International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Predicting Credit Risk Defaulting in Peer-To-Peer (P2p) Lending with Deep Learning

Victor Bwalya[1], Muwanei Sinyinda[2], Namonje Nancy[3]

[1]University Lecturer, School of Engineering and Technology, Mulungushi University
[2]College Lecturer, Information Communications Technology, Copperbelt University
[3]Senior university Lecturer (Head of Depa, School of Enginnering and Technology (SET), Mulungushi University

ABSTRACT

Peer-to-peer (P2P) lending has gained popularity as a method of lending money in recent years. It has emerged as a disruptive force in the financial industry, connecting borrowers directly with lenders. However, accurate credit risk assessment is crucial to P2P lending's viability and success. The complex and dynamic relationships included in P2P lending data may be difficult for traditional credit risk defaulting models to capture. In the context of peer-to-peer lending, this study suggests a novel strategy to improve credit risk default prediction by utilizing deep learning techniques with feature engineering. Due to the burdensome application processes and excessive interest rates, many customers are unable to get personal loans from banks. Because of this, potential borrowers are still looking for other ways of getting a loan, and that's where P2P lending comes in. The primary focus of this research was to develop a deep learning model specifically tailored for predicting credit risk defaulting in P2P lending by making use of feature engineering techniques, assess the model's ability to capture complex patterns, and compare its performance against traditional credit scoring models. The research methodology involved the collection of historical loan data from a representative of P2P lending platform, comprehensive data preprocessing (feature engineering techniques), and the implementation of deep learning model - CNNs (Convolution Neural Networks).

When training the model, predictive accuracy and risk sensitivity were optimized. Performance criteria like accuracy, precision, recall, F1-score, and AUC-ROC were used to assess the trained proposed model. 99.93% had the highest accuracy rate, which was superior compared to the other models. AUC was 90.96%, recall was at 72%, and finally precision was at 88%. The findings indicated that the suggested CNN-based prediction model performed better when incorporated with feature engineering techniques, to predict credit risk defaulting in peer-to-peer lending. This will also help reduce human bias in credit decision-making, leading to fairer outcomes for borrowers from diverse backgrounds.

Keywords: Peer-to-Peer (P2P) lending, Deep Learning, Convolutional Neural Network (CNN), Credit Risk Scoring, Model evaluations, Feature Engineering Techniques

1. INTRODUCTION

Peer-to-peer (P2P) lending platforms have emerged as a transformative force in the financial sector, reshaping the lending landscape by connecting borrowers directly with individual lenders. (Wang et al, 2018) States that, the success and sustainability of P2P lending heavily depend on effective credit risk assessment. Traditional credit scoring models may face challenges in capturing the intricate and dynamic relationships inherent in P2P lending data (Machado et al, 2022) (Bastani et al 2019). This research proposes an innovative approach by anchoring deep learning techniques to enhance credit risk prediction in the context of P2P lending.

Credit-risk models are used to estimate customers' worthiness to receive credit. The higher the credit risk

score, the lower the risk and the better the terms of the loan that will be offered to the customer (Marcos and Salma 2022).

Credit risk predictions are estimated by using information about the potential customers' payment history, types of credit taken, current debt and length of credit, as well as demographic and other behavioral information (Thomas et al, 2005)

Typically, a lending firm will utilize credit scoring to determine a borrower's credit risk before approving a loan; this involves filtering out applicants who don't meet certain credit score requirements.

Should it fall short of the required score, the borrower will not be qualified for the loan. In this way, a borrower's credit score is crucial for all forms of lending financing, including peer-to-peer lending. P2P lending may increase the risk of default because it reaches out to borrowers with poor credit, but it is unclear how to calculate this risk.

For this reason, credit risk assessment remains essential in peer-to-peer lending, and accurate assessment may increase the likelihood that well-credit borrowers may utilize P2P lending. The anticipated significance of this research lies in providing a more accurate and robust credit risk assessment tool for P2P lending platforms, thereby enhancing risk management practices in the financial industry (Dastile et al, 2021).

The study also contributes to a broader understanding of the applicability of deep learning techniques for financial risk assessment. Results are expected to show that the proposed deep learning model with feature engineering techniques outperforms traditional credit scoring models, demonstrating the potential of advanced ML / DL methods to augment risk prediction in P2P lending (Khan et al, 2023). The timeline encompasses data collection, data preprocessing (feature engineering techniques), model development, training, evaluation, and comparison, with a comprehensive budget allocated for data acquisition, computational resources, and necessary tools. Ethical considerations prioritize compliance with data protection regulations, obtaining required permissions, and ensuring privacy and confidentiality (Khan et al, 2023).

