INM427 Neural Computing – Research Paper

Vivek Kanna Jayaprakash
MSc Data Science
Vivek.Jayaprakash@city.ac.uk

I. Brief Description and Motivation of the Problem

Description: Weather forecasting is one of the earliest known predictions which were in practise known to mankind and it still in use in the current world and has grown leaps and bounds through the means of various technology advancement to make the prediction accurate and reliable. In the current world, AI and Neural networks are vastly used In the prediction of weather. In this case we try to identify the rain for tomorrow based on certain attributes of Temperature, pressure humidity, Wind gust speed, wind speed, pressure rainfall collected and the rainfall of today.

Motivation: Prediction of rain is necessary for each and every being on the face of the world and it is important in every walk and profession due to the immediate effect it can have on situations. Even though there are high quality metrological departments present in every part of the world there is a lack of details recorded in many small low per capita countries which only records the basic attributes of temperature, pressure, humidity, wind speed etc and fail to record various attributes like evaporation, wind direction, sunshine time with direction, cloud positioning etc. This is mainly due to the lack of due to the lack of equipment needed for survey. So we are using this model to predict the rain for tomorrow using 8 basic attributes which are very commonly recorded worldwide. This will serve as a tool for the prediction of rain using Neural networks across anywhere on earth.

II. Description of the Dataset Including Datatypes

The dataset is taken from the Kaggle.com weather in Australia dataset contributed by Joe Young and Adam young [2].This particular location was chosen mainly because of the vast different aspects of climate nature of the region. Being an island, Most of the centre of Australia has a dessert or arid climate and the northern part has a beautiful tropical climate and there parts where it get high cold as well. This vast contrasts in the data will suit for our research as to match with the countries we are targeting.

The dataset consists of 145460 rows 22 dependent variable and 1 target variable. As per plan we keep the major features recorded worldwide and drop the rest. Considering that the null values are less that the 10% of the entire dataset we drop the same. The cleaned dataset consists of 119865 rows and 8 dependant variable and 1 target variable. On identifying the attributes on correlation matric we understand that the attributes aren't correlated with each other.

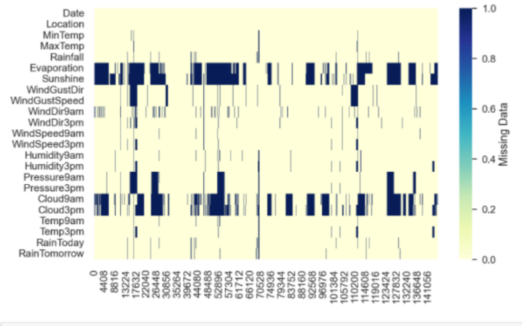

Figure 1: Heatmap for the missing values

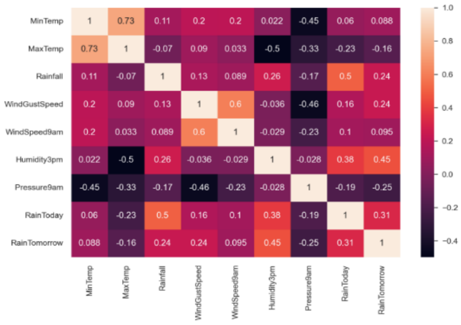

Figure 2: correlation matrix of the cleaned dataset

The count plot demonstrates that Most of the variable are skewed and it is very closely related to each other Some of the variables re expedient and are time related.
Figure 3 demonstrates that there is class imbalance in the dataset. We maintain the same without using max min scalar or smote .

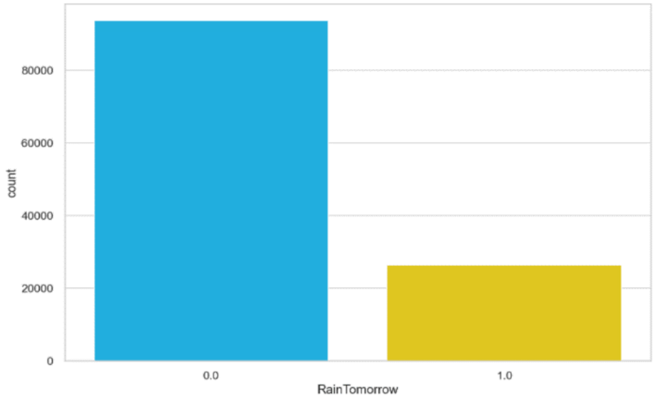

Figure 3: Count plot denoting the balance of the dataset.

