Data Mining & Machine Learning II Terminal Assignment (2020)

Student-Name: Richard Saldanha.

Student-id: x18183034@student.ncirle.ie

Department: School of Computing -Master of Science in Data Analytics- September 2019 (Full Time).

Module: Data Mining & Machine Learning II.

Group: MSCDA-A.

Date of Submission: 11th May 2020

Q1.Discuss the following with respect to complex data mining/machine learning projects?

1.1. Dimensionality Reduction Techniques

- Dimensionality reduction is a useful technique which is employed to carry out various activities such as discovering nonlinear, nonlocal relationships in data which are not needed in the feature space of the dataset. By reducing the number of dimensions in our dataset we can easily visualize out data in either a two-dimensional space or a three-dimensional space. More over if we do not perform dimensional reductions on our datasets then in the stage when we have to train our model with many unwanted features it would consume more time and along with that the space complexity of the system would also increase which can some time lead to situation like overfitting.
- Dimensionality reduction techniques work around an objective which is to search for only the small essential set of features which has the potential to describe the entire dataset and thus it has two fundamental aspects which are feature selection and feature extraction. Feature selection is nothing but searching for the most relevant feature to a problem and we generally do this based on our intuition about which features should we include and would be useful to produce fruitful results from our analysis while feature extraction is nothing but searching for new features after transforming the data from a high dimensional space to a lower dimensional space.
- There are various techniques which are employed for dimensionality reduction and they are as follows:

1. **Finding the missing values**: If the dataset under observation which is chosen for analysis has too many missing values then we just drop these variables from our datasets which does not have any significant contribution.
2. **Applying Variance filter**: In the event for exploratory data analysis this technique is used to identify and drop constant variables from the dataset which does not have any impact on the target variable as these variables have low variance we remove or we can drop these variables from our datasets.
3. **High Correlation filter**: The existence of multi collinearity can occur when in the datasets there are some pairs or even all independent variables which are having a strong correlation between then due to which it impacts the results of the analysis so we employ this technique in order to drop highly correlated variables from the datasets in order to achieve better results.
4. **Random Forest**: This technique is used when we must find out the importance of features in the dataset and retain only the essential features which will make a significant contribution and provide better results of the analysis.

5. **Backward Feature Elimination** and **Forward Feature Selection:** This technique is used when we must perform dimensionality reduction on small datasets since they consume a lot of computational time.
6. **Factor Analysis**: In this technique we group the variables which are having a high rate of correlation between them and due to this group formation there would be a low rate of correlation between the groups and such group are called as factor and they are small in numbers compared to the original dimensions of the data. These factors can sometime be difficult to observe.
7. **Linear Discriminate Analysis (LDA):** This technique is one of the popular techniques which is employed for supervised dimensionality reduction which creates a linear combination of the number of independent features in the dataset relative to the data being described due to this reason it yields larger mean differences between the described classes.
8. **Neural Network (NN):** Neural Networks also sometimes referred to as connectionist learning network because of the way in which the three layers input, hidden and output layers are connected is also employed for dimensionality reduction since they exhibit many advantages such as high tolerance with respect to noisy data, ability to classify patterns even when they have not been trained on that data also they use parallelization technique in order to speed up the computational processes.
9. **Principal component Analysis (PCA):** This technique transforms the variables into a new set of variables which are a linear combination of the original variables which are called as the principal components which are nothing but eigen pairs that describe the direction in the original feature space with the greatest variance in the data. This technique is an orthogonal, linear transformation technique that transforms the data to a new coordinate system such that the greatest variance by some projection of the data lies on the first principal component, the second greatest variance on the second component and so on. This technique is a popular technique for transforming a dataset onto a lower dimensional space for visualization and future exploration.
10. **Independent Component Analysis (ICA):** The major difference between PCA and ICA technique is that PCA looks for factors which are uncorrelated while ICA looks for factors which are independent. If two variables are uncorrelated, it means that there does not exist linear relation between them and if they are independent, it means they are not dependent on others.
11. **Single Value Decomposition (SVD):** This technique is used to decompose the original dataset into its constituents, resulting in dimensionality reduction. SVD decomposes the original variables into three constituent matrices. This essential use of this technique is to remove redundant features from the dataset. It uses the concept of Eigenvalues and Eigenvectors to determine those three matrices.
12. **t-distributed stochastic neighbor embedding (t-SNE):** This technique is suitable and employed to search for patterns easily in a non-linear way and like PCA it is a good technique for performing visualization of data. This technique is capable to retain both the local and global structure of the data at the same time as well as it can calculate the probability of similarity of points in high dimensional space as well as in low dimensional space.