In a nutshell, this research addresses a critical need in the P2P lending landscape by introducing an advanced credit risk defaulting model powered by deep learning with feature engineering techniques. The findings aim to contribute to the sustainable growth of P2P lending platforms, benefiting both lenders and borrowers through improved risk assessment practices. Convolutional Neural Network (CNN) was used for our model development. CNN is a type of deep learning model that is particularly well-suited for processing and analyzing different types of datasets. CNNs have also been successfully applied to other types of data, such as time-series data in financial forecasting or sequential data in natural language processing (Swee et al, 2023) (Chen et al, 2019).

The remaining part of this paper is organized as follows: Part 2 covers the related works section which is the literature review, while the materials and methods are described in Part 3. Part 4 presents the results, followed by the discussion in Part 5 and then finally conclusions in Part 6.

2. RELATED WORKS

This section explores existing methodologies for credit risk prediction in P2P lending and examines the limitations they possess, laying the foundation for the proposed research on integrating deep learning techniques. (Bhatore et al, 2020) In this survey paper, the authors performed a systematic literature review on existing research methods and ML techniques for credit risk evaluation. A total of 136 papers on credit risk evaluation published between 1993 and March 2019 were reviewed.

They examined the limitations of recent studies and research trends, concentrating on the effects of hyper parameters on machine learning approaches used to assess credit risk. After the review they found out that Ensemble and Hybrid models with neural networks and SVM are being more adopted for credit risk predictions, NPA prediction and fraud detection. They also figured that lack of comprehensive public datasets continues to be an area of concern for researchers.

(Dastile et al, 2021) Had a thorough investigation of systematic machine learning and its application in credit risk. Nevertheless, the role of deep learning models in credit risk hasn't been fully expressed.

(Kriebel et al, 2022) Proposed a deep learning and several other techniques to extract credit-relevant information from user-generated text on Lending Club. The authors figured that even short pieces of user-generated text can improve credit default predictions significantly. Compared with other approaches that use text, deep learning outperforms them in almost all cases.

(Kriebel et al, 2022) A comparison of six deep neural network architectures, including state-of-the-art transformer models, finds that the architectures mostly provide similar performance. This indicates that in this credit scoring context, less sophisticated approaches (like average embedding neural networks) perform on par with more sophisticated approaches (like the transformer networks BERT and RoBERTa)

(Tripathi et al, 2020) The authors proposed a novel activation function and an evolutionary approach to get optimized weights and biases by utilizing Bat optimization algorithm. Further, the simulations were performed on bench-marked (four) credit scoring datasets with various activation functions. Simulation results demonstrate that the suggested Evolutionary ELM (EELM) is more suitable for credit risk evaluation.

(Tripathi et al, 2020) ELM is a single hidden layer feed forward neural network and its performance depends upon activation function, weights and biases assigned to hidden neurons. (Peter et al, 2018) In order to estimate the likelihood of a loan default, the authors developed a binary classifier using real data and deep learning models. In this model, the authors identified 10 most important features from the named models and then they were used in the modeling process to test the stability of binary classifiers by comparing their performance on separate datasets. They deduced that the tree-based models are more stable than the models based on multilayer artificial neural networks. And so, this calls for more research to be done relative to the intensive use of deep learning systems in enterprise.

(Van-Sang et al, 2019) The authors designed a credit risk model based on advanced machine learning methods, capable of assessing the returns and risks of individual loans. They also applied a selection feature method to eliminate the irrelevant features for improving the efficiency of the ML models.

The suggested methodology could increase investment efficiency when compared to the current P2P lending methods, according to experiments done on real data sets from P2P lending platforms. (Wanga et al 2020) The researchers focused on the comparative assessment of the performances of five popular classifiers involved in machine learning used for credit scoring: Naive Bayesian Model, Logistic Regression Analysis, Random Forest, Decision Tree, and K-Nearest Neighbor Classifier. The authors figured that each model has its own strength and weakness, and so they deduced that it is assertive to say which one is the best. However, the results of this experiment proved that Random Forest performs better than the rest in terms of precision, recall, AUC (area under curve) and accuracy. (Wang et al, 2023).

(Feng et al, 2020) Developed a new deep learning ensemble credit risk scoring evaluation model to deal with imbalanced datasets. an improved synthetic minority oversampling technique (SMOTE) method was developed to overcome known SMOTE shortcomings, after which a new deep learning ensemble classification method combined with the long-short-term-memory (LSTM) network and the adaptive

boosting (AdaBoost) algorithm was developed to train and learn the processed credit datasets. The performance of the suggested model and other popular credit scoring algorithms was compared using a variety of indicators on two unbalanced credit datasets.