According to paper [] scaling a dataset to a balance demonstrates that it is highly likely to prove accurate model but will not perform wisely in a real life scenario and therefore the training data samples should be increased. We therefore work with a imbalanced dataset to check the results. The data is further normalise d for the modelling and thus split into dependant and independent variables.

III. Brief Summay of the neural networks with pros and cons

1.Multi layer perceptron : Multi-layer perceptron is a type of neural network which is a subordinate of the feed forward neural network. It has three of more layers which include a input later, an output layer and hidden layers. The input layer receives the input, and the job is performed for any prediction or classification task by the mentioned output layer. The number of hidden layers depends on the data. It is a forward moving neural network. The neurons which are present in the MLP build by a learning algorithm which is back propagated. They are global approximators which can approximate any function that is continuous.

Pros :

- Can be applied and works well with large Dataset
- Solves and provides quick output and the same accuracy can be achieved with smaller data
- Universal approximators

Cons:

- The black box concept makes the user not know to which extent the dependant variable can affect the independent counterpart.
- The iteration on computation methods take too much time and are difficult
- Since this functions based on the training its quality decided if the training quality is bad.

2.Support vector machines (SVM)

Support vector machines is a supervised machine learning modelling that is mainly used the classification algorithm in many use case of two group classification problems. Once given a labelled training data for each categories the SVM models categorizes new text. Compared to newer algorithms like neural networks, they have two main advantages: which are higher speed and better performance with a limited number of samples (in the thousands). This makes the algorithm very suitable for text classification problems, where it’s common to have access to a dataset of at-most a couple of thousands of tagged samples. They are mainly used in text classification problems as they have a large amount of tagged samples. SVM work on the principle where it take the data given x y coordinates and outputs a hyperplane as two dimensional line which separates the coordinates. This hyperplane is called the decision boundary with which it cateogirs the training.

Pros:

- SVM is very advantageorus & highly effective in high dimensional spaces.
- It is effecicient when it comes to memory
- SVM works wll when there a clear class spseration
- It is Hugely effective if the number of dimensions is greater than the number of samples

Cons :

- When the target class overlap or when the data has more Nosie SVM models tend to under perform.
- As the SVM works on the principle of keeping datapoints above and below decision boundary probabilistic explanation is non existent in this form of classification.
- It is generally not suited for large sets of data
- When the number of features of each data exceed number of train samples, SVM underperforms

IV. Hypothesis statement

Both the models are supposed to perform very well considering the tasks but the SVM being a high sclable classifier is expected to perform better than the MLP model mainly because of the binary data to be classified. The time taken for approximation is supposed to be quicker in SVM and than in MLP because of the local approximation method. Our data set target variable is left as it is with the class imbalance with the due to the reason mentioned above hence the false positives will be more considering the imbalance

Hyper parameter tuning is supposed to help the accuracy but due to the complexity in the data it is expected not to give a high level of accuracy. It is also expected that with simple CPU systems the time taken for the training will be very long for both he model training.

V. Description of Choice of training and evaluation methodology:

Choice of training: We are splitting the dataset into 80 % training and 20 % test and implanting a random seed In the training process to make the training sets a random split. The test set is completely unknown in this process. We train the entire set and we tune the hypermeters separately for every set. Using grid search CV. Attempting to do various hidden layers neurons in the MLP to get the best accuracy using trial and error Drop out regularisation is not used.

Evaluation methods:
we use the methods used in nominal Neural networks projects to we calculate the precision, recall, accuracy and F1 score for the models to build the confusion matrix to find the true positives in our dataset. ROC error curve and AOC score will also be identified.

VI. Choice of parameters and experiments results

In our Multi layer perceptron model we use the pytorch library and extend the neural network model to define out targeted model to accept all the features required and output the target variable. The constructers we build will take in the dependant variable features as the input. Three neural nets are defined for our model with each layer having an input and output feature the input of the second layer is the output of the first layer and so on. The hidden layers are experimented on a trial error basis. RELu activation function is used. On attempting to identify the value of the nodes in the hidden layer we use a BCE loss function which tells the current progress of the model and we use the ADAM optimiser as it is very safe and it serves a good initial starting point for any model. In the sense of hyperparameter tuning we use various parameters like hidden layer, two different activations, learning rate and alpha values and gridSearchCV for doing the tuning.

