Generic auxiliary tasks to learn features that help regularize the network using Auto-encoder

Vinoth Vincent Raj, 1802566
University of Essex

***Abstract*—Autoencoders play a basic role in supervised and unsupervised learning and in deep learning architectures for transfer learning and one of a kind tasks. For two main reasons, the reduction of dimensionality has been an important subject of analysis in academia. Firstly, large volume of data resulting in computing costs and becomes difficult to handle. Also, mapping from higher dimensional to lower dimensional data without losing much information removes redundant dimensions in our dataset. In this paper, an approach of how autoencoders will be used for three different dataset which were collected from different domains will be discussed along with data description and experiments that will be performed.**

***Index Terms*—Autoencoders, Data Exploration, Neural Network, Stacking.**

I. Introduction

Dimensionality curse is a common engineering problem. Ineffective and unreasonable reduction of the dimensionality of the variable jeopardizes the efficiency of machine learning, the accuracy of pattern recognition and the efficiency of data mining while increasing the workload of measured data experiments to some extent. Its dimensionality must be reduced in order to handle such real - world data adequately. Reduction of dimensionality is the transformation of high - dimensional data into a significant representation of low dimensions. In particular, high - dimensional representations are generated when signals, processes, images or physical fields are sampled.

Mostly dimensionality reduction techniques has many advantages. Firstly, Visualization: For the projection of high-dimensional data onto 2D or 3D. Secondly, Compression of data: Effective storage and recovery. Lastly, Removal of noise: Positive effect on accuracy of query.The rationale for dimensional reduction can be defined as follows from a knowledge discovery perspective, the identification of a reduced set of characteristics that predict results can be very useful. For several learning algorithms, the number of features directly increases the training and/or classification time.Noisy or meaningless features may have the same impact on classification as predictive features, thus negatively affecting accuracy.

In this paper, three different datasets from different domains have been selected. The three different domains are banking, government and medicine.

In banking data, given numerous features it requires to predict the customer satisfaction. Dramatically increasing apparent customer satisfaction or decreasing dissatisfaction has been recognized as the number one management objective in the banking industry. It is vital to know about unhappy customers to create competitive leadership, especially in the after - sales sector. Therefore, before the service interaction ends, the identification of potentially unhappy customers is of great value to enable proactive actions before the customer is actually unhappy. A completely satisfied consumer is retained and therefore from a business point of view is of great interest.

In the census data, the requirement is to correctly classify peoples income above or below a fixed threshold value based on numerous features like education, age, marriage status and location In recent years, the issue of income inequality has been of great concern. Better treatment of the destitute does not seem to be the only requirements for eliminating this problem. People in america sincerely believe that economic inequality is inexcusable and requires a good share of society's wealth. Governments in different parts of the world use different treatments to tackle economic inequality, although some success. Perhaps one of the main reasons for catastrophe is to do many things that reduce productivity and result in reduced results.

In medicine, ML techniques have been popularly used in intelligent health systems in recent decades, especially in the treatment and prognosis of breast cancer . Generally, a patient's diagnostic reliability depends on the experience of a doctor The erudite healthcare system can help doctors diagnose patients more accurately or even provide more tangible baselines and help people plan their physical condition in the future. BC has always had a high incidence rate and mortality rate as the most common cancer in women. BC alone is expected to account for 25% of all new cancer diagnoses and 15% of all cancer deaths among women worldwide, according to the latest cancer statistics [5] [6]. A precise classification can also help clinicians to prescribe the most appropriate treatment scheme. Classification is a complicated problem of optimization. Researchers have used many ML techniques to solve this classification problem.

II. Background

The autoencoder algorithm [2] is part of a special family of methods for reducing dimensionality using artificial neural networks. An autoencoder can be represented as a neural network whose primary function is to learn the underlying collector or the data set's feature space. An autoencoder tries to recreate the output inputs. It strives to try and learn a reduced structure of an input by reducing its error in reconstruction. The autoencoder algorithm and its extensions [6] [10] [1] recently demonstrated a successful ability to learn useful data features that could lead to an "intrinsic data structure".