(Feng et al, 2020) The experimental test results indicated that the proposed deep learning ensemble model was generally more competitive when addressing imbalanced credit risk evaluation problems than other models. As a future enhancement the authors suggested that there be an application of the deep learning ensemble method to multi-class rather than binary class assessment; and the employment of other deep learning algorithms with excellent capabilities such as the CNN, RNN, and the DBN as the ensemble model base learners, and using other ensemble approaches such as bagging and stacking to construct the credit scoring models.

(Khan et al, 2023) Looks at the loan default prediction model based on deep learning, by establishing a network loan default prediction model through BPNN (Back Propagation Neural Network). It compares the effectiveness of traditional machine learning techniques and deep learning BPNN by evaluating loan default probabilities when dealing with huge amounts of data. The results indicated that BPNN performed better than other traditional ML techniques, as its accuracy was at 98.01%.

(Cheng et al, 2021) Applies a deep learning multiple kernel classifier as a state-of-the-art technique, which is efficient in coping with deep structure and complex data in credit risk assessment. The results indicated that deep learning multiple kernel classifier outperforms conventional and ensemble models. The findings proved that DL models generate more reliable performance on default and prepayment risk than conventional models.

The authors recommended to build a hybrid classification process model focused on further features to assess a credit risk model as a future enhancement.

(Bussmann et al, 2020) The paper proposes an XAI model that can be used in credit risk assessment and in measuring the risks that arise when credit is borrowed employing peer to peer lending platforms. The model applies correlation networks to Shapley values so that AI predictions are grouped according to the similarity in the underlying explanations. As a future enhancement the authors suggested that there should be an extension with the proposed methodology to other datasets, especially imbalanced ones, for which the occurrence of defaults tends to be rare and even more than what observed for the analyzed data.

Traditional credit scoring models have been the cornerstone of risk assessment in P2P lending. They leverage statistical and machine learning techniques to analyze borrower attributes and historical repayment data. However, studies (Swee et al, 2022) highlight their limitations in capturing non- linear relationships and temporal dependencies. (Chen et al, 2019) Proposes other deep learning methods such as convolutional neural Network (CNN) may be used to extract features and the recurrent convolutional Neural network (RCNN) to be applied to enhance the prediction accuracy with respect to the sequenced Loan applications.

(Xolani Dastile and Turgay Celik, 2021) The proposed method converts tabular datasets into images and thus allowing the application of 2D CNNs in credit scoring. Each pixel of the image corresponds to a feature bin of the tabular dataset. The predictions from the 2D CNNs were explained using state-of-the art explanation methods. Furthermore, explanations were evaluated using a sanity check methodology and performances of the explanation methods were compared quantitatively.

The proposed explainable deep learning model outperforms the other credit scoring methods on publicly available credit scoring datasets. (Xolani Dastile and Turgay Celik, 2021) As a future enhancement the authors proposed that there be validating and evaluating performances of the explanations by using domain

experts in the field of credit scoring such as the credit risk analysts.

They also suggested that let there be a technique on comparing the performances of the methods that perform tabular data conversion into images. Further, more work was needed on the proposed framework that it can be extended to other deep learning architectures such as CNNs, ResNet, Deep Graph Neural Networks and others.

(Mhlanga 2021) The author discovered that artificial intelligence and machine learning have a strong impact on credit risk assessments using alternative data sources such as public data to deal with the problems of information asymmetry, adverse selection. This will allow lenders to do serious credit risk analysis, to assess the behavior of the customer, and subsequently to verify the ability of the clients to repay the loans, permitting less privileged people to access credit.

Therefore, (Mhlanga 2021) recommends that financial institutions such as banks and credit lending institutions like peer-to-peer lending facilities to invest more in artificial intelligence and machine learning to ensure that financially excluded households can obtain credits on merit. The author also recommends that governments ensure that emerging economies be able to start investing in AI and machine learning. This is important because many emerging economies find it hard to do serious investments due to the costs related to these technologies.

(Khatir et al, 2022) The authors used five different ML classifiers, and each of them is combined with different feature selection techniques and various data-balancing approaches. According to the empirical analysis of a retail credit bank dataset, they found that the best combination is given by RF, random forest recursive feature elimination and random oversampling. As a future enhancement the authors proposed the use of a larger real dataset to double-check the accuracy of the models. Another issue that was figured and open to further research is the use of different credit categories to test the models. In future research, the authors planned to extend the investigation to corporate defaults.