In the SVM modelling since because of the type of the model, it creates it own path hence the choice of parameters are limited but during out hyperparameter tuning process we using rbf kernel in our approach and build using the gridsearchCV method.

The difference between the train and test set accuracies denote the bias and variance in the dataset which is clearly considered. Hyperparameter tuning in the dataset isn't fully doing its job variations will be discussed below

VII. Analysis and Critical Evaluation

Analysis : Both the type of neural networks performs to build a model for identifying the rain for tomorrow with accuracies matching the literatures. Figure [4], [5] of the confusion matric of both the Multi layer perceptron and the Support vector machine consists of true positive values which helps us understand that our model is very much efficient at prediction where there will be no rain for tomorrow but lags when predicting if there will be actually rain for tomorrow. This is mainly due to the less amount of data we had for the rain for tomorrow when compared to no rain for tomorrow as denoted in the figure.[3] Therey causing the impact point of class imbalance.

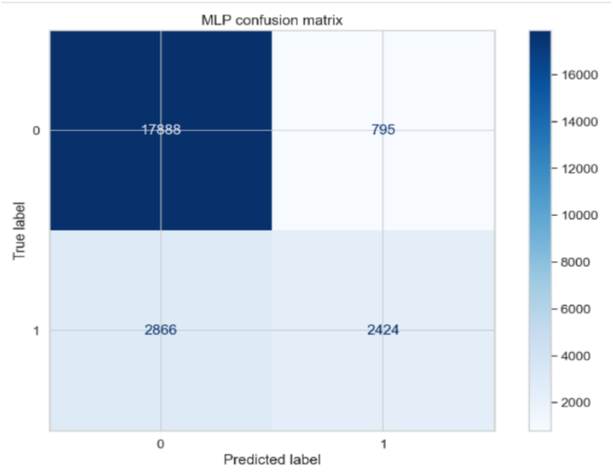

Figure 4: MLP confusion matrix

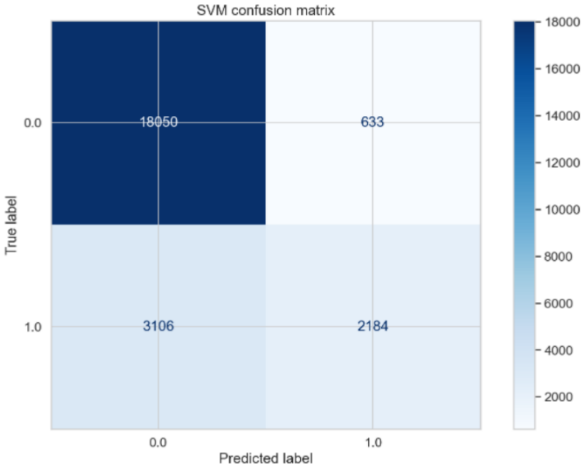

Figure 5 : SVM confusion matrix

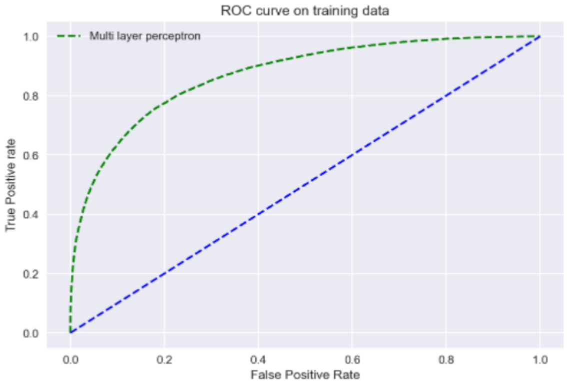

Figure 6 : ROC curve on MLP

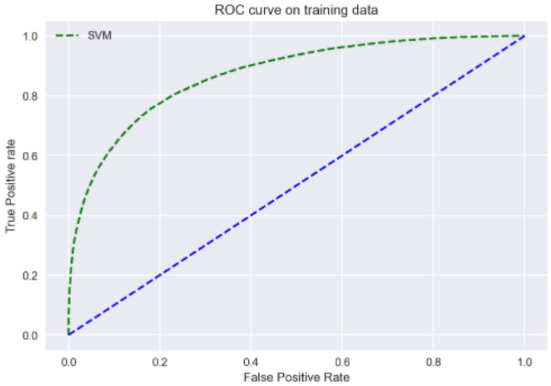

Figure 7: ROC curve on SVM

The findings are similar in both models but the table 1 denotes that since the major target for this project is to identify the rain for tomorrow and that is lightly better for the SVM as based on the recall scores recorded. This is also proven by the higher F1 scores in the target variable. The constast between the testing and testing accuracies are to be noted which tends to suggest that model shows bias and variance

