CM4107 CW Part2: Comparative Study

Darie-Dragos Mitoiu

School of Computing, RGU, Aberdeen, UK

*Abstract* — In this paper an analysis related to the accuracy and error rate results produced of three machine learning algorithms from Scikit-Learn will be discussed in the context of text classification. The machine learning algorithms that will be discussed are the Recurrent Neural Network (RNN) model acquired from the Python Keras Library, the Multi-Layer Perceptron (MLP) model acquired from Python Scikit-Learn Library and the Support Vector Machines (SVM) model also acquired from the Python Scikit-Learn Library. The type of data that will be analysed in this comparative study it is related to the common text classification problem. The analysis will consist of two different type of problems being investigated, these problems are represented in the form of datasets related to IMDB Movies reviews and sentiment and the identification of SPAM Text Messages in a binary manner. The analysis of the IMDB Movies data has the goal to identify the sentiment present on the review in cause, the sentiment can be either positive or negative, this classification will be represented in a binary manner. The analysis of the SPAM Text Messages data has the goal identify a section of text as either spam or not, this classification or identification of the spam text will be done in a binary manner. The results for the IMDB Movie data represented in the form of accuracy and error rate are approximative 80% accuracy and 20% error rate. The results for the SPAM data, similar to the IMDB Movie data it is represented in the form of accuracy and error rate with values of approximative 90% accuracy and 10% error rate. It is safe to say that the RNN, MLP and the SVM models have performed better on the SPAM Text Messages dataset comparing to the IMDB Movie dataset. This paper also contains a discussion related to Explainable AI topic and the Ethics related to AI projects when it comes to their involvement and impact on human life.

# Comparative Study using ScikitLearn

The aim or ultimate goal of this comparative study is to evaluate the accuracy and error rate produced by three machine learning algorithms from the Python Scikit-Learn and Keras Libraries. This paper will present a comparative study related to three machine learning algorithms in the context of text classification, the three machine learning algorithms are from the Python Libraries called Scikit-Learn and Keras. The machine learning algorithms chosen for this comparative study are the Recurrent Neural Network (RNN) model, Multi-Layer Perceptron (MLP) model and the Support Vector Machines (SVM) model. The Recurrent Neural Network (RNN) algorithm has been chosen as Jason Brownlee argues that this type of Neural Network used in the form of Long Short-Term Memory type of Recurrent Neural Network will produce good results on text classification or language translation problems. The Multi-Layer Perceptron algorithm has been chosen as this algorithm it is an Artificial Neural Network (ANN) with multiple layers and it can be defined as a classic algorithm in the Neural Networks area. The Support Vector Machine (SVM) algorithm has been chosen in this study as this algorithm it uses a discrimination approach in order to solve classification problems.

This paper will present the analysis of two datasets, these two datasets are an IMDB Movies dataset containing reviews and the sentiments for the reviews in cause and a SPAM Text Messages dataset containing possible spam text and the identification of such spam text for each record in cause. The Recurrent Neural Network (RNN) it is a type of Neural Network that presents the potential to dominate difficult machine learning problems that require sequences of inputs. The Recurrent Neural Network (RNN) presents connections that have loops which allow the addition of memory and other information to networks over time. The addition of memory to the networks over time allowing the networks to generalize and learn across sequences of inputs despite individual patterns. The type of Recurrent Neural Network presented in this paper it is the Long Short-Term Memory Network. This type of Recurrent Neural Network it is a powerful one that it is particularly efficient when stacked into a deep configuration related to problems such as language translation. In this paper the Recurrent Neural Network called Long Short-Term Memory from the Python Keras Library will be discussed. This type of Recurrent Neural Network has been used in association with the hyper-parameters called epochs and batch size and other parameters such as embedding size and the text data number of words limitation. The value used for the hyper-parameter called epochs in this study it is equal to 3. The value used for the hyper-parameter called batch size in this study it is equal to 64. The embedding size parameter value used for this study it is equal to 32. The text words limitation used for this study it is equal to 5000. The Grid Search Cross Validation was not used for the Recurrent Neural Network (RNN), the parameters used are static. The Multi-Layer Perceptron (MLP) algorithm can be defined as a classic Artificial Neural Network (ANN) with multiple hidden layers, the Multi-Layer Perceptron Classifier algorithm having similar attributes to the classic Artificial Neural Network (ANN), some example of those similar attributes would be the hidden layer sizes and the learning rate hyper-parameters. The hidden layers sizes or values of the Multi-Layer Perceptron used in this study are 10, 15, 20, 25 and 30. The learning rates used in this study are 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0. The hidden layer sizes and the learning rate values will be passed to a Grid Search Cross Validation function in order to obtain the best or optimal combination of these hyper-parameters for the Multi-Layer Perceptron (MLP) Classifier algorithm. The Support Vector Machines (SVM) machine learning algorithm can be defined as an algorithm that will a binary classification on a given dataset by using a plane that will separate the classes in cause, the separation of classes will be done linearly if this operation it is possible, if this operation is not possible the algorithm can use a higher-dimensional space to allow the finding of a separator. The parameters used in this study for the Support Vector Machines are the C parameter and the Gamma parameter, the values used are between 0.001 and 100.0 for both parameters.

