, using cross validation with 10 folds, the best classifier using the scenario in which an automated feature selection is done based on correlation, the best classifiers were either the *SVM* or *Logistic Regression* classifier. Using the dataset where the feature selection is done manually, the best classifier proved to be the *Neural Network* classifier.

8 Conclusion and future work

This work proved that without a doubt, using real data about *MRI* scans on patients, and through some data preparation steps beforehand, that it's possible to apply machine learning algorithms in order to classify correctly whether an adult suffers from dementia or not. Every model that was consider proved to be efficient in this case, which may be due to the lack of data, where whether *holdout* was used or *cross validation*, or even the dataset with an automatic feature selection based on correlation or the dataset with a manual feature selection, the accuracy of the classifiers was always higher than 89%, with a maximum average *accuracy* of 94.5% with Naive Bayes using a train/test split on the dataset with an automated feature selection and 95.41% also using Naive Bayes on the dataset with a manual feature selection. And 94.49% using cross validation on both scenarios (datasets), where the best classifier in the first scenario was either the *Logistic Regression* or *SVM classifier*, and on the second scenario the *Neural Network* classifier. Consequently, the best possible classifier is either *SVM* or *Logistic Regression* classifier with a 10 folds cross validation using a feature selection based on correlation, since these results are more adequate due to the fact that using *cross validation* every single instance is both used for training and testing. Using a train/test split on a small dataset like the one that was used for this study, can produce exaggerated results, due to the fact that there are not enough instances to both train and test the classifier. Also, while using a manual feature selection is sufficient in most cases, removing correlated attributes is also very important when it comes to preparing the data for the classifiers, since attributes with high correlation often induce a *bias* when they are applied to the classifiers, which it can often inflate results when it comes to the *overall* performance of the classifiers. Hence, and since the dataset was already small and even 10 instances smaller after the data preparation, with only about 363 instances, more data should be supplied in order to test and train the classifiers more adequately, before deploying it in a real world situation. Classifiers like Neural Networks also in most cases, should have some kind of *hyper parameter tuning*, in order to find and use the best parameters for the given data, instead of executing the classifier with the standard parameters. The results, presented in the last section, however do prove that the steps taken in the data preparation were adequate, and the whole process came together and supplied good results for the problem at hand, and there is without a doubt real potential to develop and implement a classifier in a real world situation to predict dementia in adults.

References

[1] The magnitude of dementia occurrence in the world. *Alzheimer Disease and Associated Disorders*. Wimo A, Winblad B, Aguero-Torres H, von Strauss E, 2003.

[2] Machine Learning Techniques for the Diagnosis of Alzheimer's Disease: A Review. *ACM Transactions on Multimedia Computing, Communications and Applications*.
Tanveer, M. Richhariya, Bharat Khan, Riyaj Rashid, A.H. Prasad, Mukesh Khanna, Pritee Lin, Chin-Teng, 2020.

[3] A machine learning based approach for detection of alzheimer's disease using analysis of hippocampus region from MRI scan. *2017 International Conference on Computing Methodologies and Communication (ICCMC)*.
A. Raut and V. Dalal, 2017.

[4] Classification of Alzheimer's Disease using Machine Learning Techniques.
Shahbaz, Muhammad and Ali, Shahzad and Guergachi, Aziz and Niazi, Aneeta and Umer, Amina, 2019.

[5] Dementia: timely diagnosis and early intervention. *BMJ*.
Robinson, Louise and Tang, Eugene and Taylor, John-Paul, 2015.

[6] WHO: Dementia,
`https://www.who.int/news-room/fact-sheets/detail/dementia`

[7] Japanese Alzheimer's Disease Neuroimaging Initiative,
`https://www.gaaindata.org/partner/J-ADNI`

[8] CRISP-DM – a Standard Methodology to Ensure a Good Outcome,
`https://www.datasciencecentral.com/profiles/blogs/crisp-dm-a-standard-methodology-to-ensure-a-good-outcome`

[9] Introduction to Machine Learning Algorithms: Linear Regression,
`https://towardsdatascience.com/introduction-to-machine-learning-algorithms-linear-regression-14c4e325882a`

[10] Decision trees a simple way to visualize a decision,
`https://medium.com/greyatom/decision-trees-a-simple-way-to-visualize-a-decision-dc506a403aeb`

[11] Understanding Random Forest,
https://towardsdatascience.com/understanding-random-forest-58381e0602d2

[12] Naive Bayes Classifier,
https://towardsdatascience.com/naive-bayes-classifier-81d512f50a7c

[13] Support Vector Machine — Introduction to Machine Learning Algorithms,
https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47

[14] Hold-out vs. Cross-validation in Machine Learning,
https://medium.com/@eijaz/holdout-vs-cross-validation-in-machine-learning-7637112d3f8f

[15] Logistic Regression for Machine Learning,
https://machinelearningmastery.com/logistic-regression-for-machine-learning/

[16] Classification: Precision and Recall,
https://developers.google.com/machine-learning/crash-course/classification/precision-and-recall/

[17] ML Metrics: Sensitivity vs. Specificity ,
https://dzone.com/articles/ml-metrics-sensitivity-vs-specificity-difference

[18] 5 Reasons why you should use Cross-Validation in your Data Science Projects
https://towardsdatascience.com/5-reasons-why-you-should-use-cross-validation-in-your-data-science-project-8163311a1e79