1.2. Feature Engineering

- Feature engineering can be defined as the process which is involved in the transformation of raw data into meaningful features that gives a better representation of the underlying problems to the machine learning predictive models which results in an improved versions of the model accuracy on the dataset chosen for the analysis.
- Features are nothing but a numeric representation of an aspect of raw data and in the machine learning pipeline it sits in between the data and models and this feature is taken as the input to the machine learning models applied on the datasets under consideration in order to derive meaningful insights or to make better prediction analysis.
- Feature engineering is a crucial step involved in data pre-processing as a lot of time is spend on finding the right features which will fit the machine learning models because if it is not done properly then the machine leaning models will not yield better and high accuracy results.
- Feature engineering can also be considered as a feature construction step in which we either decompose our data under analysis or we aggregate our data in its raw format in order to better describe the underlying problems.
- The structure of the data can be in any format since in real time scenario data does not exist in one format only. The data could be either in a tabular format, it could be images, documents or any text format. A tabular format data is described of having rows and columns where columns are nothing, but the features and rows are the number of observations. In case of images the special observation like line in an image could be its feature and for videos a character image would be the feature of the video data. In the event of performing natural language processing or text analysis on data which are of the form of a document or tweets from social media posts the word count could be a feature while in speech recognition the utterance of words are observation and its features might be a single word or a special sentence depending upon the problem taken for analysis.
- Feature engineering can be done either by incorporating the traditional approach using domain knowledge which is a time consuming process, it can cause errors and it becomes difficult when dealing with time series based datasets and so due to this reason it can lead to invalid machine leaning solutions and so due to this reason it is necessary that we incorporate automated feature engineering technique to overcome this problems and this can be achieved by using an open source python library which is “Featuretools” as this library is very useful in various applications like customer segmentation.
- Since feature engineering is a tedious task and an important step otherwise it will produce misleading results using the automated feature engineering technique such as Featuretools library by python we can build many valid features for any datasets of the relational dataset format and applying them would help us in building high performance machine learning models for predictions or classification tasks thereby helping us to tackle many machine learning problems with ease and produce good results for analysis.

1.3. Dataset Size

- One of the most crucial element for conducing any analysis, it could be prediction, classification, regression or even time series analysis using any techniques and methodologies which fall under the two broad categories of the field of data science or data analysis which are machine learning and deep learning and if we do not have the right type and the right amount of data for analysis then these machine learning or deep leaning models cannot bear the right fruits of the results from the analysis.
- In context of the size of the data the volume of the data you are using to carry out any analysis like classification, prediction, regression and many others depend on two factors the first being the complexity of the problem under study and the second is the complexity of the algorithm chosen for analysis.
- It is a matter of concern that we should know about the size of the training datasets in order to best fit our model. There are various methods one can employ to calculate the size for a sample and one such method is the statistical heuristic method and it is quite suitable to deal with problems based on classification as it used as a function for the number of classes, input features or model parameters. There are several ways in which the statistical heuristic function can be applied such as a factor of the number of classes in which suppose we have 10 classes then we should have x independent examples for these classes where x could be any positive number of examples, factor of the number of input features in which for a set of input features there should be x% more examples than the input features and in the event for factor of the number of model parameters there should be x independent examples for each of the parameters in the model.
- It is evident that when we are applying machine learning algorithms we need sufficient amount of data for training the model as well as for testing the model so for that roughly around tens or thousands of record would be sufficient, however if we are applying deep learning algorithms like CNN, RNN or LSTM well in that case the volume of the data or the dataset size should be huge for providing good accuracy of the models.
- There are various data sources which are licensed as open source to use which provides a plethora of datasets for analysis in different sizes and file format. When we select a dataset for performing either classification or prediction or any other analysis it is important that the right dataset is chosen which will address the problem and which is sufficiently large otherwise if the dataset is not large enough after data pre-processing stage let's assume that a few couple of records are removed and there is only few records now left to apply the machine learning or deep learning models and to draw out the results so in that case there is a chance that the model will not train properly as there is only small amount of data which is considered for training dataset and no sufficient data for testing the model so in that case there can arise a problem of accuracy with the model's performance for addressing the problem so the size of the datasets really matters for analysis.
- There is also one way which could be employed which is nothing but using a visualization technique like using a simple line plot graph which will actually show the dependence of dataset size on the performance of the model in which the x-axis would represent dataset

size for training i.e. the training dataset and on the y-axis the model performance which will rough estimate of how much data is required and could help us to decide on the datasets size for analysis.