The three selected machine learning algorithms called Recurrent Neural Network (RNN), Multi-Layer Perceptron (MLP) and the Support Vector Machines (SVM) were used on two different datasets related to the common text classification problem. The first dataset it is related to IMDB Movie Reviews, this dataset containg 50000 records and 2 columns or features. The 2 features present in this dataset are the review of a possible movie represented as a string and the sentiment of the review in cause which it is also represented as a string but in a binary manner as either positive or negative. The sentiment feature has been converted from a string to an integer in order to represent the positivity of the movie review in cause as either 1 if positive or 0 if negative. The Multi-Layer Perceptron (MLP) and the Support Vector Machines (SVM) algorithms have been trained tested on a smaller sample of the IMDB Movie dataset, the training data sample having only 25000 instances and the testing data containing the same number of instances. The Recurrent Neural Network (RNN) has been trained using the whole IMDB Movie Review dataset. The Spam Text Messages dataset contains 5574 records and 2 features, the features being the spam text and the identification of the message in cause as either spam or not, these messages were converted to integers as 1 for span and 0 for ham.

## Visualisation of Results

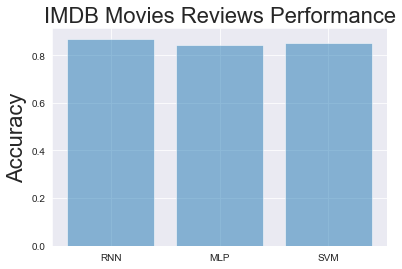


Fig. 1. RNN, MLP and SVM Accuracy for the IMDB Movies Data

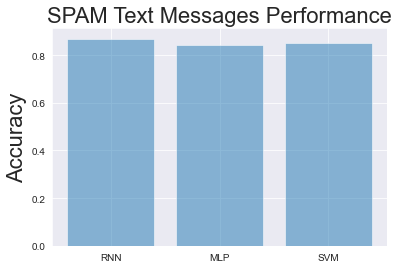


Fig. 2. RNN, MLP and SVM Accuracy for the SPAM Data

## Discussion of Results

The IMDB Movies Reviews and the Spam Text Messages datasets were pre-processed by making use of a function that eliminated irrelevant text elements such as HTML keywords and other non-words elements, characters and symbols that did not bring value to the text classification operation later on. Once the non-essential words have been eliminated by the pre-processing function, the last instructions executed by the function were to add an emoticon at the end of each text in cause and to convert the text to lower case characters. A Term Frequency-Inverse Document Frequency (TFIDF) Vectorizer was created from the Sci-kit Learn Python Library in order to transform text to feature vectors that can be used as input later on. A TFIDF Vectorizer object was then passed to a function that has the role to create a pipeline from a vectorizer associated with the model in cause, meaning a pipeline object was created for the Multi-Layer Perceptron (MLP) model and the Support Vector Machines (SVM) model, an exception being the Recurrent Neural Network (RNN) model which was evaluated using static parameters. A Grid Search Cross Validation object was then created for each model using their distinct pipeline and parameters. In order to evaluate the performance in the form of accuracy and error rate for the Multi-Layer Perceptron (MLP) model and the Support Vector Machines (SVM) model a Grid Search Cross Validation was performed using a series of parameters for the models in cause. The Recurrent Neural Network (RNN) was used on both IMDB Movies Reviews dataset and the Spam Text Messages dataset with static hyper-parameters, meaning there was no use of the Grid Search Cross Validation function in order to find the optimal parameters.