An autoencoder learns to squash data into a short code from the input neural network layer and then uncompress

the data into something that closely represents the raw data. This enables the autoencoder to reduce dimensionality by trying to learn to ignore noise, for example.The encoding layer encodes the entire image into a matching code. The decoding layers will start to learn to decode the depiction of the learned code as closely as possible in its original form. More recently, autoencoders have once again taken center stage in the "deep architecture" approach,in which autoencoders, especially in the form of restricted Boltzmann machines, are stacked and trained in an unsupervised manner, followed by a supervised algorithm to train the encoded layer. Forecasting is an effective method to predicting bank clients ' level of customer satisfaction. The models documented in the research literature for bank client satisfaction are labeled into two main groups. Statistical models such as structural equation models and regression models are included [7][8][9][10] in the first category. There are some examples in this category. The main downside for this category of designs is the lack of decent predictive accuracy, now since the models have a linear structure, while customer satisfaction has a highly non-linear trend with the factors that influence them.

Artificial intelligence models like artificial neural networks (ANN) are included in the second group. ANN is an efficient process for modeling the actions of nonlinear structures with a simple system that makes them computationally efficient. At 15:51 2nd August 2015 (PT) telecommunications and healthcare services [12][13] ANN was used to anticipate customer satisfaction in non-banking organizations such as the automotive industry. No study of ANN modeling of bank customer satisfaction has been conducted in the literature. This study develops the first case to the best of the authors ' knowledge.

A recent analysis [14] mentions a generalized autoencoder (GAE), which concentrates on altering autoencoders to take into account the co relation between data by introducing a weighted relational function. In particular, the weighted distances between remodelled and unique instances are minimized. Even though ensuing applications[15][16] verify that the actual data link can enhance autoencoder overall performance in the decrease of dimensionality, the Generalized Autoencoder model seems to have some disadvantages. The current rebirth of interest in autoencoders is due to the advancement of training of deep architectures, since traditional gradient-based optimisation techniques are are are are are not efficient when hidden layers are stacked with non-linearities several times. In 2006, Hinton[17] trained a deep network architecture empirically by sequentially optimizing each layer's weights in a RBM (Restricted Boltzman machine. Bengio succeeded in training a packed autoencoderwith a comparable greedy layer approach on the MNIST dataset. This training approach addresses the issue of non-convex optimization that prevents deep network structure. Later studies show that the stacked autoencoder algorithm can learn substantive elements and therefore perform better in high dimensional data for classification.

Autoencoder is becoming increasingly popular as the training efficiency improves. It was soon found that the weights of neural network layers layers rise steeply due to matrix multiplication as layers are stacked, and then the imports of these weights also become bigger than the previous input characteristics. This overfitting problem leads to the fact that depictions of deep layers are much more likely to rely on the network topology than on the original input functions. Regularization is added in the autoencoder loss function to impose a penalty on large weights was added by Goodfellow et al.[19] Vincent et al.[20] advocated a Denoising Autoencoder (DAE) to resolve this problem by introducing input noises. The allocation of neural network hidden layer representations is not regulated by past studies. Variational Autoencoder[21] is therefore proposed to produce required representation allocations in neural network hidden layers. Autoencoders can be refereed as an unsupervised process to reduce dimensions and the decreased high-level features typically produce the critical insights from the source data.Hence, This does not make autoencoders susceptible to minor changes.

Contractive Autoencoder to make them resilient to minor variations was proposed by Rifai et al. [2]. Many GAE applications [23] prove that the maintenance of data relationships can produce good results, but outcomes depend heavily upon how distance weights are defined and selected. Due to the assignment of high weights to chosen relationships, the technique of robust, pre-defined distance weights is very arbitrary and can be biased in converting GAE into a supervised model. We therefore advocate a Relation Autoencoder (RAE) to reduce dimensionality by minimizing both the total loss of data characteristics.

III. Methodology

For this research purpose, three datasets have been chosen as discussed already. This section will explain the about the data in brief and also explore the data with the intuitive data analysis. The major results of the analysis will also be discussed in this section.

A. Census data

The information for our research were accessed from the Machine Learning Repository of the University of California Irvine (UCI). Barry Becker actually extracted it from the 1994 census database. The data set contains data from 48,842 unique records and 14 features for 42 countries. There are 14 features consisting of 6 continuous and 8 categorical features containing information on age, education, nationality, family status, employment, employment, gender, race, working hours per week, capital loss and capital gain, as shown in Table 1. The binary label in the data set is the level of income of that predicts whether or not a person earns more than 50 thousand dollars per year on the basis of the given set of features. Numerical column values have no missing values in the dataset. The scales of each column are different. Hence it is advised to do a data normalization before working on the dataset.

