College of Engineering
Northeastern University

IE 7615: Neural Networks and Deep Learning
Final Project Report

BANKRUPTCY PREDICTION USING NEURAL NETWORKS

Bhopatrao, Rutvik   Seelam, Siva Teja Reddy   Rastogi, Piyush

ABSTRACT

Prediction of Bankruptcy is one of the major business classification problems. It is the task of predicting bankruptcy and various measures of financial distress of firms. The main aim of such projects is to help the investors and creditors in evaluating the likelihood of a firm going bankrupt. Our problem statement focuses on developing a predictive model that combines various econometric parameters which allow foreseeing the financial condition of a firm. Our dataset has synthetic features which were used to reflect higher-order statistics. The dataset we used for this project involves Polish companies. The data was collected from Emerging Markets Information Services (EMIS), which is a database containing information on emerging markets around the world. The bankrupt companies were analyzed in the period of 2000 to 2012, while the still operating companies were evaluated from 2007 to 2013. In this project, we have documented our observations as we try to explore, build and compare some of the widely used models such as Logistic Regression, Support Vector Machines and Neural Networks with/without dropout technique. This report contributes to a thorough understanding of the features of the tools used to develop bankruptcy prediction models and their related shortcomings. We begin by doing data preprocessing and exploratory analysis where we impute the missing values. In the end, we analyze and evaluate the performance of the models on the validation datasets. The average accuracy for all the 3 models ranges from 65% to 93%. The classification accuracy and validation test results indicate that Support Vector Machines model outperforms the other two models. Towards the end, we discuss the challenges we faced and suggest ways to improve the prediction, including the scope for future work.

SOURCE

UCI Machine Learning Repository:
https://archive.ics.uci.edu/ml/datasets/Polish+companies+bankruptcy+data#

INTRODUCTION

Financial distress and then the consequent failure of a business is usually an extremely costly and disruptive event. Statistical financial distress prediction models attempt to predict whether a business will experience financial distress in the future. Bankruptcy prediction holds an important role in economic decision making. Be it a large or small company, any kind of business concerns creditors, investors, local communities and industry participants. It also influences global economy to a large extent. Therefore, high social and economic costs created by bankruptcies have attracted attention of researchers for better understanding of the causes and factors and eventually prediction of financial distress [2]. The surge in the research is a function of the availability of data, for public firms which went bankrupt or did not, numerous accounting ratios that might indicate danger can be calculated, and numerous other potential explanatory variables are also available. Consequently, the area is well-suited for testing of increasingly sophisticated, data-intensive forecasting approaches [3].

The history of bankruptcy prediction includes application of numerous statistical tools which gradually became available and involves deepening appreciation of various pitfalls in early analyses. Interestingly, research is still published that suffers pitfalls that have been understood for many years. Bankruptcy prediction has been a subject of formal analysis since at least 1932, when FitzPatrick published a study of 20 pairs of firms, one failed and one surviving, matched by date, size and industry, in The Certified Public Accountant. He did not perform statistical analysis as is now common, but he thoughtfully interpreted the ratios and trends in the ratios. His interpretation was effectively a complex, multiple variable analysis.

The purpose of the bankruptcy prediction is to assess the financial condition of a company and its future perspectives within the context of long-term operation on the market [4]. It is a vast area of finance and econometrics that combines expert knowledge about the phenomenon and historical data of prosperous and unsuccessful companies. Typically, enterprises are quantified by numerous indicators that describe their business condition that are further used to induce a mathematical model using past observations [5].

There are different issues that are associated with the bankruptcy prediction. Two main problems are the following: First, the econometric indicators describing the firm's condition are pro- posed by domain experts. However, it is rather unclear how to combine them into a successful model. Second, the historical observations used to train a model are usually influenced by imbalanced data phenomenon, because there are typically much more successful companies than the bankrupted ones. As a consequent, the trained model tends to predict companies as successful (majority class) even when some of them are distressed firms. Both these issues mostly influence the final predictive capability of the model.