	Precision		Recall		F1 Score		Accuracy	ROC
	No rain	Rain	No rain	Rain	No rain	Rain		
MLP	0.86	0.75	0.96	0.45	0.91	0.56	85%	87.5%
SVM	0.86	0.78	0.96	0.47	0.92	0.57	85%	87.4%

The accuracy scores of the training set is very import for ideation that the acquires have a stayed almost constant even after hyperparameter tuning but very slightly have increased on tuning the hyperparameters for the same.

The ROC curves values show very little difference between the models But as AOC score denotes a slight increase in the value. From this we understand that both the models perform similar to the nature of the data however on paper SVM suggests having made a very minimal accuracy increase in predicting the rain for tomorrow.

Critical analysis :

- The rain prediction is a highly technical standout point which involve various metrological attributes combing together to make accurate prediction. Our idea of predicting rain based on the 9 features alone even though they constitute 95% of the required features for predicting rain, still the method should be researched practically before put into practise.
- The use of imbalanced data was preferred mainly because of the accuracy errors due to over-sampling or under sampling of the dataset. But however as our model denotes the amount of true rain for tomorrow is predicted very less mainly because of the less amount of training data in the true rain set. Ideally more data should be collected for making the right prediction in the sense. Therefore making the model very conservative to pick ideal rain for tomorrow however our model is very efficient in predicting the no rain for tomorrow which is ideally need for predicting the rain too. SMOTE oversampling technique could have been used.
- SVM modelling took 6+ hours to train before of the non presence of pytorch configuration on the NVIDIA GPU which make the model to run on base CPU which in a industrial point of view is waste of resource and loss financially. This time taken to run caused the work to be completed at a early ratee which made the project less efficient.

- The hyperparameter tuning doesn’t improve the accuracy in a huge scale. Class imbalance is the cause of the factor here too.
- The function of MLP are majorly based on the amount of hidden layer and are said to perform better on the number hidden layers included. If the set data is improved with more hidden layer we may get a very better result in the future.

VIII. CONCLUSION LESSONS LEARNED REFERENCE AND FUTURE WORK

Conclusion: Based on the data set provided we are able to establish our models and find out if tomorrow will be a rainy day or not. We are able to predict that will be no rain tomorrow on very high accurate scale but the recall and precision values denotes that our model wont be efficient to and will predict less actually ran days mainly due to the imbalance.
MLP and SVM are the models used and both perform very well however as we can conclude that the SVM models performed slightly better than the MLP on the random sampled test set.

Lessons learned : Always balance a dataset. It is highly essential to balance a dataset to get very accurate results of the same. Even though under sampling or over sampling will be provide a model bias which wont be practical in many real world use cases we have to make attempts at least by collecting more data to make suit that dataset in a balanced manner to suit the target variable.

Due to the complexity SVM modelling took more time to build compared to MLP. The MLP always works on the concept of hidden layers hence on a computations scale it is recommend to use MLP and it performs better than SVM when the amount of hidden layers are more and adjusted perfectly.

Future plan : Further training data to be collection to standardise and to have a clearly balanced dataset. Tensor flow and Keras can be used mainly due to the availably for enhancing GPU performance and code simplicity. SMOTE generalising techniques could have been used to make the performance better. Optuna hypermeter tuning set will be researched and put into use.Various other modelling algorithms can be used to find how different models handle the dataset.

REFERENCES

[1]. Blaiech, A., Khalifa, K., Boubaker, M. and Bedoui, M., 2012. Implemntation of a Multi-Layer Perceptron Neura Networks in Multi-Width Fixed Point Coding. *International Journal of Modeling and Optimization*, pp.280-283.

[2]. Ertekin, Ş., 2009. *Learning in extreme conditions: Online and active learning with massive, imbalanced and noisy data*.