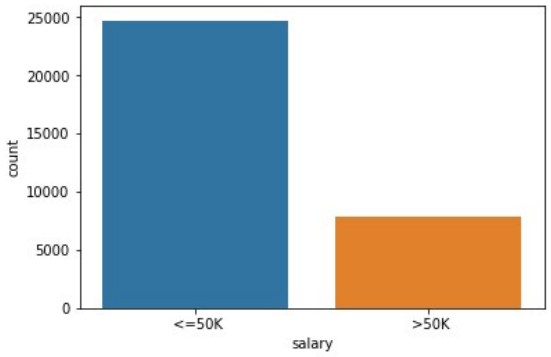

Fig. 1. Count of Salary feature which is the target to predict for the adult census data.

B. Breast Cancer

Machine learning algorithms have been trained to recognize breast cancer by using Wisconsin Diagnostic Breast Cancer data (WDBC) [26]. The data consists of variables calculated from a digitalised picture of a fine needle aspirate (FNA) of a breast mass, according to[26].

The data source contains 569 data points with 31 features: 357-Benign and 212-Malignant, Benign instances are regarded as non - cancerous, i.e. non - life intimidating. However ,it could become a cancer status on a couple of occasions. An immune system termed ”sac” ordinarily separates benign tumors from many other cells and can be easily removed from the body.Malignant cancer begins with an unusual cell death and can actually spread or invade adjacent tissue quickly. The nuclei of malignant cartilage are usually much larger than in normal cells, which in possible future phases can be life - threatening.

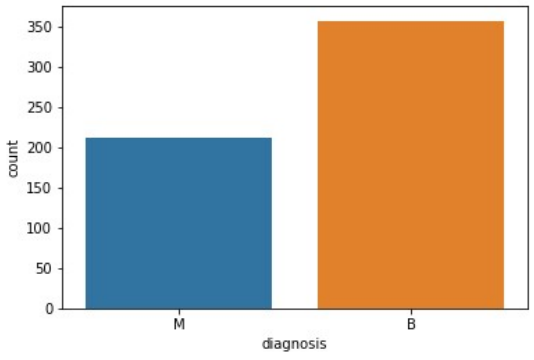

Fig. 2. Diagnosis values count for Malignant and Benign.

Some of the key features are radius, texture, compactness, concavity, perimeter, symmetry, smoothness, concave points, and fractal dimension. Each feature has three relevant information[26]: (1) average, (2) standard error, and (3) ”worst” or largest (mean of error).

C. Santander Customer Satisfaction

In this study, customer satisfaction prediction is the objective in of Santander Bank, a large company focuses primarily on the northeastern United States market Through a Kaggle competition (Santander, 2015), the aim is to find a suitable model to predict whether a certain customer will be unhappy with other qualities in the future. The introduction of this model can ensure that Santander is assertive.

The dataset consists of 371 anonymous columns and 76020 instances. The target variable of whether a customer is satisfied defined as a binary outcome in target column. The target is highly imbalanced as shown in the figure with 96% belonging to one class and the rest in other class. Based on some initial analysis we could find that unhappy customers have less products with the bank. Since the variables are anonymous it will be useful to plot feature importance from any model to get the variables contributing more to the target variable. Usually, in banking terms, it may be mortage value of the customer in the bank or the value of the customer The target

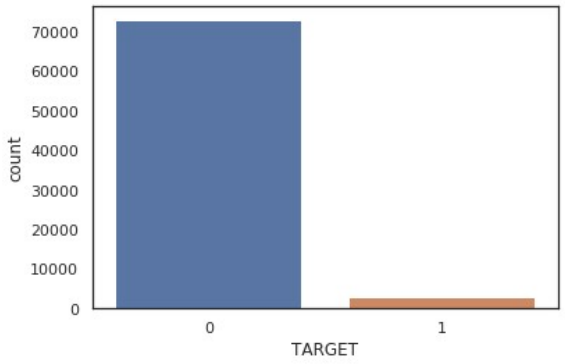

Fig. 3. Customer satisfaction count for santander data

variable is highly imbalanced as we could see in the figure above. Introducing SMOTE technique to do a oversampling could help the data to sample the target values equally which would be helpful for the machine learning algorithm to learn.