To speak about the modern methods of approach for the field of bankruptcy prediction, it is worth noting that survival methods are now being applied. Option valuation approaches involving stock price variability have been developed. Under structural models, a default event is deemed to occur for a firm when its assets reach a sufficiently low level compared to its liabilities. Neural network models and other sophisticated models have been tested on bankruptcy prediction. Modern methods applied by business information companies surpass the annual accounts content and consider current events like age, judgements, bad press, payment incidents and payment experiences from creditors.

LITERATURE REVIEW

The idea of using machine learning to predict bankruptcy has previously been used in the context of Predicting Bankruptcy with Robust Logistic Regression by Richard P. Hauser and David Booth [1]. This paper uses robust logistic regression which finds the maximum trimmed correlation between the samples remained after removing the overly large samples and the estimated model using logistic regression [1]. This model has its limitation. The value of this technique relies heavily on researchers' abilities to include the correct independent variables. In other words, if researchers fail to identify all the relevant independent variables, logistic regression will have little predictive value [7]. Its overall accuracy is 75.69% in the training set and 69.44% in testing set.

First attempts of the formal bankruptcy prediction trace back to the beginnings of the 20th century when first econometric indicators were proposed to describe predictive abilities of business failure (Fitzpatrick, 1932; Merwin, 1942; Winakor & Smith, 1935) [6]. The sixties of the twentieth century brought a turning point in the survey of the early recognition of the business failure symptoms. First of all, the work of Beaver (1966) initiated application of statistical models to the bankruptcy prediction. Following this line of thinking, Altman (1968) proposed to use multidimensional analysis to predict corporate bankruptcy that was further developed by others (Altman & Loris, 1976; Blum, 1974; Deakin, 1972; Edmister, 1972; Ketz, 1978; Koh & Killough, 1990; Laitinen, 1991; Libby, 1975; Meyer & Pifer, 1970; Pettway & Sinkey, 1980; Rujoub, Cook, & Hay, 1995; Sinkey, 1975; Wilcox, 1973) [7]. In parallel, a great interest was paid to the generalized linear models that can be used in both decision making and providing certainty of the prediction (Aziz, Emanuel, & Lawson, 1988; Grice & Dugan, 2003; Hopwood, McKeown, & Mutchler, 1994; Koh, 1991; Li & Miu, 2010; Ohlson, 1980; Platt & Platt, 1990; Platt, Platt, & Pedersen, 1994; Zavgren, 1983; Zmijewski, 1984) [8]. Additionally, the generalized linear models are of special interest because estimated weights of the linear combination of economic indicators in the model can be further used to determine importance of the economic indicators.

Since nineties of the 20th century artificial intelligence and machine learning have become a major research direction in the bankruptcy prediction. In the era of increasing volumes of data, it turned out that the linear models like the logistic regression or logit (probit) models are unable to reflect non-trivial relationships among economic metrics. Moreover, the estimated weights of the linear models are rather unreliable to indicate the importance of the metrics.

In order to obtain comprehensible models with an easy to understand knowledge representation, decision rules expressed in terms of first-order logic were induced using different techniques, naming only a few, like rough sets (Dimitras, Slowinski, Susmaga, & Zopounidis, 1999)[9] or evolutionary programming (Zhang et al., 2013). However, the classification accuracy of the decision rules are very often insufficient, therefore, more accurate methods were applied to the bankruptcy prediction. One of the most successful model was support vector machines (SVM) (Shin, Lee, & Kim, 2005). The disadvantages of SVM are that the kernel function must be carefully hand-tuned and it is impossible to obtain comprehensible model.

A different approach aims at automatic feature extraction from data, i.e., automatic non-linear combination of econometric indicators, which alleviates the problem of a specific kernel function determination in the case of SVM. This approach applies neural networks to the bankruptcy prediction (Bell, Ribar, & Verchio, 1990; Cadden, 1991; Coats & Fant, 1991; Geng, Bose, & Chen, 2015; Koster, Sondak, & Bourbia, 1991; Salchenberger, Cinar, & Lash, 1992; Serrano-Cinca, 1996; Tam, 1991; Tam &