1.4. Computational Cost

- In Machine learning, the computational cost involved for fitting a single observation say x0 is O(N) flops but except in the case for oversimplified cases like square kernels. The M basis functions costs O(M) for one evaluation and it is expressed as ~ O(logN) on the other hand these basic function methods have an initial cost of at least O(NM^2+ M^3).
- The computational cost is essentially that of a regression problem of this size, which is O(p^4 2^p) and is manageable when the value of p is less than 20.
- In machine learning or deep learning approach it is utmost important to know the potential and computational efficiency of the algorithms so that we can make effective use of the computational resources like RAM, Storage space. Once we can analyse the algorithm, we then can measure its potential when we use it on larger datasets.
- Big O and Big Ω are the two universal symbols which are used to describe the worst case and best-case scenarios which describes the computational complexity of the machine learning or deep learning algorithms and we can measure an algorithm based on how it handles these inputs by considering a generalized function say f(n) where f is the function and n is the inputs.
- There are several common computational complexity classes like constant time O(1), logarithmic time O(log n), linear time O(n) and many more and which are arranged as per the computational speed of importance. Regardless of the dataset size the constant time is the fastest which takes in one operation or one unit of resources to deal with.
- When we perform machine learning and deep neural network analysis the complexity of the models and its time and speed involved in processing are mostly proportional to the size of the input data provided. While the (Big O) complexity measuring algorithm is the key element for measuring the efficiency of an algorithm over time and space measurements there are several additional factors which needs to be considered for efficiently training the machine learning models or deep neural network model.

Q2. Review the Review of classical dimensionality reduction and sample selection methods for largescale data processing.

1. Structure Title: The authors (Xu *et al.*, 2019) of this paper have given a title which is "Review of classical dimensionality reduction and sample selection methods for large-scale data processing" which seems to be slightly incorrect because they have not mentioned or specified as a part of the title about whether it is a review for machine learning algorithms or deep learning algorithms which can sometimes be misleading to the intended audience so a correct title according to me should be "A review of classical dimensionality reduction and sample selection techniques for processing large-scale data in the field of Artificial Intelligence."

2. Abstract: The authors (Xu *et al.*, 2019) of this paper in the abstract section have firstly described about the importance of samples and high dimensional attributes in various fields of areas such as data mining, pattern recognition and many others. Then the authors have discussed about the overview, the application and combination of use of dimensionality reduction and sample selection methods with different machine leaning and deep learning methods. They also highlighted and discussed that how an application framework that combines these methods in context of two aspects which are sequential and simultaneous would provide ideal results for both the aspects and concluded that these two methods are efficient and effective techniques for processing large scale data. They also discussed their future work as a part of their research on the review of dimensionality reduction and sample selection methods for processing large scale data.
3. Introduction: The authors (Xu *et al.*, 2019) of this paper in the introduction section have discussed about the negative influential impact of training samples such as more storage requirement, increasing complexity and decreasing prediction accuracy rate of machine learning models and have further more discussed about how high dimensional data can face problems when analysed using machine learning models if the important parameter of samples which are its columns also known as features or attributes are not selected. The authors have also discussed that sample selection and dimensionality reduction techniques are proposed to deal with problems faced by samples and features and have cited papers to support their point. The authors have discussed about various sample selection methods and have also shared a light on dimensionality reduction methods and its impact if not done correctly and have also mentioned about the relevant works done by other authors and with that laid down their motive behind conducting this research and have claimed to be novel approach in this domain. Lastly to an end with the introduction section the authors have decided to organize their research paper in such a way that each topic is divided into different sections so that the intended audience would get a clarity about the research topic.
4. Graphical abstracts and/or highlights: The authors (Xu *et al.*, 2019) of this paper have neither summarized nor visualized their core findings as a separate dedicated section either presented at the first page or at the last page of the paper from their review on methods for dimensionality reduction and sample selection which should have been done which would be beneficial to the intended audience in order to make notes and to have an eye catcher summary either through a visual representation or through a set of bullet points of the results or findings of the research work after reading the entire paper.
5. Methodology : This research paper is a review based research so in terms of the data required for reviewing the classical dimensionality reduction and sample selection methods the authors (Xu *et al.*, 2019) have used a total of 106 research papers. Adopting this methodology for only review purpose seems suitable since the work done in the past by the different researchers over a couple of years can help to give a good review about the two major aspects of processing huge volume of data which are dimensionality reduction and sample selection methods. Since the scope of this research work is to only review so according to me if the author would have considered limited number of papers say around