D. Data Pre-processing methods

Data pre-processing techniques includes data preparation, and data transformation into a suitable form for mining purposes. The goal of data pre-processing is to reduce the dimensions of the data, clean the data, find patterns in the data, normalize data and remove outliers. Since we are dealing with three sets of data. It is highly obvious that the features have different scales and each features can be ordinal, categorical and numerical. Also, to process any data it has to be converted to numerical format for the algorithms to access. So some of the pre-processing methods which will be done for the three data sets are summarized in stages.

Data Cleaning: The instances in the data might have incomplete values, noise and outliers. The first step in data pre-processing is to correct the inconsistent data as these inconsistent data might affect the performance of our algorithm.

Therefore, it is advised to clean the data before passing it to algorithm

Converting categorical (text) values into numerical values : Many of the variables are categorical in all three data sets which are numeric. The categorical variables are converted into numeric values.

Removing unrelated columns: The variable which is not contributing to the target can be removed using various feature selection techniques

Handling null values: After doing all the above data cleaning steps, separates data frames for the feature and target data are generated.

Plot correlation matrix: The next technique which could be implemented is the correlation plot across all the variables. This would eventually show us the correlation between each variable and the target variable also the correlation between independent variables.

Label encoding: Since our dataset contains text as values in some of the rows. Especially in breast cancer dataset the target value has two categories in text. Label encoder encodes the values between 0 to 1 for two classes and it depends on the number of classes a feature has. The drawback of label encoding is that it gives weight for higher numbers and may derive false assumptions.

One Hot encoder: Census data has countries column which has to be one hot encoded. It takes a column with categorical data that has been encoded by the target and then divides the column into several columns. Depending on which column has what value, the numbers are replaced by 1s and 0. We'll get four new columns in our example, one for each country-Japan, the United States, UK, Greece and China

Normalization: Normalization is the adjustment of values calculated on different scales, often before averaging, to a common scale. Normalization can refer to more robust modifications in more complex cases in which the motive is to align the entire probability distribution of normalized values

Multicollinaearity: Multicollinearity happens in our data when variables rely heavily on each other. Multicollinearity occurs when your model contains several variables that are not only correlated with your target variable, but also with each other. In simple words, if you have factors that are a little redundant, it results. So columns with high correlation between each other in independent variables will be removed as it wont contribute much to the model and might increase the comlication of the model.

IV. Experiments

For all the three datasets, an individual autoencoders will be trained first to find the best parameter for the autoencoder algorithm to replicate out input data. Then the encoder part which is the compressed learned features of all the input will stacked with a neural network to predict the classes of the target. But this steps will be experimented by taking smaller to all data points in the dataset. Their performace will be measured for different sample proportions and the autoencoder bottleneck neuron counts.The following sppraches will be taken. Firstly training a autoencoder and passing the well trained auto encoder as input to an artificial neural network and predict the target variable.

A. Training Auto encoder

An autoencoder is a neural network that learns efficient data coding by replicating the structure of its input. The objective of an autoencoder is to try to learn to comprise a set of data, typically to reduce dimensionality, by training the network to ignore the signal of "noise."

A greedy layer-wise training is a great way to get a stacked autoencoder decent parameters. Train the first layer of unprocessed input to obtain parameters W (1,1), W (1,2), b (1,1), b (1,2). To turn the raw input into a vector that activates the hidden units, use the first layer. Train the next layer of this vector for parameters W (2,1), W (2,2), b (2,1), b (2,2). Repeat for consequent layers using each layer's output as an input for the further layer. This technique trains each layer's parameters separately while making the remaining model parameters constant. To produce good outcomes, fine-tuning u after this training phase has been completed. This

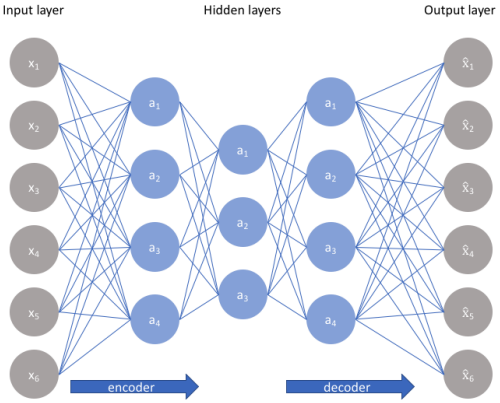

Fig. 4. Architecture of a deep autoencoder.

type of network consists of two parts:

Encoder: Encoder part of the network deforms the input into a depiction of latent space. The encoding functionality h= f (x) can represent it.