Kiang, 1992; Wilson & Sharda, 1994; Zhang, Hu, Patuwo, & Indro, 1999) [10]. The main problem of the neural networks lies in the fact that they can fail in case of multimodal data. Typically, the econometric metrics need to be normalized/standardized in order to have all features of the same magnitude. This is also necessary for training neural networks so that the errors could be backpropagated properly. However, the normalization/standardization of data do not reduce the problem of the data multimodality that may drastically reduce predictive capabilities of the neural networks. That is why it has been advocated to take advantage of different learning paradigm, namely, the ensemble of classifiers (Kittler, Hatef, Duin, & Matas, 1998) [11]. The idea of the ensemble learning is to train and combine typi- cally weak classifiers to obtain better predictive performance. First approaches but still very successful were bagging (Breiman, 1996)[12] and boosting (Freund & Schapire, 1996; Friedman, 20 01; 20 02; Zi˛eba, Tomczak, Lubicz, & ´Swi˛atek, 2014)[13]. The idea of boosting was further developed to the case of unequal classification costs (Fan, Stolfo, Zhang, & Chan, 1999) and imbalanced data (Galar, Fernandez, Barrenechea, Bustince, & Herrera, 2012)[14]. Recently, the boosting method was modified to optimize a Taylor expansion of the loss functions, an approach known as Extreme Gradient Boosting (Chen & He, 2015a) that obtains state-of-the-art results in many problems on Kaggle competitions. 1 Recently, it has been shown that the ensemble classifier can be successfully applied to the bankruptcy prediction (Nanni & Lumini, 2009)[15] and it significantly beats other methods (Alfaro, García, Gámez, & Elizondo, 2008)[16].

The recent work within the field of Bankruptcy and Insolvency Prediction compares various differing approaches, modelling techniques, and individual models to ascertain whether any one technique is superior to its counterparts. Jackson and Wood (2013) provides an excellent discussion of the literature to date, including an empirical evaluation of 15 popular models from the existing literature. These models range from the univariate models of Beaver through the multidimensional models of Altman and Ohlson, and continuing to more recent techniques which include option valuation approaches. They find that models based on market data—such as an option valuation approach—outperform those earlier models which rely heavily on accounting numbers. Zhang, Wang, and Ji (2013)[17] proposed a novel rule-based system to solve bankruptcy prediction problem.

METHODOLOGY AND APPROACH

From this section, we focus on step and step approach of accomplishing the benchmarking results. The next sections focus on understanding the dataset, dataset features, instances, data preprocessing, etc. Then we deal with the problems found in the dataset like missing values, and data imbalance. Next, we introduce models we are going to work on the models and how we train them. In the end, we analyze and evaluate the performance of the models using metrics like accuracy, precision and recall.

DATA

For this project, we use the Polish bankruptcy dataset, hosted by the University of California Irvine(UCI) Machine Learning Repository, which is a huge repository of freely accessible datasets for research and learning purposes intended for the Machine Learning/ Data Science community. The data was collected from Emerging Markets Information Service (EMIS), which is a database containing information on emerging markets around the world. The bankrupt companies were analyzed in the period 2000-2012, while the still operating companies were evaluated from 2007 to 2013. The dataset is very apt for our research about bankruptcy prediction because it has highly useful econometric indicators as attributes (features) and comes with a huge number of samples of Polish companies that were analyzed in 5 different timeframes: Based on the collected data five classification cases were distinguished, that depends on the forecasting period:

1. 1st year: The data contains financial rates from 1st year of the forecasting period and corresponding class label that indicates bankruptcy status after 5 years.
2. 2nd year: The data contains financial rates from 2nd year of the forecasting period and corresponding class label that indicates bankruptcy status after 4 years.
3. 3rd year: The data contains financial rates from 3rd year of the forecasting period and corresponding class label that indicates bankruptcy status after 3 years.
4. 4th year: The data contains financial rates from 4th year of the forecasting period and corresponding class label that indicates bankruptcy status after 2 years.
5. 5th year: The data contains financial rates from 5th year of the forecasting period and corresponding class label that indicates bankruptcy status after 1 years.