(Lin et al, 2022) The authors developed a penalized deep learning model to predict default risk based on survival data. As opposed to simply predicting whether default will occur, they focused on predicting the probability of default over time. Moreover, by adding an additional one-to-one layer in the neural network, they achieved feature selection and estimation simultaneously by incorporating an L1-penalty into the objective function. An analysis of a real-world loan data and simulations demonstrate the model's competitive practical performance, which suggests favorable potential applications in peer-to-peer lending platforms. (Lin et al, 2022) Future work can be extended to other data settings, such as truncated data and interval censored data. Further, when several commercial datasets are available for analysis, it can be important and challenging to recognize and accommodate cross-data heterogeneity. We leave it to future research to study the prediction of default risk in an integrative analysis framework.

(Machado and Holmer, 2022) Examined the performance of the machine learning boosting algorithms XGBoost and CatBoost, with logistic regression as a benchmark model, in terms of assessing credit risk. These methods were applied to two different data sets where grid search was used for hyperparameter optimization of XGBoost and CatBoost. According to their results, the machine learning boosting methods outperformed logistic regression on the test data for both data sets and CatBoost yield the highest results in terms of both accuracy and AUC. The authors also admitted to the fact that they lacked computational power required to be able to get optimal set of hyperparameters.

(Machado and Holmer, 2022) They also suggested that more powerful algorithms can be implemented and try them on various sets of data, containing both numerical and categorical variables. it would also help in yielding optimal results. Hence the need for the proposed model using deep learning.

(Li et al 2019) The authors proposed a new default risk prediction model based on heterogeneous ensemble learning. The classifiers included: extreme gradient boosting (XGBoost), a deep neural network (DNN) and logistic regression (LR). They were used simultaneously with a liner weight ensemble strategy. In particular, the proposed model can process missing values. After generating discrete and rank features, the model was able to add missing values to the model for self-training. Then, the hyperparameters were optimized by the XGBoost model to improve the performance of the prediction model. Finally, compared with the benchmark model, the proposed method significantly improves the accuracy of the prediction results and so it solved the class-imbalance problem. (Nikita et al, 2022) The research paper made three (3) major contributions: Revisiting statistical fairness criteria and examine their adequacy for credit scoring, cataloging algorithmic options for incorporating fairness goals in the ML model development pipeline and Empirically comparing multiple fairness processors in a profit-oriented credit scoring context using real-world data.

The empirical results substantiate the evaluation of fairness measures, identify suitable options to implement fair credit scoring, and clarify the profit-fairness trade-off in lending decisions (Nikita et al, 2022). They found that multiple fairness criteria can be approximately satisfied at once and recommend separation as a proper criterion for measuring scorecard fairness.

(Simon, 2021) The author developed auxiliary vector machine models, decision trees, and random forests. He then compared their prediction accuracy with benchmarks based on logistic regression models. Which showed that the performance of Random Forest is better than other models. In addition, the performance of the support vector machine model is poor when using linear and non-linear kernels. The results also deduced that financial institutions could create value. Improve standard predictive models by researching more machine learning and / or deep learning techniques.

(Simon, 2021) One of the challenges the author faced came from when the sample contained a few default values, which made it difficult to construct a reliable scorecard. The other obstacle came from data set acquisitions due to the sensitivity of privacy laws regarding the data. Solution in this kind of situation is the use of artificial data to overcome these limitations and create special conditions for studying specific issues.

(Bertrand 2020) In this paper, the author was trying to analyze how social biases are transmitted from the data into banks loan approvals by predicting either the gender or the ethnicity of the customers using the exact same information provided by customers through their applications. and the assumption which we made was that if social biases were not included, then factors characterizing credit scores would be sufficiently different from those that can characterize either men or women. If the data used to score customers can be used to predict any sensitive information, and if the data is socially biased, then the credit score will also be biased. (Bertrand 2020) The author proposed a technique of Unbiasing either the data set or the model. Though will have to ensure the issue is carefully addressed considering that unbiasing a data set is likely to engender an opposite bias.

(Schmitt 2022) The two models currently competing for the pole position in credit risk management are deep learning (DL) and gradient boosting machines (GBM). (Schmitt 2022) The author benchmarked the named algorithms in the context of credit risk scoring using three (3) different datasets with different features to account for the reality that model choice is often dependent on the underlying characteristics of the dataset.

The experiment deduced that GBM proved to be more powerful than DL and has also the advantage of speed due to lower computational requirements. based on this research study both algorithms can be

considered state-of-the-art for binary classification tasks on structured datasets. The author also proposed that a future research should focus on transparency, auditability, and explainability of machine learning models in credit scoring (Bücker, Szepannek & Biecek, 2022; Dastile & Celik, 2021).