30 related papers as part of their related work and performed an experiment by considering a publicly available dataset and apply the respective dimensionality reduction techniques which are feature extraction and feature selection methods as well as sampling selection method and apply machine learning or deep learning models to the dataset the results obtained through the analysis would not only have a theoretical evidence from past research work but also a practical evidence through the results from the experiment. In this research work there is enough information for conducting review and a graphical representation in the form of figure to apply combination of the data pre-processing methods, but no methodology defined to implement it practically. The authors have made sure that each section and their respective sub-section is reviewed step by step through past papers and figures and the need to propose a new approach which is a combination of sample selection and dimensionality reduction method is well explained and reviewed through past papers as well as through diagrammatic representation.

6. Results: When compared to other research papers it is evident and observed that after conducting analysis on a particular field of study using different respective methodologies the researchers often create a separate section called as 'Results' which contains the results achieved from the analysis which makes the life easy for the intended audience. The authors (Xu *et al.*, 2019) in their paper have not provided any dedicated section which documents the results obtained from the review of the classical dimensional reduction and sample selection method which should have been done which makes life easy to the readers otherwise they would have to fetch for what was the results obtained by rereading the paper over and over again which is a time consuming process. Despite not having any dedicated section for results from each section they got valuable results of the review which are:
    1. Sample selection in machine learning means to get rid of noisy data and analysing large scale unprocessed samples could lead to several issues like overfitting and dealing with such samples can be a time-consuming process so the need of sample selection is required.
    2. To overcome overfitting of the training model dimensionality reduction is a mandatory and it is seen widely in various deep learning and machine learning applications for pre-processing of datasets.
7. Conclusion/Discussion: The authors (Xu *et al.*, 2019) have finally concluded their critical review on dimensionality reduction and sample selection technique by first discussing about the problems such as overfitting, high and low computational speed and complexity faced by large scale data and then emphasized that the main contribution of their research towards the body of knowledge is to provide a summary about the different applications of dimensionality reduction technique and sample selection methods when applied with different flavours of machine learning and deep learning algorithm. They also discussed about the contribution of CNN in various fields of study one of it being image classification and recognition and stated that despite being a good way of processing large scale data there does not exist a universal method of dimensionality reduction and sample selection technique which is confident enough to tackle all problems faced by large scale datasets.

The authors have given reference to two figures which are (Fig. 9) and (Fig. 10) and with that have put forward their point about if a combined version of both the methods are used simultaneously then probability of yielding good performance is at a high rate and this approach was put forward by them because from the evidence gathered from previous work by different researchers it was found that either one of the method is used at a time but not both the methods used together simultaneously. The review and suggestion made by the authors about having a combined versions of dimensionality reduction methods and sample selection method would surely open doors towards improvising the processing of large-scale datasets.

8. Language : The authors (Xu *et al.*, 2019) have used simple English language with appropriate grammar consideration in different sections of the paper with a thought process that the intended audience would be able to understand the research work clearly. Apart from the use of simple English language the authors have also visually expressed their review on dimensionality reduction and sample selection methods in the form of figures which provides a good overview of both the dimensionality reduction and sample selection methods.
9. Previous Research : The authors (Xu *et al.*, 2019) have not created any dedicated section like 'Related Work' or 'Literature Review' which is evident in most of the research papers but they have correctly cited 106 research papers as per the IEEE format to support their research review on classical dimensional reduction and sample selection methods for processing large scale data which is observed in different sections of the research papers. The selection of research papers does not only limit to review of classical dimensional reduction and sample selection methods but they have also made sure that whatever they have written in different sections of the paper has a support proof of evidence in the form of giving reference to the piece of work from where they have referred. Apart from citation they have also expressed their gratitude towards three contributors by creating a separate section known as 'Acknowledgements' for supporting them in their research work which is a good gesture of appreciation.

Q3. Discuss the following data mining/machine learning techniques, critically highlighting and assessing both potential benefits and potential limitations:

3.1. Ensemble Approaches

- The ensemble approach means "combining" multiple machine learning models which is one of the most effective techniques for delivering better results in machine leaning analysis.
- The ensemble approaches use a combination of multiple learning algorithms to address the statistical or machine leaning based problems in order to achieve a better predictive results than could have been achieved from if we have considered any of the single machine learning algorithms alone for the analysis. The combination method used when dealing with numerical values is unweighted average and for categorical values, we make use of majority count or simply voting. Apart from that when we train a leaner model in order to

combine the individual leaners then such combing is called as stacking or combing by layer technique in which the output from the individual learners are used for training the data for the next layers.