Dataset Characteristic	Multivariate			
Number of features	64			
Feature characteristics	Real values			
Missing Values	Yes			
Associated tasks	Classification			
Number of Instances	Data	Total Instances	Bankrupt Instances	Non-Bankrupt Instances
	1st year	7027	271	6756
	2nd year	10173	400	9773
	3rd year	10503	495	10008
	4th year	9792	515	9227
	5th year	5910	410	5500

Table 1: Summary of the Polish bankruptcy dataset

ATTRIBUTES

attr1 – net profit/total assets
attr2 – total liabilities/total assets
attr3 – working capital/ total assets
attr4 – current assets/short term liabilities
attr5 – [(cash + short-term securities + receivables - short-term liabilities) / (operating expenses - depreciation)] * 365
attr6 - retained earnings / total assets
attr7 - EBIT / total assets
attr8 - book value of equity / total liabilities
attr9 - sales / total assets
attr10 - equity / total assets
attr11 - (gross profit+extraordinary item+financial expenses)/total assets
attr12 - gross profit/short term liabilities
attr13 - (gross profit+depreciation)/sales
attr14 - (gross profit+interest)/sales
attr15 - (total liabilities * 365) / (gross profit + depreciation)
attr16 - (gross profit + depreciation) / total liabilities
attr17 - total assets/total liabilities
attr18 - gross profit/total assets
attr19 - gross profit/sales
attr20 - (inventory*365)/sales
attr21 - sales(n)/ sales(n-1)
attr22 - profit on operating activities/total assets
attr23 - net profit/sales
attr24 - gross profit in 3 years/total assets
attr25 - (equity-share capital)/total assets
attr26 - (net profit+depreciation)/total liabilities
attr27 - profit on operating activities/financial expenses
attr28 - working capital/fixed assets
attr29 - logarithm of total assets
attr30 - (total liabilities-cash)/sales
attr31 - (gross profit + interest) / sales
attr32 - (current liabilities * 365) / cost of products sold
attr33 - operating expenses / short-term liabilities
attr34 - operating expenses / total liabilities
attr35 - profit on sales / total assets
attr36 - total sales / total assets
attr37 - (current assets - inventories) / long-term liabilities
attr38 - constant capital / total assets
attr39 - profit on sales / sales
attr40 - (current assets - inventory - receivables) / short-term liabilities
attr41 - total liabilities / ((profit on operating activities + depreciation) * (12/365))
attr42 - profit on operating activities / sales
attr43 - rotation receivables + inventory turnover in days
attr44 - (receivables * 365) / sales

attr45 - net profit / inventory
attr46 - (current assets - inventory) / short-term liabilities
attr47 - (inventory * 365) / cost of products sold
attr48 - EBITDA (profit on operating activities - depreciation) / total assets
attr49 - EBITDA (profit on operating activities - depreciation) / sales
attr50 - current assets / total liabilities
attr51 - short-term liabilities / total assets
attr52 - (short-term liabilities * 365) / cost of products sold)
attr53 - equity / fixed assets
attr54 - constant capital / fixed assets
attr55 - working capital
attr56 - (sales - cost of products sold) / sales
attr57 - (current assets - inventory - short-term liabilities) / (sales - gross profit - depreciation)
attr58 - total costs /total sales
attr59 - long-term liabilities / equity
attr60 - sales / inventory
attr61 - sales / receivables
attr62 - (short-term liabilities *365) / sales
attr63 - sales / short-term liabilities
attr64 - sales / fixed assets
attr65 (class) – the response variable Y:0=did not bankrupt; 1=bankrupt

DATA ASSESSMENT

First, we look at the statistics of the missing values. To test, we plot of the nullity matrix for the 1st year dataset that explains the sparsity of 1st Year data. The nullity matrix gives us a data-dense display which lets us visually pick out the missing data patterns in the dataset. We notice that the features attr21 and attr37 have the highest number of missing values. Then, we explore to see the correlation among various features in the 1st Year data as an example. The range of this nullity correlation is from -1 to 1. This is how we check the correlation among the features with respect to the missing values. Then, we check how much data is the missing values.

We further summarize the populations of class labels in each dataset. Looking at the class label, we can see that there are a minority of bankrupt class labels as compared to non-bankrupt class label. The table shows the data imbalance for the datasets.