A Grid Search Cross Validation using 3 folds was performed for the Multi-Layer Perceptron (MLP) model and the optimal grid parameters for the hidden layers size and the learning rate found from 50 candidates and 150 fits are: the hidden layer size with a value of 25 and the learning rate with a value of 0.1 for the IMDB Movie the model evaluated with these parameters was able to produce an accuracy of 0.834 from the Cross Validation best score and an accuracy of 0.844 when used on the movies testing data. A Grid Search Cross Validation was performed for the Multi-Layer Perceptron model for the Spam Data and the optimal parameters for the hidden layer size layers and the learning rate that were found from 50 candidates and 150 fits are: the hidden layer size value equal to 10 and the learning rate value equal to 0.6. A Grid Search Cross Validation using 3 folds was performed also for the Support Vector Machines (SVM) model and the optimal grid parameters for the C parameter and the Gamma parameter found from 18 candidates and 54 fits are: the C parameter with a value of 10.0 and the Gamma parameter with a value of 0.1 for the IMDB Movie data and last but not the least the model evaluated with these parameters was able to produce an accuracy of 0.831 from the Cross Validation best score and an accuracy of 0.852 when used on the movies testing data. A Grid Search Cross Validation was performed for the Support Vector Machines (SVM) model for the Spam Data and the optimal parameters for the C parameter and the Gamma parameter that were found from 18 candidates and 54 fits are the C parameter value equal to 10.0 and the Gamma parameter value equal to 0.1, which produced an accuracy of 0.981 for Cross Validation and 0.987 for testing.

# Explainable AI (XAI)

## Overview of explainable XAI

The term called “Explainable AI” or “XAI” refers to the Artificial Intelligence theory and development of computer systems which are able to perform specific tasks that would normally require human intelligence to do so, but in association with a method that allows a human being to understand the process that allows a machine to learn from a specific set of information given by its creator. The “Explainable Ai” also implies the transformation of Artificial Neural Networks (ANNs) in association with their visualisations of the model’s specifications into a readable format by a human being. In other words, All the decisions of the Artificial Intelligence algorithm must be transparent, giving the ability to human practitioners to dissect and understand the algorithm’s behaviour and not allow the algorithm to execute into a black box manner.

## Legal, ethical and Social implications of XAI

The Artificial Intelligence projects are an important element in the day-to-day life these days for human beings and because of this some legal authorities and regulators are asking for the Artificial Intelligence projects to become more explainable, at least to a certain level of understanding by a regular person that does not specifically present any technical knowledge or skills in this area of interest (Jaffe 2020). The law will imply the provision of information related to the implementation of the Artificial Intelligence algorithms, this information must be related to the two main elements, these elements are: what information is the Artificial Intelligence algorithm has concluded about certain problem domains and how the Artificial Intelligence algorithm has got to that conclusion. These 2 main elements will bring a series of legal issues that can be represented largely by the key law called the General Data Protection Regulation 2018.

The black box in which the Artificial Intelligence algorithms are performing their activities in order to make decisions for a specific problem domain will bring into discussion laws related to Data Protection as there is a significant increase of desire that people present for the right to know how companies are holding and processing their personal information. This black box manner will also imply the companies and regulators to find out a method to solve the problem of the processing and storage of data as the Artificial Intelligence algorithm will process information related to a person in order to avoid any legal problems due to laws such as the General Data Protection Regulation 2018 which protects the information of customers, users or projects subjects.

According to Jaffe (2020), when it comes to legal problems caused by the accusations of bias, particularly in the employment contexts, a defendant in a discrimination suit will need to provide information related to what elements did result in a possible employment decision or what elements did not motivate the employment decision. A situation like this will require the Artificial Intelligence algorithm are used on data from resumes the algorithms need to justify the decision made about a specific candidate, meaning it has to give enough information on how it got to that result of selection or rejection.