- The driving force behind the existence of an ensemble learning method is that there does not exist any single machine leaning or deep leaning algorithm which claims to be the most accurate and optimized for any situation since every algorithm is designed to address different types of problems since their algorithmic design structure like input parameters, representation of data, training of data varies from one algorithm to another which is the base line of the No Free Lunch Theorem and the solution to this is the Ensemble Learning approach in which when we combine a set of base leaner models we will get results of higher accuracy.
- There are two major aspects which formulates the creation of ensemble methods they are machine learning or deep learning models which are trained on a given problem for analysis and a decision rule which determine how the results of those models would be combined to produce the final output. During the observation of different models trained on the same problem for analysis if it is found that if the models are diverse in nature means they are uncorrelated then if the models are weak to deliver better performance for the analysis their combined or ensembled versions would outperform in performance results put forward by individual models.
- The underlying models are also known as base learners which are derived from a training data by a base learning algorithm. There are two types of ensembles that could be generated during this process which are Homogenous ensemble which relies on a single base learning algorithm and the other one is Heterogenous ensembles which does not rely on a single base learning algorithm.
- There are two techniques which are employed for Ensemble approach which are Bagging which is also known as (Boot Strap Aggregation since there is involvement of bootstrap sampling and aggregating) technique and Boosting technique. In Bagging technique, suppose if we have our dataset say ‘D’ which has many rows and columns and if we have a set of many base learning models say ‘M1’, ‘M2’, ‘M3’ to which a sample of dataset is provided which is ‘D’ suppose the original dataset holds ‘n’ records and we select a sample of records and provide to each of the models individually for training where the number of samples provided for each models should be less than the total number of samples, thus the models are trained from these sampled data. Now, if we provide a new data say ‘Dt’ and we are trying to predict the output then in that case we take the average of the response as the prediction results for the analysis. Random Forest classifier technique is a bagging technique where decision trees are used in which row sampling and feature sampling method is used done for generating the output. The random forest technique can be used for both classification and regression problems in which we take the majority vote counts of output from all decision tress for classification-based problems and take the average or the mean of the outputs from all the decision trees for regression-based problems.

- In Boosting technique, we carry out training on series of classifiers model emphasizing a lot on the number of instances till we get the previous one wrong and then finally we take a weighted average of the results obtained for the predictions of the results. The underline concept of the boosting technique is to identify and rectify the mistakes made by the weak classifiers so that we combine or covert all the results achieved from the weak leaners to strong learners in order to increase the accuracy of the model and get more precise and more accurate predictions. There are three types of boosting technique which are:
    1. Adaptive boosting - Combining several weak leaners in to one single learner and which can be used for both classification and regression problems.
    2. Gradient Boosting - In this method the base learners which are generated sequentially are more effective then the previous base learners. When we compare it to Adaptive boosting, we do not increment the weights of the misclassified output but in this technique, we try to optimize the loss function or errors from the previous learner, and this also can be used for classification as well as regression model.
    3. XGBoosting technique - This technique is nothing but an advanced version of the gradient boosting method in which the aim is to improve the computational speed and the performance of the gradient boosting model.
- There are many advantages or rather potential benefits of using ensemble approaches since it has the capability to improve the predictive ability of the machine learning and deep learning models and they are as follows:

1. The accuracy achieved through the ensembles methods is more when compared to the individual learners because it makes use of multiple machine learning algorithms together to perform the same task.
2. Higher consistency in the results is obtained because it avoids overfitting
3. It also reduces the bias and variance errors due to which the model is more stable and robust in nature.
4. The ensemble approach can be used for both classification as well as regression-based problems.
5. By adopting ensemble approach for our analysis, it is possible to discover the linear and non -linear properties in the datasets.

- There are also some drawbacks or limitations of using ensemble approaches and they are as follows:

1. It is necessary to understand the problem under study and then apply either Bagging or Boosting technique of the ensemble approach because we cannot apply any thing as we wish since bagging technique is suited for dealing when the machine learning models have high variance rate i.e. they would lead to overfitting where as boosting is used the machine learning models are biased in nature.
2. It is observed that adopting the ensemble approach is computationally expensive phenomenon which results in increasing processing and learning time of the models as well as it adds memory constraints to the system.