(Bastani et al, 2019) Test results showed that wide and deep learning model in combination with IHT achieves the highest performance on the loan status prediction compared to the benchmark algorithms although econometrics techniques such as Heckman two-step correction (Heckman, 1977), which are mainly designed to correct the selection bias, can also be utilized to address the imbalance problem with the IRR distribution. In as much as predicting credit risk score is our major concern, (Ampountolas et al, 2021) states that more practical measures should be used to develop the model for assisting investors in making investment decisions. In addition, the estimated performance of extreme gradient models with available data is far from meeting the requirements of investors and so massive research is needed which will enable massive returns and greater efficacy.

(Yujia et al, 2024) Evaluated the stability of LIME and SHAP on datasets of progressively increased class imbalance. The results show that interpretations generated from LIME and SHAP are less stable as the class imbalance increases, which indicates that the class imbalance does have an adverse effect on machine learning interpretability. To check the robustness of our outcomes, they also analyzed two open-source credit scoring datasets and obtained similar results.

(Yujia et al, 2024) The authors also suggested that it would be preferable to apply other imbalanced learning techniques to credit scoring, such as deep learning techniques to help in generating unbiased interpretation results. More fundamentally, it would be valuable to investigate the effects of class imbalance on the stability of interpretation methods theoretically.

(Kumar et al, 2016) states that; In as much as random forest scored above 88 % in its prediction, there is need to explore classifiers that are probability based and anomaly detection algorithms with features that correspond to gaussian distribution. (Mohan et al, 2020) recommends the use of an extensive data set to boost the model's performance and provide more accurate estimations.

(Khemakhem et al, 2017) states that P2P lending datasets may have missing values, outliers, and inconsistencies. Access to comprehensive and high-quality data can be a challenge, and so the solution is to Implement rigorous data cleaning processes, use imputation techniques, and collaborate with P2P lending platforms or financial institutions to access reliable data. Credit risk datasets often exhibit imbalances, with a minority of instances representing default cases (Khemakhem et al, 2017).

In (Marcos and Salma, 2022) they used hybrid machine learning models, by combining supervised (Adaboost, gradient boosting (GB), support vector machine (SVM), DT, RF and ANN) and unsupervised (k-Means and DBSCAN) ML models. They later compared the performance of the hybrid model to that of individual supervised ML models and deduced that the proposed hybrid model outperformed their individual supervised ML counterpart in predicting commercial customers credit score.

(Gunnarsson et al, 2021) Argued that deep learning algorithms do not seem to be appropriate models for credit scoring and so for them they had to proposed XGBoost over the other credit scoring methods only if classification performance is the main objective of credit scoring activities. One of the main reason why they proposed XGBoost over any deep learning Model is the inefficiency of most deep learning models to interpret the results and that they are "black box" in nature.

In this research paper two deep learning architectures were constructed, namely deep belief networks and multilayer perceptron networks, and compared to conventional methods for credit scoring, logistic regression and decision trees, and two ensemble methods for credit scoring, random forests and XGBoost.

The different classifiers were compared based on four performance indicators over ten data sets. This comparison highlighted the many benefits of Bayesian statistical testing procedures and secured empirical findings. The authors also suggested the use of less traditional data sources for credit scoring, e.g. unstructured data sources such as imagery or text, to enhance the richness of the considered data sources and further improve the performance of classification algorithms for credit scoring.

Imbalanced datasets can lead to biased model training, solution is to explore techniques such as oversampling, under sampling, or using advanced algorithms designed for imbalanced datasets to ensure fair representation of both classes. Comparing the performance of deep learning models with traditional credit risk models may pose challenges due to differences in interpretability and model complexity (Tse et al, 2022). Probable mitigation is to clearly define evaluation metrics, conduct thorough benchmarking, and consider both quantitative and qualitative aspects of model performance.

(Zhou et al, 2021) logistic regression model proposed provides human-interpretable information of how borrower's information and loan type affect the probability of default of loan on online P2P lending platform. (Zhou et al, 2021) Logistic regression model ensures the transparency of decision-making for loan approval and rejection which satisfy the requirements of lending.

(Moscato et al, 2021) The authors proposed a benchmarking study of some of the most used credit risk scoring models to predict if a loan will be repaid in a P2P platform. They dealt with a class imbalance problem and leverage several classifiers among the most used in the literature. Finally, the three best approaches have also been evaluated in terms of their explainability by means of different eXplainable Artificial Intelligence (XAI) tools. (Moscato et al, 2021) They also stated that future works to be devoted on improving the evaluation considering other real P2P lending platforms and to investigate more recent approaches as well as deep learning or ensemble strategies that in some cases could show better performances.

After reviewing the existing literature in relation to the proposed research topic of interest, we found several issues that needed to be addressed. In this paragraph, we will discuss some of them. First and foremost, we deduced that; there was lack of comprehensive public datasets available to use in the creation of most predicated model which continues to be an area of concern for researchers.