Decoder: Decoder segment is intended to recreate the input from the real representation of latent space. A decoding function r= g (h) can represent it.

Autoencoders are expected to keep as many information as possible whenever an input runs through the encoder and then the decoder, but they are still trained to have different nice properties in the new display. It is a feed-forward neural network that aims to minimise the loss function from the output which is the difference between the actual input and the output Once the loss is minimised it means that once autoencoder has been trained.This implies that once properly trained autoencoders are quite accurate and it will be hard to generalize data sets other than those for which they have been trained

B. Stacked Auto-Encoder

Once the autoencoder has been trained to replicate the exact data with minimum loss. The encoder pat can be passed as an input to the neural network. This functionality is basically a dimensionality reduction technique in which the encoder compress all the information into bottle neck neurons. These neurons have the compressed input information for all the data. This neurons will now be passed into a traditional artificial neural network to predict their targets as shown in the figure below.

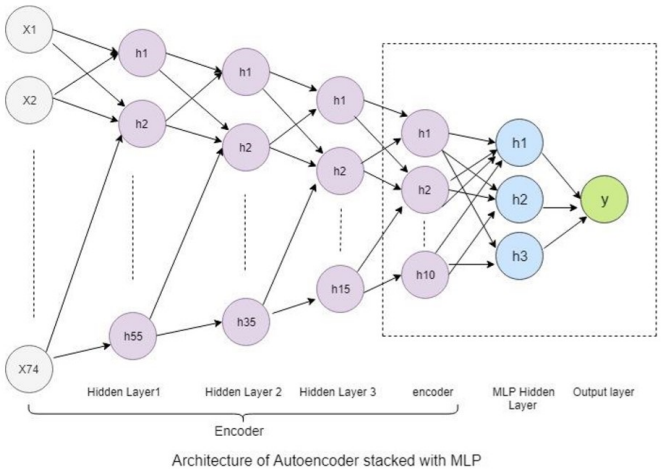

Fig. 5. Example architecture of a Stacked autoencoder with a deep neural network.

As shown in the the figure above, the hidden layers are trained by an unsupervised algorithm and then adjusted by a supervised method. Stacked autoencoder largely consists of three steps. Use raw data to train and obtain traineddata. The data trained from the previous layer is used as an input for another layer until the training has been completed. To minimize loss and update weights with the training set, use the back propagation algorithm to achieve fine tuning once all hidden layers are trained. The recent advancements in Stacked Autoendocer gives a data version with much accurate and detailed and and encouraging variable information that is used to train a algorithm in a specific context and find decent accuracy than raw data training.

C. Compare results with standard algorithm

Once obtained the results using a stacked autoencoder on different subsamples we will be comparing all the results with a standard algorithm and do a comparative study whether this method is able to compete some of the past accuracy obtained by using standard machine learning models.

Three different datasets will be used to run three different standard state of the art machine learning models and their performance with the stacked auto encoder will be compared. For example, In breast cancer data, SVM suppose to achieve better results in the past, which will be compared with the dimension reduced model. In census data , we will be comparing the model with Gradient boosting classifier algorithm which was able to achieve 89% accuracy in predicting the income published in [25]. The customer satisfaction data LGBM model will be compared against our stacked autoencoder model and will be compared with the top accuracy obtained in kaggle competitions

V. DISCUSSION

All the three datasets have imbalanced dataset so measuring the accuracy alone would not help us to access the true performance of the models. Metrics such as accuracy, accuracy, recall, F1 and Gini coefficient. F1-score is a widely used measure of the accuracy of a test, since we are dealing with classification problem. To calculate the score, it embodies both the accuracy and the recall of the test. Precision is the number of real positive factors divided by the sum of true positive and false positive factors. Similarly, the number of true positive elements divided by the sum of true positive and false negative elements, which is the total number of elements belonging to the positive class. Gini coefficient is used in problems of classification derived directly from the AUC value. It is only a ratio between the area between the ROC curve and the diagonal line and the area of the triangle

Confusion Matrix: The matrix of confusion is a NxN matrix in which the actual number of predi ctorlabels is N. The values of positive and negative and and parts and correctly recognized cases lead to accuracy.