3. The ensemble approach has a weak interpret ability because of the increased complexity in processing and due to this reason, it may become a tedious task for deriving valuable insights from the business prospects.

3.2.Unsupervised Deep Learning with Artificial Neural Networks

- In unsupervised deep learning the model attempts to describes the patterns and associations among a set of input measures since there is no guidance or supervision regarding the output measures.
- In other words, unsupervised machine learning or deep learning algorithms basically tries to learn by themselves from the environment without being trained or supervised by the environment. In unsupervised learning we train our datasets on unlabelled data without any guidance which is to discover by itself any hidden patterns or trends in the data and using the unsupervised deep learning methods we can solve clustering and regression-based problems.
- In the context of unsupervised deep learning approach with respect to artificial neural networks there are two main categories of unsupervised leaning approaches they are competitive leaning in which the resulting network is positioned to mainly clustering or classification and the second one is Hebbian leaning in which the resulting networks are positioned to measure familiarity or to project the input data onto the principal components.
- There are various algorithms employed for unsupervised learning they are K-Means, C-Means for conducting clustering analysis and algorithms such as Apriori and Association rule mining for dealing with problems based on association.
- The potential benefits of unsupervised deep leaning algorithms are as follows:

1. It is widely used as a recommended system in various ecommerce websites like Amazon, flipkart and many others. Other Applications are in credit card fraud detection, detection of Anomaly which can lead to the discovery of unusual patterns in the dataset, while another popular application of unsupervised machine learning algorithm is association mining which is a useful unsupervised mining algorithm for identifying the items of the datasets which have a strong association between them.
2. It is observed and evident that the output achieved through unsupervised machine and deep learning approaches is potential enough for making new business decisions and through employing exploratory data analysis (EDA) it will help us to derive useful insights from it.
3. For pre-processing the datasets in unsupervised deep learning artificial neural networks, the latent variable models are employed for decomposing the dataset into multiple components as well as to reduce the number of features in the dataset as well as to overcome some of the limitations in deep neural networks the winner takes it all strategy can be used.

- The potential limitations of unsupervised deep learning algorithms are as follows:

1. In terms of computational complexity, the unsupervised machine or deep learning approaches are computationally complex.

2. The accuracy of the results achieved is low compared to supervised learning because the machine learning or deep learning model is required to perform the activity without any support.
3. In unsupervised deep neural network, the model is not robust enough to withstand against any failure so if one of the layers fails to function then a whole input layers or clusters of layers are lost.

Q4. Review the Enhancing deep neural networks via multiple kernel learning?

1. Structure Title: The authors (Lauriola, Gallicchio and Aiolli, 2020) of this paper have given a title which is "Enhancing deep neural networks via multiple kernel learning" which according to my perceptive doesn't seem to be an appropriate title because it is not clear. The authors have used the term "Enhancing" deep neural network but have not mentioned whether it is the performance or the capability of deep network. Thus, according to me the title should have been "Using multiple kernel learning methods to enhance the capacity of various deep neural networks".
2. Abstract: In this section the author (Lauriola, Gallicchio and Aiolli, 2020) have first discussed about the significance of multiple kernel learning methods and deep neural networks methods as their utilization and success of using them in various applications. The authors have then described about their proposed approach and discussed about the how the resulting ensemble methodology would be represented for CNN and multi-layer perceptron's in consideration of both with random weights and full trained model. They have also mentioned that when they conducted the experiment on the different datasets under consideration it proved that the proposed approach has potential benefits thereby providing a reflection of the content in the research project work.
3. Introduction: The authors (Lauriola, Gallicchio and Aiolli, 2020) in the introduction section have first discussed about the importance of deep neural network architecture and also discussed that the deep learning models perform well for dealing with large scale problems and they have emphasized their concern towards the capability of the last hidden representation because there is a possibility that there is no optimal representation especially when dealing with small and medium sized problems and with that they have discussed the significant importance of how the multiple kernel learning (MKL) methods could be employed for enhancing the capability of deep neural networks which is the underlying intension of the authors of the project research work. Next, the authors have also discussed about proposing an ensemble method by the name of "KerNET" for optimally combining the hidden layer representation using the MKL methodology and for this purpose they have discussed to perform the experiment on two of the deep learning neural network architectures which are Multi-Layer Perceptron (MLP) and Convolutional Neural Network (CNN). The authors have cited 16 research papers in this section in order to explain and support their work with respect to the intension and the proposed methodology and with that they have stated the main intension of this research project which is to understand and observe the benefits of the proposed methodology which is (KerNET) framework when used in synergy with different deep learning neural networks like CNN, MLP and Random Weighted Neural Network (RWNN) which are further assessed using medium sized datasets. Lastly in the introduction section the authors have concluded their introduction on the research topic and have organized the remaining part

of the paper in to five sections with their respective subsections for easy reading and understanding of the research work.