Dataset	No of total instances	No of bankrupt instances	No of non bankrupt instances	Percentage of minority class samples
1st year	7027	271	6756	3.85%
2nd year	10173	400	9773	3.93%
3rd year	10503	495	10008	4.71%
4th year	9792	515	9277	5.25%
5th year	5910	410	5500	6.93%

Table 2: Assessing the data imbalance for all the datasets

To deal with the imbalance technique, we use the SMOTE technique, which is a widely used oversampling technique.

DATA MODELLING

We look at the various classification models that we have considered for training on the Polish bankruptcy datasets to achieve the task of coming up with a predictive model that would predict the bankruptcy status of a given unknown company with an appreciable accuracy. We have considered the following 4 models:

1. Logistic Regression
2. Support Vector Machine
3. Neural Network
4. Neural Network with dropout technique

Logistic Regression Classifier

Logistic regression is a linear model for classification. It is also known as logit regression, maximum-entropy classification (MaxEnt) or the log-linear classifier. In this model, the probabilities describing the possible outcomes of a single trial are modeled using a logistic function.

$$\text{LogitClassifier}(x) = min\|w\|_1 + C\sum_{i=1}^{n}\left(\log\left(\exp\left(y(x_i^T w + C)\right) + 1\right)\right)$$

Introduced L2 penalty into the model, which calculates Logitclassifier values for datapoints.

Support Vector Machine

Specifically, we use support vector classify (SVC), a subcategory of SVM, in this task. It constructs a hyper-plane, as shown in Figure 1, in a high dimensional space which is used for classification. Generally, a good separation represented by the solid line in Figure 1 means the distance(the space between the dotted lines) to the nearest training data points (the red and blue dots) of any class (represented by the color red and blue) is the largest. This is also known as functional margin.

With training vectors in two classes and a vector,

$$x_i \in R^p, i = 1 \dots n, y \in \{1, -1\}^n$$

respectively, SVM aims at solving the problem:

$$\min_{wb\zeta} \frac{1}{2}\omega^T\omega + \mathbb{C}\sum_{i=1}^{n}\zeta_i$$

subject to

$$y_i(\omega^T \phi(x_i) + b) \geq 1 - \zeta_i$$

Its dual is

$$\min_{\alpha} \frac{1}{2} \alpha^T Q - e^T \alpha$$

subject to

$$y^T \alpha = 0, 0 \leq \alpha_i \leq C, i = 1 \dots n$$

where e is a common vector, C>0 is upper bound, Q is n by n positive semidefinite matrix, $Q_{ij} \equiv y_j y_j k(x_i . x_J)$, and $K(x_i, x_j) = \phi(x_i)^T \phi(x_j)$ is the kernel.

Here the function implicitly maps the training vectors into a higher dimensional space.

The decision function is:

$$sgn(\sum_{i=1}^{n} y_1 \cdot \alpha_i k(x_i, x) + \rho)$$

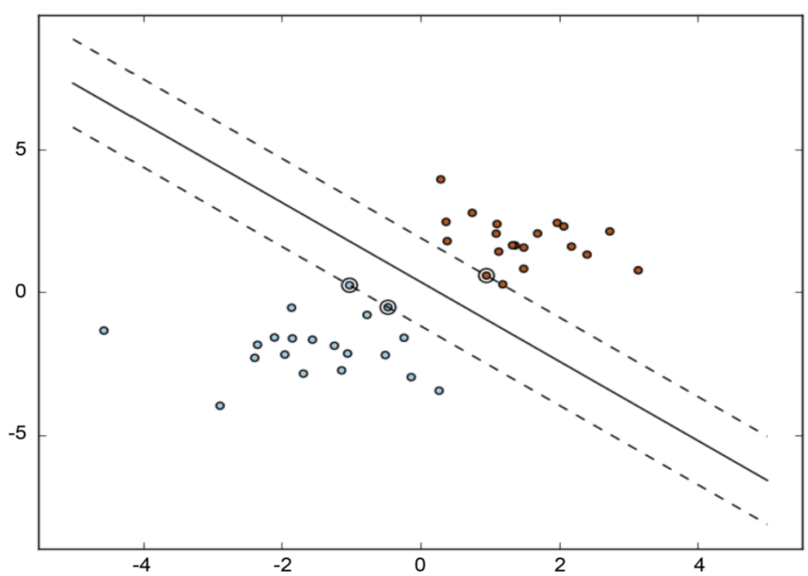

Figure 1. SVM model [3]