There was also luck of incorporating feature engineering techniques on the available datasets, which can improve the accuracy of the predication model. We also deduced that the role of deep learning models such as convolutional neural Network (CNN) may be used to extract features and the recurrent convolutional Neural network (RCNN) to be applied to enhance the prediction accuracy with respect to the sequenced Loan application in the field of credit risk defaulting hasn't really been fully expressed, and so, this calls for more research to be done relative to the intensive use of deep learning systems in enterprise.

One of the authors also suggested that a hybrid classification process focused on further features to assess a credit risk model be built. There is also a need to have techniques of comparing the performances of the methods that perform tabular data conversion into images. There is also need for government agencies to ensure that emerging economies be able to start investing in AI and machine learning. This is important because many emerging economies find it hard to make serious investments due to the costs related to these technologies. There is also a need to study the prediction of default risk in an integrative analysis framework. in that there is need to extend this to other data settings such us truncated data and censored data.

Other researchers also had limited computational power which was needed to get optimal set of hyper parameters. There is also a need to have more powerful algorithms can be implemented and try them on various sets of data, containing both numerical and categorical variables.

Finally, P2P lending data is often temporal, and capturing time dependencies is crucial for accurate credit risk prediction and so the solution is to utilize models like CNNs that are designed to capture sequential dependencies in time-series data. Implement appropriate time-series validation techniques. In this research paper we will try to produce a detailed analysis and give a comprehensive comparison among methods with the sole purpose of improving existing results

3. METHODOLOGY

Predicting credit risk defaulting in peer-to-peer (P2P) lending using deep learning involves exploring complex neural network architectures to analyze large volumes of data and extract patterns for accurate credit risk assessment.

We had to incorporate feature engineering techniques to reduce overfitting and improve the model performance. To fully utilize deep learning models for credit risk default prediction in peer-to-peer lending, feature engineering was crucial. As it's a known fact, applying feature engineering techniques, can enhance the predictive power of ML / DL models and improve its accuracy.

They can also improve model performance, interpretability, and resilience by creating characteristics that are both relevant and informative. From data collection to model building, evaluation, and ethical issues, this research design offers an organized approach for carrying out the study.

Researchers can modify this design in accordance with the details of their investigation and the attributes of the P2P lending data they are utilizing.

The DFD below shows the activities that will be undertaken to accomplish the task at hand, Depicting Data preprocessing with feature engineering techniques

By following this proposed methodology, we could systematically address key aspects of credit risk default prediction in P2P lending using deep learning models.

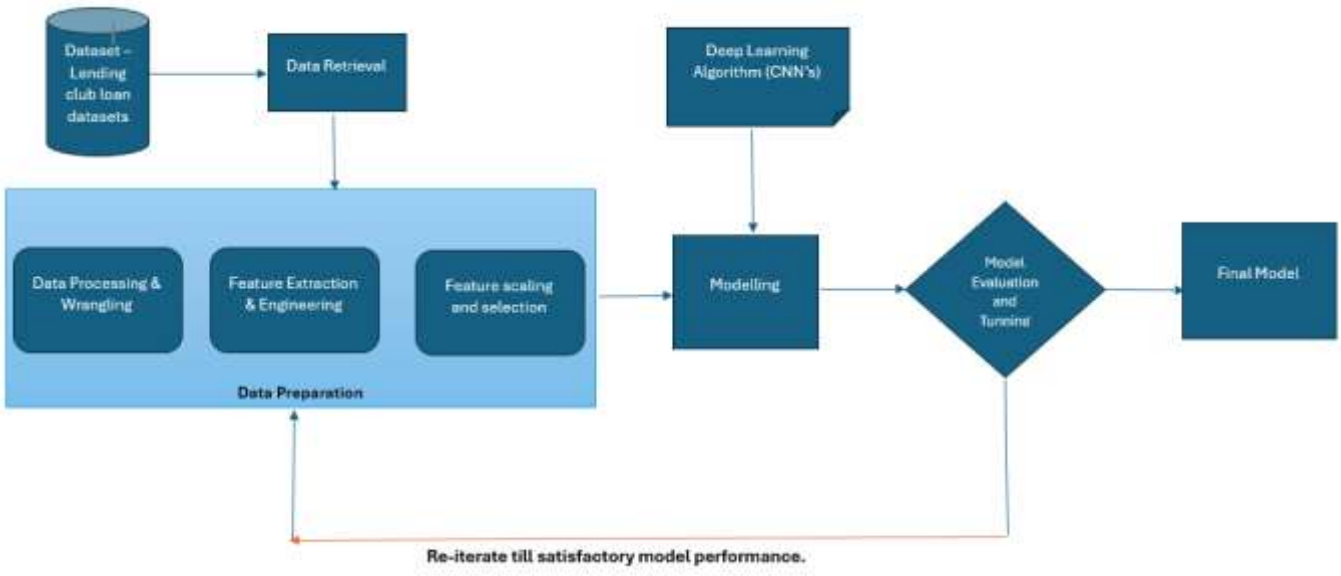

Figure 1: DFD Depicting Data preprocessing with feature engineering techniques.