	P' (Predicted)	N' (Predicted)
P (Actual)	True Positive	False Negative
N (Actual)	False Positive	True Negative

Fig. 6. Confusion matrix.

True Positives (TP): True positives are situations in which the real data class was True and the prediction is True Ex: The scenario in which a person has diabetes (1) the model categorizing his case as diabetes (1) is True positive.

True Negatives (TN): True negatives are situations in which the real data point class was False and the predicted is also False

False positives (FP): If the the the real data class was False and and and the prediction is 1 (True) The wrong thing is that the model was wrongly and positively predicted because the predicted class was positive.

False negatives (FN): If the the real data point class is True and and and and the prediction is False, there are situations of

false negatives. The worst case isis that the model incorrectly and negatively predicted because while the class predicted was negative.

Precision : The number of correct positive outcomes is neatly divided by the number of positive results the classifier predicts.

$$Precision = \frac{tp}{tp + fp}$$

Recall : The number of correct positive outcomes is split by the number of all relevant samples (all samples to be identified as positive).

$$Recall = \frac{tp}{tp + fn}$$

Classification Accuracy: When we use the phrase accuracy, we generally mean accuracy. It is the ratio of the right number of forecasts to the total number of input datasets. $Accuracy = \frac{tp + tn}{tp + tn + fp + fn}$

F1 Score: Harmonic mean between precision and recall is F1 score. The range for F1 Score is[0, 1] High accuracy but lower recall, which gives you an extremely accurate range, but then misses a large number of hard to classify instances. The higher the F1 score, the better our model's performance. It can be expressed in mathematical terms as:

$$F = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

ROC-AUC Curve: ROC is a curve of probability and AUC is a degree of separability or measurement. It shows how much model can differentiate between classes. Higher the AUC, the model is better. ROC is plotted with True positive rate on y-axis against the false positive rate on x-axis.The best rule of decision is sensitivity and 1-specificity low. It is a guideline that most true positives are predicted to be positive and few true negatives are expected to be positive.

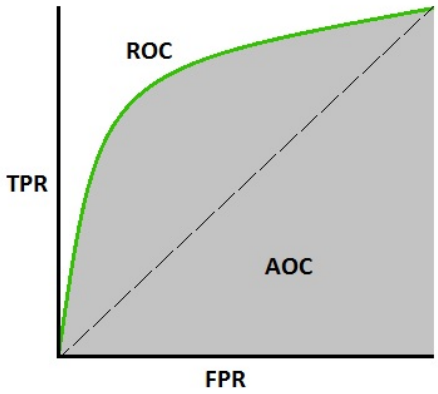

Fig. 7. AUC-ROC Curve.

VI. CONCLUSION

In this paper, I detailed the steps and algorithms to try out the proposed methods which by learning from features like dimensionality reduction technique and regularize the networks. A comparative report can also be generated to compare the performance of the proposed model with other state of the art model and their performances can be competed and evaluated using the evaluation metrics discussed. A further analysis can also be made by fine tuning the parameters of Autoencoder and the machine learning model to find its optimal parameter to report a meaningful insights in comparing the algorithms.

REFERENCES

[1] S. Rifai, P. Vincent, X. Muller, X. Glorot, and Y. Bengio., *Contractive auto-encoders: Explicit invariance during feature extraction. International Conference on Machine Learning*, 2011. 2, 3

[2] D. Rumelhart, G. Hinton, and R. Williams. *Learning internal representations by error propagation. Parallel Distributed Processing.* Vol 1: Foundations. MIT Press, Cambridge, MA, 1986. 2

[3] H. Lee, C. Ekanadham, and A. Ng. *Sparse deep belief net model for visual area v2. Advances in Neural Information Processing Systems*, 2008. 2

[4] G.E. Hinton and R.R. Salakhutdinov. *Reducing the dimensionality of data with neural networks.* Science, 313(5786):504, 2006.