Neural Network with Dropout technique

Neural networks' inputs are modelled as layers of neurons. Its structure is shown in the latter figure. As shown in Figure 2, the formal neuron uses n inputs x1, x2,…xn to classify the signals coming from dendrites, and then synoptically weigh correspondingly with w1, w2, … wn that measure their permeabilities. Then, the excitation level of the neuron is calculated as the weighted sum of input values:

$$\xi = \sum_{i=1}^{n} x_i \omega_i$$

f in Figure 3 represents activation function.

When the value of excitation level x reaches the threshold h, the output y (state) of the neuron is induced.induced. This simulates the electric impulse generated by axon.

Dropout is a technique that further improves neural network's accuracy. In Figure 3, let L be the number of hidden layers, l∈{1…L} the hidden layers of the neural network, z(l) and y(l) the vectors of inputs and outputs of layer l ,respectively.

W(*l*) and *b*(*l*) are the weights and biases at layer *l*. For *l*∈{0′···,*L* −1} and any hidden unit i, the network then can be described as:

$$z^{(i+1)} = w^{(i+1)\cdot} + b^{(i+1)}$$
$$y^{(l+1)} = f\left(z^{(l+1)}\right)$$

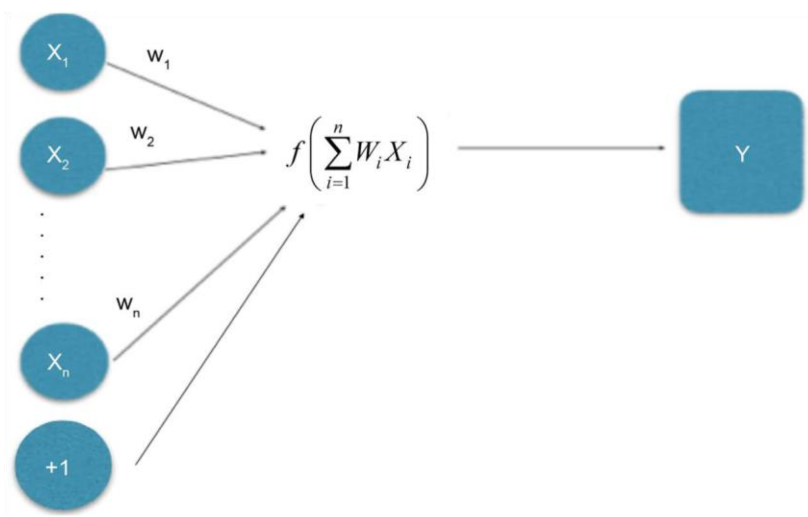

Figure 2. Neural network model [13]