## Explaiability and Transparency on chosen algorithms

The aim or ultimate goal of this study was to evaluate the accuracy and error rate produced by three machine learning algorithms from the Python Scikit-Learn and Keras Libraries. In the first section of this paper, it was presented a comparative study related to three machine learning algorithms in the context of text classification, the three machine learning algorithms are from the Python Libraries called Scikit-Learn and Keras. The machine learning algorithms that were chosen for this comparative study are the Recurrent Neural Network (RNN) model, Multi-Layer Perceptron (MLP) model and the Support Vector Machines (SVM) model. The Recurrent Neural Network (RNN) algorithm has been chosen as Jason Brownlee argues that this type of Neural Network used in the form of Long Short-Term Memory type of Recurrent Neural Network will produce good results on text classification or language translation problems. According to Rudin and Radin (2019) when talking about black box machine learning algorithms we refer to models created directly from data which means the human beings do not have a clear understanding about the behaviour of these algorithms, this applies even to those that implemented such algorithms, meaning that they do not poses an understanding of the factors and variables used by the algorithms in order to make predictions.

It is safe to say the complexity of the algorithms that present an architecture based on Neural Networks such as the Multi-Layer Perceptron (MLP) would not be very transparent as the nature of the algorithm adds a certain level of complexity when talking about Explainable AI, as specific computation will be performed on each layer of the Multi-Layer Perceptron (MLP) in order to provide predictions for a problem domain.

Similar to the Multi-Layer Perceptron (MLP), the Recurrent Neural Network (RNN) algorithm it has an architecture based on the classical Artificial Neural Network (ANN) algorithm, which means that this algorithm will also bring a level of complexity which can categorise the algorithm as not very transparent due to the recurrent nature of computation in order to make predictions for a problem domain, in addition to the complexity added by the recurrent behaviour, the algorithm complex behaviour may also contribute to a difficult manner in which it could be explained.

The Support Vector Machine (SVM) algorithm has been chosen in this study as this algorithm it uses a discrimination approach in order to solve classification problems. The discrimination approach in which this algorithm separates the classes of a specific dataset for a problem domain it is done by making use of linear plane that separates the classes in cause. It is safe to say that due to a simpler architecture the Support Vector Machines (SVM) it is more transparent due to the architecture it poses in order to make predictions for a problem domain and because of this transparency, the Support Vector Machines (SVM) it has some advantage when it comes to Explainable Ai comparing to other algorithms that are different from an architecture point of view. An exception made to the lack of transparency and a difficult manner in which the algorithm can be explained it is the situation when the Support Vector Machines (SVM) cannot separate the classes of a specific dataset when using a linear plane.

# ethical AI

## Artificial Intelligence Key Ethical Issues

According to the European Commission (2020), a trustworthy Artificial Intelligence project must follow the 3 principles of being lawful, ethical and robust. The main key ethical problems represented in the form of guidelines that must be respected when creating Artificial Intelligence Systems or using such systems in order to make them trustworthy are represented by the 7 key requirements, these requirements are:

1. Human agency and Oversight,
2. Technical Robustness and Safety,
3. Privacy and Data Governance,
4. Transparency,
5. Diversity, Non-Discrimination and Fairness,
6. Societal and Environmental well-being,
7. Accountability.

The Human agency and Oversight requirement or element refers to the power given to the human beings by the Artificial Intelligence projects in order to make informed decisions and fostering fundamental rights in association with the proper mechanism needed to secure the human-in-command approaches. The Technical Robustness and Safety requirement or element refers to the insurance that the Artificial Intelligence projects are resilient and secure, in other words the AI must present a decent accuracy in order to be reliable. The Privacy and Data governance requirement or element refers to ensuring full respect for privacy and data protection and ensuring legitimised access to the data in cause. The Transparency requirement or element refers to the Artificial Intelligence models, the data they are using and the overall system to be transparent and not to hide sections of the project by giving full information related to the decisions made by the AI. The Diversity, Non-Discrimination and Fairness requirement or element refers to unfair bias that must be avoided and eliminated by existent when talking about AI projects which should be available to all human beings despite other differences. The Societal and Environmental well-being requirement or element implies the AI projects to help all human beings, meaning the AI system must be sustainable and to take into account the environment by considering other living beings. Last but not the least the Accountability requirement or element implies a mechanism that must be put in place to ensure an assessment of the AI systems and to redress the system if adequate.

## Artificial Intelligence Risks

The risks presented by the AI systems could be represented by 2 key elements, these elements are: security and safety. The safety element refers to possible accidents caused by unpredictable negative behaviour of the AI projects. The security element refers to the economic point of view and military assets where the AI projects could end up in the hand of bad actors which would use the systems to cause harm to other human beings.