[5] Chen, W.; Zheng, R.; Baade, P.D.; Zhang, S.; Zeng, H.; Bray, F.; Jemal, A.; Yu, X.Q.; He, J. *Cancer statistics in China, 2015*. CA Cancer J. Clin. 2016, 66, 115132. [CrossRef] [PubMed]

[6] Torre, L.A.; Bray, F.; Siegel, R.L.; Ferlay, J.; Lortet-Tieulent, J.; Jemal, A. *Global cancer statistics,* 2012. CA Cancer J. Clin. 2015, 65, 87108. [CrossRef] [PubMed]

[7] W. H. Cantrell, and W. A. Davis, "Amplitude modulator utilizing a high-Q class-E DC-DC converter", *2003 IEEE MTT-S Int. Microwave Symp. Dig.*, vol. 3, pp. 1721-1724, June 2003.

[8] Kaura, V. (2013), *Antecedents of Customer Satisfaction: A Study of Indian Public and Private Sector Banks, International Journal of Bank Marketing,* Vol. 31, No. 3, pp.167186.

[9] Ibok, I.N., and John, A.S. (2013), *Investigating Customer Satisfaction Driven Values in the Retail Banking Industry, International Journal of Finance and Accounting,* Vol. 2, No. 5, pp. 292-296.

[10] Singh, J., and Kaur, G. (2011), *Customer Satisfaction and Universal Banks: An Empirical Study, International Journal of Commerce and Management,* Vol. 21, No. 4, pp. 327-348.

[11] Fatima, J. and Razzaque, M.A. (2010a), *Service Quality, Customer Involvement and Customer Satisfaction: A Case Study of Retail Banking in Bangladesh,* Journal of Business and Policy Research, Vol. 7, No. 2, pp. 135146.

[12] Cui, Q. A., Wang, X., Li, H. J., and Kang, X. (2011), *Using PCA and ANN to Identify Significant Factors and Modeling Customer Satisfaction for Complex Service Processes*, in International Conference on Industrial Engineering and Engineering Management, China, 2011, IEEE, pp. 1800-1804.

[13] Goode, M.M.H., Davies, F., Moutinho, L., and Jamal, A. (2005), *Determining Customer Satisfaction from Mobile Phones: A Neural Network Approach,* Journal of Marketing Management, Vol. 21, No. 7-8, pp. 755-778.

[14] W. Wang, Y. Huang, Y. Wang, and L. Wang, *Generalized autoencoder: a neural network framework for dimensionality reduction,* in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops,2014, pp. 490497.

[15] Z. Camlica, H. Tizhoosh, and F. Khalvati, *Autoencoding the retrieval relevance of medical images,* in Image Processing Theory, Tools and Applications (IPTA), 2015 International Conference on. IEEE, 2015, pp. 550555

[16] L. Meng, S. Ding, and Y. Xue, *Research on denoising sparse autoencoder,* International Journal of Machine Learning and Cybernetics, pp. 111, 2016.

[17] G. E. Hinton and R. R. Salakhutdinov, *Reducing the dimensionality of data with neural networks,* Science, vol. 313, no. 5786, pp. 504507, 2006.