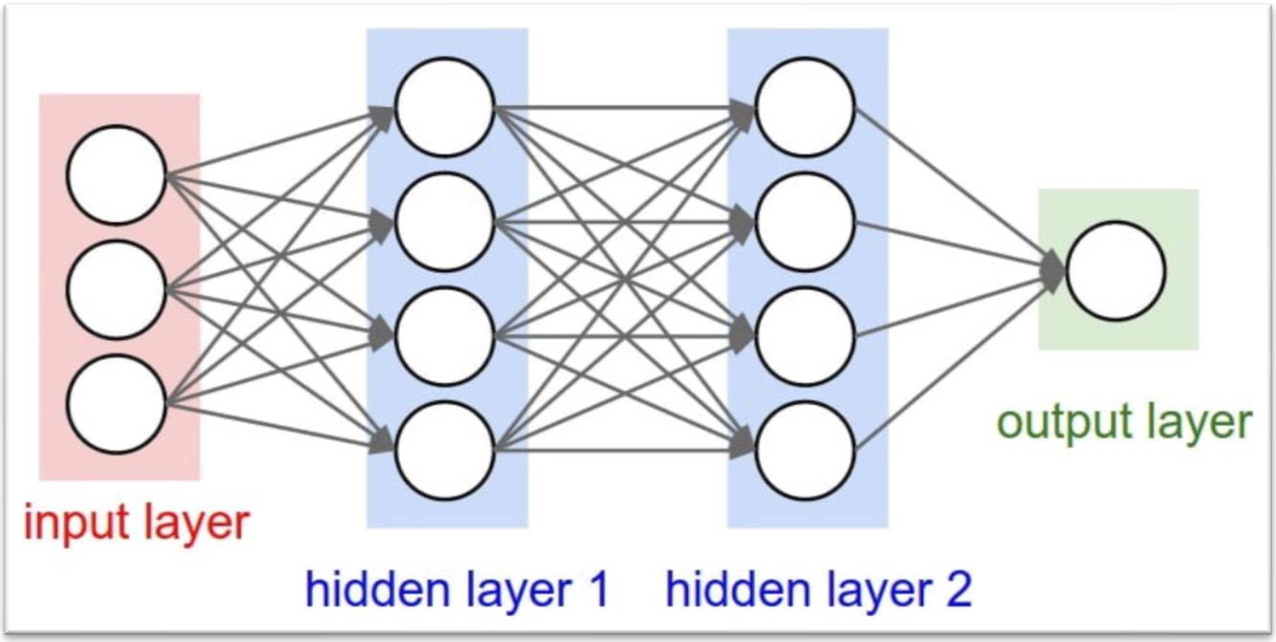

Figure 3. Artificial neural network [3]

where f is any activation function.
With dropout, the feed-forward operation becomes: $r^{(l)}$-Bernoulli(p),

$$y^{(l)} = r^{(l)} y^{(l)}$$
$$z^{(l+1)} = w^{(1+1)} y^{l} + b^{(l+1)}$$

PARALLEL COMPARISON

Logistic Regression(L2)

The accuracy by the Logistic regression model is good but lowest when we consider the other models. The accuracy for all the 5 datasets were found to be,

1st year: 75.9203036053131,
2nd year: 75.9043250327654,
3rd year: 74.9539828625833,
4th year: 74.86044928522804,
5th year: 72.72419627749576.

SVM (Linear kernel)

The accuracy for this model increases when truncate increases in a SVM model. This model gives us the highest accuracy amongst all the models which was found to be,

1st year: 95.82542694497154,
2nd year: 96.06815203145479,
3rd year: 94.9857188194224,
4th year: 94.96255956432948,
5th year: 92.78059785673999.

NEURAL NETWORK (Activation = Softmax, Num_Classes = 2, Optimiser = Adam, Loss = Categorical _Crossentropy, Metrics = Accuracy)

When other things in the model hold the same value, dropout technique of 0.5 gives us the highest accuracy amongst the neural network models which was found to be,

1st year: 85.9203036053131,
2nd year: 86.26474442988204,
3rd year: 85.0809266899397,
4th year: 85.13274336283186,
5th year: 82.94980259447264.

CODE

1. We load all the libraries.
2. Then we load the raw data (.arff files) as pandas dataframes and assign the new column headers to them. Although the features are numeric and class labels are binary, in the dataframes, all the values were stored as objects. So we converted them to float and int values respectively.
3. Now we start the data analysis. Firstly, we see how much of data is missing in each dataframe and look at the nullity (sparsity) by generating the nullity matrix and nullity heatmaps respectively.
4. Then we perform imputation of the missing data using Mean, k-NN, EM and MICE imputation techniques and generate fresh dataframes of imputed data.
5. We apply SMOTE oversampling on all these imputed dataframes to get fresh dataframes of imputed-and-oversampled dataframes and store them in a dictionary.
6. We create models.
7. We iterate over all the models. In each model, we iterate over all oversampled dataset collections. Each collection has 5 dataframes corresponding to 5 years' data.

CONCLUSIONS

Logistic regression, Support vector machine and Neural network with dropout are three relatively new models applied in bankruptcy prediction problems. Their accuracies outperform those of the three older models (inductive learning algorithms, genetic algorithms). The improved aspects include the control for overfitting, the improved probability of finding the global maxima, and the ability to handle large feature spaces. This paper compared and concluded the progress of machine leaning models regarding bankruptcy prediction and checked to see the performance of relatively new models in the context of bankruptcy prediction that have rarely been applied in that field. However, the three models also have drawbacks. SVM does not directly give probability estimates but uses an expensive five-fold cross-validation instead. Also, if the data sample is not big enough, especially when outnumbered by the number of features, SVM is likely to give bad performance. With dropout, the time to train the neural network will be 2 to 3 times longer than training a standard neural network. And this can be a problem when the most relevant information only makes up a small percent of the input. The solutions to overcome these drawbacks are yet to be found.