## Artificial Intelligence Machine Morality

##### References

1. Jason Brownlee (2016, July 7). *Crash Course in Recurrent Neural Networks for Deep Learning* [Online]. Avaiable: <https://machinelearningmastery.com/crash-course-recurrent-neural-networks-deep-learning/>
2. Jason Brownlee (2016, May 17). *Crash Course On Multi-Layer Perceptron Neural Networks* [Online]. Avaiable: <https://machinelearningmastery.com/neural-networks-crash-course/>
3. Jason Brownlee (2020, August 15). *Support Vector Machines for Machine Learning* [Online]. Avaiable: [https://machinelearningmastery.com/  
   support-vector-machines-for-machine-learning/](https://machinelearningmastery.com/support-vector-machines-for-machine-learning/)
4. Jason Brownlee (2019, January 25). *Understand the Impact of Learning Rate on Neural Network Performance* [Online]. Avaiable: <https://machinelearningmastery.com/understand-the-dynamics-of-learning-rate-on-deep-learning-neural-networks/>
5. Jason Brownlee (2020, August 21). *How to Control the Stability of Training Neural Networks With the Batch Size* [Online]. Avaiable: <https://machinelearningmastery.com/how-to-control-the-speed-and-stability-of-training-neural-networks-with-gradient-descent-batch-size/>
6. Harshita Pandey (2020, May 9). *Spam Classification with Tensorflow-Keras* [Online]. Avaiable: <https://medium.com/analytics-vidhya/spam-classification-with-tensorflow-keras-7e9fb8ace263>
7. Kevin Casey (2019, May 22). *What is explainable AI?* [Online]. Avaiable: <https://enterprisersproject.com/article/2019/5/what-explainable-ai>
8. Peter Jaffe (2020, September 17). *Are we being too hard on HAL? Some thoughts on the legal need for “explainable” artificial intelligence* [Online]. Avaiable: [https://www.lexology.com/library/detail.aspx?g=  
   0ee05837-3eb5-41ea-bd16-6d8f2d502c3f](https://www.lexology.com/library/detail.aspx?g=0ee05837-3eb5-41ea-bd16-6d8f2d502c3f)
9. Rudin, C., & Radin, J. (2019). Why Are We Using Black Box Models in AI When We Don’t Need To? A Lesson From An Explainable AI Competition. *Harvard Data Science Review*, *1*(2). <https://doi.org/10.1162/99608f92.5a8a3a3d>
10. Bryson, J.J. Patiency is not a virtue: the design of intelligent systems and systems of ethics. *Ethics Inf Technol* **20,**15–26 (2018). <https://doi.org/10.1007/s10676-018-9448-6>
11. Gunkel, D.J., Bryson, J. Introduction to the Special Issue on Machine Morality: The Machine as Moral Agent and Patient. *Philos. Technol.* **27,**5–8 (2014). <https://doi.org/10.1007/s13347-014-0151-1>
12. Bryson, J.J. and Kime, P.P., 2011, June. Just an artifact: Why machines are perceived as moral agents. In *Twenty-second international joint conference on artificial intelligence*.
13. European Commision 2020. *Ethics guidelines for trustworthy AI* [Online]. Avaiable from: <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai> [Accessed: 16 December 2020]
14. Wiratunga, N., 2020. Artificial Neural Nets (ANNs), CM4107 [Powerpoint Presentation]. Artificial Neural Nets (ANNs). Advanced Artificial Intelligence. Robert Gordon University. School of Computing. 12 December. Available: [http://campusmoodle.rgu.ac.uk/mod/resource  
    /view.php?id=3687426](http://campusmoodle.rgu.ac.uk/mod/resource/view.php?id=3687426)
15. Wiratunga, N., 2020, RNN. Aberdeen. Robert Gordon University
16. Wiratunga, N., 2020, TextAnalysis-part1-IMDB. Aberdeen. Robert Gordon University
17. Wiratunga, N., 2020, TextAnalysis-part2-SMS Text Detection. Aberdeen. Robert Gordon University
18. Wiratunga, N., 2020, ScikitLearn-Evaluation-part1. Aberdeen. Robert Gordon University
19. Wiratunga, N., 2020, ScikitLearn-Evaluation-part2. Aberdeen. Robert Gordon University
20. Wiratunga, N., 2020, ScikitLearn-Evaluation-part3. Aberdeen. Robert Gordon University
21. G. Eason, B. Noble, and I.N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529-551, April 1955. (*references*)