[18] H. Larochelle, D. Erhan, A. Courville, J. Bergstra, and Y. Bengio, *An empirical evaluation of deep architectures on problems with many factors of variation,* in Proceedings of the 24th international conference on Machine learning. ACM, 2007, pp. 473480.
[19] I. Goodfellow, H. Lee, Q. V. Le, A. Saxe, and A. Y. Ng,*Measuring invariances in deep networks, in Advances in neural information processing systems*, 2009, pp. 646654.
[20] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.- A. Manzagol, *Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion, The Journal of Machine Learning Research,* vol. 11, pp. 33713408, 2010.
[21] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, *Generative adversarial nets, in Advances in Neural Information Processing Systems*, 2014, pp. 26722680.
[22] S. Rifai, P. Vincent, X. Muller, X. Glorot, and Y. Bengio, *Contractive auto-encoders: Explicit invariance during feature extraction,* in Proceedings of the 28th international conference on machine learning (ICML-11), 2011, pp. 833840.
[23] Z. Camlica, H. Tizhoosh, and F. Khalvati, *Autoencoding the retrieval relevance of medical images, in Image Processing Theory,* Tools and Applications (IPTA), 2015 International Conference on. IEEE, 2015, pp. 550555.
[24] H.-C. Shin, M. R. Orton, D. J. Collins, S. J. Doran, and M. O. Leach, *Stacked autoencoders for unsupervised feature learning and multiple organ detection in a pilot study using 4d patient data,* IEEE transactions on pattern analysis and machine intelligence, vol. 35, no. 8,pp. 19301943, 2013.
[25] Navoneel Chakrabarty, Sanket Biswas, *A Statistical Approach to Adult Census Income Level Prediction*
[26] William H Wolberg, W Nick Street, and Olvi L Mangasarian. 1992. *Breast cancer Wisconsin (diagnostic) data set.* UCI Machine Learning Repository [http://archive. ics. uci. edu/ml/] (1992).

VII. PLAN

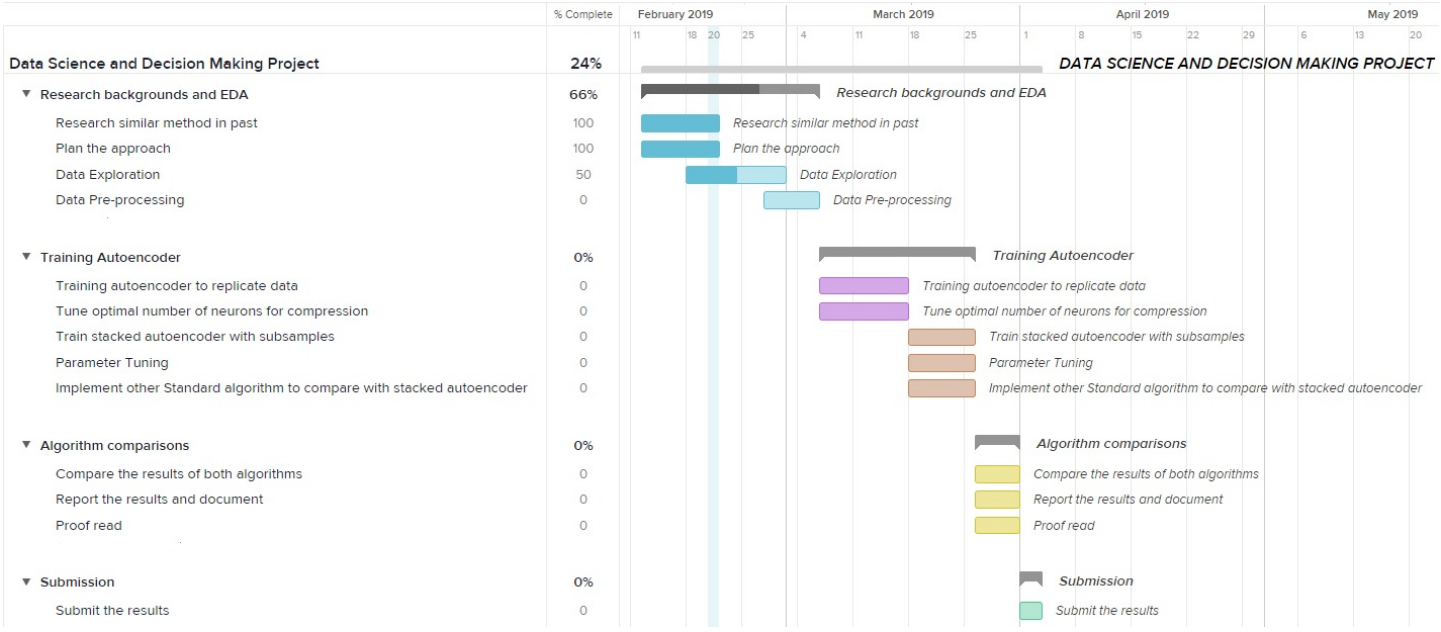

Fig. 8. Proposed timeline to complete the project

VIII. CODE REPOSITORY

I have placed the basic data analysis code in the github repository mentioned below https://github.com/vincentkr18/ce888labs/tree/master/Project$_1$