4. Graphical abstracts and/ or highlights: The authors (Lauriola, Gallicchio and Aiolli, 2020) of this paper have not given any summary, visual representation neither they have tabulated their final core findings as a separate dedicated section either presented at the first page of appearance like a rectangular area in the middle of the paper or at the last page of the paper they could have created a separate section called as "Highlights" and mention the key points of the research project work right after the "declaration of competing interest" and "Acknowledgement" section which would be beneficial to the intended audience in order to make notes and to give an eye catcher noteworthy summary of the ability of different machine learning kernel methods to enhance the capacity of the different deep learning neural networks.
5. Methodology: The authors (Lauriola, Gallicchio and Aiolli, 2020) in their paper have very well documented their research paper by dividing it into sections and subsections and not only that have given a small precise note of what each sub section will explain which is a very good practise they have followed which would help the intended audience to read as well as to make notes regarding the research work effectively and efficiently. Apart from the other three sections the two most important sections are section three and section four which gives an in-depth explanation about how the experiment will be conducted, evaluated for the proposed model. Starting with the third section the authors have explained the architectural framework of the proposed model which is the KerNET framework and have created two sub section (Section 3.1) which explains how the instances of the KerNET framework is used with three different feed forward based deep learning neural network architecture which are MLP, RWNN and CNN as well as the process of extracting the intermediate representation; while in (Section 3.2) they authors have well explained a mechanism involved in combining the hidden layer representation. For easy understanding and for the purpose of explanation the authors have created three illustrative representation of the process which are (Fig. 1), (Fig. 2) and (Fig. 3). When we consider the fourth section the authors have further divided it into five sub-sections. The first sub-section (Section 4.1) gives us information regarding the datasets used which tells us from where the data is collected. (Section 4.2)- In depth knowledge about the details of implementation, validation procedures and network configuration. (Section 4.3)- Information regarding performing the experiment under an empirical comparative perception. (Section 4.4) - results produced and (Section 4.5) - analysis of the contribution of each layer. Thus, all this valuable information provided by the authors is enough to replicate the research work since the authors have answered to all the queries that would arise from the reviewer's information pack regarding the methodology section.
6. Results: The results obtained from the analysis by the authors (Lauriola, Gallicchio and Aiolli, 2020) are portrayed in the subsection titled "Results" of the main section which is "Experimental assessment" which is a good practice as it followed by many researchers and it is evident in most of the research papers which will help the intended audience to understand what are the findings from and observations from the research work. The authors have clearly laid down their findings in a logical sequence by creating four individual tables based on the average score of the test accuracy earned, standard deviation of the three deep learning neural network architecture which are MLP, RWNN, CNN and when applied on the 6 different datasets having different characteristics. They found that

in all aspect of consideration the highest test accuracy is achieved by the proposed method which is the KerNET framework i.e. the KerNET framework outperformed the other individual deep neural network architectures framework and also it has the ability to learn a final representation which reflects the complexity of the problem under consideration.