4. EXPERIMENTS

Convolutional Neural Networks (CNNs) is one type of deep learning technique which has demonstrated promising results in enhancing predictive accuracy and capturing intricate patterns in borrower data when used to focus on credit risk defaulting in peer-to-peer (P2P) lending. The below stated steps were undertaken in predicting credit risk defaulting using CNNs and ANNs in peer-to-peer lending. Please note that the task was implemented in google colab. Google Colab, short for Google Collaboratory, is a cloud-based platform provided by Google that allows users to run and execute Python code in a browser environment. It is particularly popular among data scientists, researchers, and developers for its ease of use, free access to GPU and TPU resources, and collaboration features. Below are the six (6) main steps that were undertaken to accomplish the task at hand.

1. We started by Loading a CSV file dataset from Kaggle using Pandas.
2. Secondly, we applied feature engineering techniques so as to improve on the dataset downloaded from Kaggle.
3. We then created three groups of data which were train, validation, and test sets.
4. After which we defined and trained a model using Kera's (also included setting class weights).
5. We then Evaluated the model using various metrics (which included: Accuracy, Precision, F1 and recall).
6. Selected a threshold for a probabilistic classifier to get a deterministic classifier.
7. And finally we compared with class weighted modelling and oversampling

4.1 Data Exploration and Preprocessing

Feature Engineering: We had to Identify relevant features such as borrower demographics, loan attributes, repayment history, credit utilization, etc. and transformed the raw data into suitable formats for our proposed deep learning models.

- **Data Normalization:** We also had to scale and normalize numerical features like monthly income to ensure uniformity and enhance model training efficiency.
- **Handling Missing Data:** Statistical techniques such as the mean, median, or mode for numerical data were used to fill in the missing values. In certain instances, we utilized a unique category named "Unknown" in place of the most common categories when it came to classifying features.
- **Flagging:** To show whether a value was missing, we created binary features, which in turn assisted the model in identifying patterns linked to missing data.
- **New Features Created:**
    1. A New feature was created by determining the credit utilization ratio this was done by dividing the total credit limit by the person's income.
    2. Another feature was introduced by performing the Debt-to-Income Ratio which was done by computing the ratio of total monthly debt payments to monthly income.
    3. Delinquency Features were also created by creating features indicating the number of delinquent accounts or the time since the last delinquency.
    4. Credit Age was another feature that was introduced, this was derived by the number of years since the borrower first took credit.
- **Binning and Discretization:**

We had to create what is referred to as age bins, to do this we had to put ages in group ranges (e.g. 18 – 25, 26 – 35 etc.). We also did Income and Loan Amount Binning were continuous variables like income

and loan amounts were put into categories (e.g., low, medium, high) so we could capture non-linear relationships.

- **Behavioral Features:**

1. Payment History: we included features such as the proportion of on-time payments, late payments, and the maximum lateness.
2. Loan Purpose: we had to analyze and categorize the purpose of the loan, which might correlate with the default risk.

- **External Data Sources:**

we also looked at other external data sources like economic indicators were we had to incorporate external economic indicators such as unemployment rates, inflation rates and growth of GDP, which in turn affected the credit risk. The other external data source we looked at was location-based data where we used geographic information which included regional economic conditions.

- **Text features:**

this was another feature that we explode to improve the predictability of the model. Two items were looked at in this scenario, namely Loan Description and Employment Title. Regarding the loan description, we made use of the Natural Language Processing (NLP) techniques to extract sentiment and key topics. While on the employment title, we had to extract and encode information from the borrower's employment title. We began by utilizing pandas to import significant libraries into the colab. It is impossible to exaggerate the significance of Pandas in Google Colab when it comes to data processing and analysis in Python. Pandas are widely available and pre-installed in Google Colab, facilitating users' utilization of its functions without requiring extra setup.

Whether importing information from Google Drive, doing database queries, or interacting with other datasets, Pandas makes data processing and analysis chores easier, which makes it a vital tool in Colab for researchers, analysts, and data scientists (Si Shi et al, 2022).

4.2 Clean, Normalize / split the imported Data:

The data imported had three (3) major issues. The time and amount column were too vague to be used in that manner. In that, the time had no significance or meaning at the time. secondly the amount column was to wide and so we had to log it out to reduce its range, and finally the loan_status column had three main instance, namely fully paid, defaulted and ongoing.