[3]. Kaggle.com. 2022. *Rain in Australia*. [online] Available at: <https://www.kaggle.com/datasets/jsphyg/weather-dataset-rattle-package> [Accessed 8 May 2022].Kahira, A., Nguyen, T., Gomez, L., Takano, R., Badia, R. and Wahib, M., 2021. An Oracle for Guiding Large-Scale Model/Hybrid Parallel Training of Convolutional Neural Networks. *Proceedings of the 30th International Symposium on High-Performance Parallel and Distributed Computing*,.

[4]. Kumar, P., Sharma, N. and Rana, A., 2012. Handwritten Character Recognition using Different Kernel based SVM Classifier and MLP Neural Network (A COMPARISON). *International Journal of Computer Applications*, 53(11), pp.25-31.

[5]. Medium. 2022. *Top 4 advantages and disadvantages of Support Vector Machine or SVM*. [online] Available at: <https://dhirajkumarblog.medium.com/top-4-advantages-and-disadvantages-of-support-vector-machine-or-svm-a3c06a2b107> [Accessed 8 May 2022].

[6]. Pointer, I., 2022. *What is Apache Spark? The big data platform that crushed Hadoop*. [online] InfoWorld. Available at: <https://www.infoworld.com/article/3236869/what-is-apache-spark-the-big-data-platform-that-crushed-hadoop.html#:~:text=Apache%20Spark%20is%20a%20data,with%20other%20distributed%20computing%20tools.> [Accessed 4 May 2022].

[7]. to Travel, C., 2022. [online] Available at: <https://www.climatestotravel.com/climate/australia> [Accessed 8 May 2022].

[8]. Torres, F., Trinidad, J. and -Ochoa, J., 2022. *An Oversampling Method for Class Imbalance Problems on Large Datasets*,.

Appendix :

IX. Glossary

Absorbing state: On which a neural network for a long time, tends to work

Activate : causing a response from the neuron

Activity : The amount of potential in which one neuron can arrive from its location to another

Architecture : The way of neurons which are connected to make a neural network

Back-error propagation: the algorithm which allows error between the output of the layer and its desired output

Bias : allows for identifying the neuron threshold

Binary classification : Outputs two or more sub classes that are high manually exclusive.

Boltzmann Machine : An algorithm for learning the probability between the distributions on a set of inputs

Confusion matrix : A NXN table to denotes how successful a classification matrix has performed

Dataframe: a type of pandas representation of data

Down sampling : for training a model the samples are decreased

Energy function: The numerical assignment in which the stability of neural net is indicated

Feature detection neurons : Neuron that are trained to detect any input aspects

Generalisation : The Ability of the neural network through which it can attain and give response to which it was never exposed.

Gradient : Partial vector which is a derivative

Hyperparameter : Tweaks the training to make it more accurate

Hyperplane : Splitting the data into two pieces

Learning law : Rule for shaping the changing the connection weights

Linear separability : The presence of a plane by which it can be sperate into a set of point in a space

Local minimum : The error value or an energy function

Neural assemblies : Active neuron sets

Output layer : A neuron set which gives output to specialty class

Recurrent network : The network in which active is fed back into the hidden layer

Training set : A set which is used for training

Test set: A set used for testing

Transfer function: The response of neuron usually a step function

Under sampling : Reducing an imbalanced dataframe to created a balanced dataset

Weight : linear feature model coefficient.

X. IMPLEMENTATION DETAILS:

Data processing EDA and Feature engineering: The data is imported using Python language using panads, NumPy library and clear process and all the feature engineering is done using the same.

Visualization : All the pre processing visualisation are done using seaborn library. missingno library is also used in some cases to identify null values. Various types of count plot and pair plots are used.

Data normalisation: Max min scalar function is used in this section

Model selection are based on simple techniques which are done on most cases as per the literature the easiest fitting classification models are chosen

In pytorch the training data is converted to tensors and we do the same and we plant a random seed in the training data

Model building :

MLP: BCE loss criterion ana adam optimiser is used for this neural net for the functionally

SVM: SK learn library is used for importing SVM into the system for the modelling

Hyperparameter tuning: In the case of tuning Grid search CV is used in both cases and the confusion matix is build and the precision recall scores are implemented using the classification report library

Finally the reports are done in MS word and converted to PDF and Pictures are snipped using the snipped tool. Redme file is devolved on notepad.