7. Conclusion/Discussion: After applying the proposed methodology to three feed forward deep learning neural network based models and performing critical analysis, assessment and based on the results achieved the authors (Lauriola, Gallicchio and Aiolli, 2020) have justified that the proposed ensemble framework "KerNET" is deemed suitable for enhancing the capability such as the accuracy score, robustness against overfitting of the deep neural network models and also their critical research project work has added value to the overall body of knowledge as well as added significant knowledge to domain specific application. The authors have not indicated the results which are related to the earlier research work but they have discussed that the proposed approach is suitable to work with any deep learning neural network model which is the expectation and as part of their future roadmap they would like to conduct it on large scale applications as well as investigate its implication on Deep Recurrent Neural Network.
8. Language: The authors (Lauriola, Gallicchio and Aiolli, 2020) have made use of simple English language with appropriate grammar consideration in different sections and sub sections of the paper with a thought process that the intended audience would be able to understand the research work clearly. But it has been observed that they have also used jargons and synonyms in their sentences which according to my perspective sometimes may be difficult for the intended audience to understand at first and due to this reason may lose focus on the research work if the word or the sentence they have not followed is not searched for on google or any other respective medium. Apart from the use of English language the authors have also visually expressed their research work on the enhancement of deep neural networks using multiple kernel learning methods in the form of figures like process flow diagram, line graph and bar charts representation which gives a good explanation of the different procedures or process flow involved in the enhancement of the capacity of the deep neural networks using multiple kernel learning methods.
9. Previous Research: The authors (Lauriola, Gallicchio and Aiolli, 2020) have well documented their research paper into sections and subsections for the better understanding of the intended audience and have created a dedicated section called 'Related Work' which forms a part as a subsection of section 2 .This good practice is observed and evident in most of the research papers which has been followed by these researchers. The authors have correctly cited 41 research papers as per the IEEE format to support their research work on enhancing the capability of deep neural networks via multiple kernel learning methods which is observed in different sections as well as subsections of the research paper. Out of 41 research papers they have reserved 12 research papers numbered from 21 to 32 for related work sections which are the works of previous researchers who have mainly emphasized their research only on the strong bond or synergy between the different kernel methods and deep neural networks. The authors have not only restricted themselves to the selection of papers regarding the main topic of research work but have also included other researcher papers which are related to the main topic and domain area of research work and have made sure that whatever is written in each section and subsection has a proof of evidence to it in the form of giving reference to the piece of work from where they have been referred. Apart from research papers the authors have also cited 3 links and one note

information about the different resources used in the research work like website information from where the datasets are retrieved, information to the 'Keras' library which is used in the project work from its original website, information about the system which is used for conducting the research project work and a GitHub link to the Multiple Kernel Learning (MKL) package used in the project work all of this they have cited in the footnote section which is an appropriate and a correct way of citing the references which are not research papers. The authors have also declared of no known competing interest that could have influenced the research work and also the author Ivano Lauriola have expressed his gratitude by creating a separate section known as 'Acknowledgements' for the support provided to him for his contribution towards the research work.

The references that are used for this terminal assessment assignment are as follows:

1. Dimensionality reduction techniques
   - https://www.analyticsvidhya.com/blog/2018/08/dimensionality-reduction-techniques-python/
   - (Cse, 2018)- Cse, H. O. D. (2018) 'A Review on Dimension Reduction Techniques in Data Mining', 9(1), pp. 7–14.
2. Feature Engineering
   - https://machinelearningmastery.com/discover-feature-engineering-how-to-engineer-features-and-how-to-get-good-at-it/
   - Feature engineering for machine learning – book
     Zheng, A. and Casari, A. (2018) *Feature engineering for machine learning, O'Reilly Media*. doi: 10.13140/RG.2.1.3564.3367.
3. Dataset Size
   - https://machinelearningmastery.com/much-training-data-required-machine-learning/
   - https://www.altexsoft.com/blog/datascience/preparing-your-dataset-for-machine-learning-8-basic-techniques-that-make-your-data-better/
4. Computational Cost
   - The Elements of Statistical Learning Data Mining,Inference,and Prediction Second Edition -Book
     Hastie, T., Tibshirani, R. and Friedman, J., 2009. *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media.
   - https://www.youtube.com/watch?v=YoZPTyGL2IQ
   - https://medium.com/@augustinechang/computational-complexity-theory-in-ml-modeling-66b5a5fa610f
5. Question No: 2 Xu, X. *et al.* (2019) 'Review of classical dimensionality reduction and sample selection methods for large-scale data processing', *Neurocomputing*. Elsevier B.V., 328, pp. 5–15. doi: 10.1016/j.neucom.2018.02.100.
6. Ensemble Approaches
   - https://towardsdatascience.com/simple-guide-for-ensemble-learning-methods-d87cc68705a2
   - https://medium.com/@aravanshad/ensemble-methods-95533944783f
   - https://www.youtube.com/watch?v=kho6oANGu_A

- https://www.youtube.com/watch?v=KIOeZ5cFZ50&t=3s
- https://www.youtube.com/watch?v=nxFG5xdpDto

7. Unsupervised Deep Learning
   - https://www.guru99.com/unsupervised-machine-learning.html
   - https://www.datasciencecentral.com/profiles/blogs/unsupervised-learning-an-angle-for-unlabelled-data-world
   - http://intellspot.com/unsupervised-vs-supervised-learning/
8. Question 4: Lauriola, I., Gallicchio, C. and Aiolli, F. (2020) 'Enhancing deep neural networks via multiple kernel learning', *Pattern Recognition*. Elsevier Ltd, 101, p. 107194. doi: 10.1016/j.patcog.2020.107194.