FUTURE SCOPE

In this section, we intend to discuss the future work in bankruptcy prediction. So far, in our experiments, we have dealt with the synthetic features—an arithmetic combination of core econometric features. But it is also possible to gather more core features and hence synthesize more synthetic features by varying the arithmetic operations performed on these core features. Also, it is possible to synthesize more features considering the current synthetic features as base features. Doing so may result in better prediction of bankruptcy, but it must be thoroughly validated by domain experts, as to whether such highly complex synthetic features would be meaningful, in terms of financial economics. It is also feasible to reduce the dimensionality of features. But for a dataset like the Polish bankruptcy data we have seen so far, with such high missing data, it becomes difficult to rank the features and perform feature extraction. The features dropped in such manner might bear significant impact in the prediction, had the data not been so much sparse. So, the takeaway is that, if the data to be collected in the future, pertinent to bankruptcy prediction, is made sure to less sparse, it is possible to apply all the techniques mentioned in the future scope so far, and hence obtain better predictive models.

REFERENCES

[1] Hauser, R.P. and Booth, D. (2011) Predicting Bankruptcy with Robust Logistic Regression. Journal of Data Science, 9, 565-584.

[2] Kim, M.-J. and Han, I. (2003) The Discovery of Experts' Decision Ruels from Qualitative Bankruptcy Data Using Genetic Algorithms. Expert Systems with Applica-tion, 25, 637-646,

[3] Pedregosa, et al. (2011) Scikit-Learn: Machine Learning in Python. Journal of Machine Learning Research, 12, 2825-2830.

[4] Sirvastava, N., et al. (2014) Dropout: A Simple Way to Prevent Neural Networks from Overfitting. Journal of Machine Learning Research, 15, 1929-1958.
[5] Dev, D. (2017) Deep Learning with Hadoop. Packet Publishing, Birmingham, 52.
[6] Nielsen, F. (2001) Neural Networks—Algorithms and Applications. https://www.mendeley.com/research-papers/neural-networks-algorithms-applicatio ns-5/
[7] Robinson, N. (n.d.) The Disadvantages of Logistic Regression. http://classroom.synonym.com/disadvantages-logistic-regression-8574447.html
[8] Sima, (1998) Introduction to Neural Networks. Technical Report No. 755.
[9] Baldi, P. (2012) Autoencoders, Unsupervised Learning, and Deep Architectures. Journal of Machine Learning Research, 27, 37-50.
[10] Martin, A., Uthayakumar, J. and Nadarajan, M. (2014) Qualitative Bankruptcy Data Set, UCI. https://archive.ics.uci.edu/ml/datasets/qualitative_bankruptcy
[11] Laitinen, E. K. (1991). Financial ratios and different failure processes. Journal of Business Finance & Accounting, 18, 649–673
[12] Zhang, Y., Wang, S., & Ji, G. (2013). A rule-based model for bankruptcy prediction based on an improved genetic ant colony algorithm. Mathematical Problems in Engineering, 2013
[13] Wikipedia contributors. "Bankruptcy prediction". Wikipedia, The Free Encyclopedia, <https://en.wikipedia.org/wiki/Bankruptcy_prediction>
[14] Constand, R. L., & Yazdipour, R. (2011). Firm failure prediction models: a critique and a review of recent developments. Advances in Entrepreneurial Finance (pp. 185–204). Springer
[15] Merwin, C. L. (1942). Financing small corporations in five manufacturing industries. NBER Books p. 1926-1936. New York: National Bureau of Economic Research, Inc.
[16] Shapiro, A. (1983). The role of structured induction in expert systems. Ph.D. Thesis, University of Edinburgh
[17] Sinkey, J. F. (1975). A multivariate statistical analysis of the characteristics of problem banks. The Journal of Finance, 30, 21–36