We then had to transform the target variable loan_status to numerical by giving the loan_status valued ‘fully paid’ a value of ‘0’ while those valued ‘defaulted’ a value of ‘1’. We had to divide the dataset into training, validation, and test datasets after it had been cleaned. Before dividing the data samples into training, validation, and test sets, we had to make sure they were randomly shuffled. This helped prevent any inherent ordering or biases in the data from affecting the model's performance.

The validation set was used during the model fitting to evaluate the loss and any metrics. The test dataset was used to evaluate the final performance of the trained model after hyper parameter tuning and model selection.

In our case, the test set was used only at the very end to assess how well the model generalizes to new data, having been completely ignored during the training process. Most of our data, known as the training dataset, was utilized to train the model's parameters (weights and biases) during the optimization procedure. This was crucial for imbalanced datasets since the scarcity of training data poses a serious risk of overfitting.

After the splits were done – we had to verify whether they were of the same distribution. As can be seen from the above statements – this seemed about right. We then had to normalize the input features using sklearn Standard Scaler, by setting the mean to 0 and standard deviation to 1. At the end of it - all we wanted to predict was a class label 0 or 1, no default or default respectively. This is also known as deterministic classifier. To get a label prediction from our probabilistic classifier, one needs to choose a probability threshold T. The default is to predict label 1 (default) if the predicted probability is larger than threshold T = 50%. The below metrics implicitly make use of this probability.

- False negatives and false positives are samples that were incorrectly classified.
- True negatives and True positives are samples that were correctly classified.
- Accuracy is the percentage of examples correctly classified > True samples / Total samples.
- Precision is the percentage of predicated positives that were correctly classified > True positives / True positives + False positives (Apostolos et al, 2021).
- (Peter et al, 2018) Recall is the percentage of actual positives that were correctly classified > True positives / True positives + False negatives.

The Accuracy metric was not of help for the task in question. Reason being we could have 99.8 % + accuracy by predicting False all the time. The following are other metric's that we considered that also looked at all possible choices of thresholds:

- (Marc et al 2022) AUC refers to the Area Under the Curve of a Receiver Operating Characteristic curve (ROC-AUC). This metric is equal to the probability that a classifier will rank a random positive sample higher than a random negative sample.
- AUPRC refers to Area Under the Curve of the Precision-Recall Curve. These metric computes precision-recall pairs for different probability thresholds (Apostolos et al, 2021).

4.3 Deep Learning Models:

- Convolutional Neural Networks (CNN): We applied CNN architectures for feature extraction from structured dataset related to credit risk prediction, borrower documents and financial statements.
- We also incorporated attention mechanisms in our proposed model to focus on important features and prioritize relevant information during credit score default prediction.
- Auto-encoder architectures were used for feature learning and dimensionality reduction, which improved model performance and generalization

4.4. Building The Model:

We started by creating a 1D CNN architecture and we used pooling layers after a few convolutional layers to extract features. Added dense (fully connected) layers with appropriate activation functions for classification. We also Included dropout layers to reduce on overfitting. To guarantee that each batch had a reasonable probability of having a few positive samples, we made sure the model was fitted using a larger than normal batch size of 2048. We would have been less likely to have no default transactions to draw lessons from if we had chosen a smaller batch size. Additionally, we had to assemble the CNN model with the proper optimizer (Adam), metrics (accuracy, AUC-ROC), and loss function (binary cross-entropy in our instance for binary classification).

4.5 Model Optimization:

- Hyper parameter Tuning: we had to optimize hyper-parameters such as learning rate, batch size, number of layers, activation functions, dropout rates, etc., to improve model convergence and performance.

- Regularization Techniques: We Applied regularization techniques like L1 and L2 regularization, dropout regularization, and batch normalization to prevent overfitting and enhance model robustness.
- Finally, we employed evaluation metrics such as accuracy, precision, recall, F1-score, ROC-AUC, and calibration curves to assess model performance and validate credit risk score default predictions.

4.6. Return Model:

After testing the model, we figured that the output wasn't what we had expected, and so we had to do some tuning, for this we used validation dataset for hyper parameter tuning and early stopping to prevent overfitting. We continued to monitor training/validation loss and accuracy during the training. The correct bias was derived from the below:

$$p_0 = pos/(pos + neg) = 1/(1 + e^{-b_0})$$

$$b_0 = -log_e(1/p_0 - 1)$$

$$b_0 = log_e(pos/neg)$$

Before proceeding we had to confirm that the careful bias initialization helped. To do so we had to train the model for 30 epochs, with and without this careful initialization and there after compare the losses:

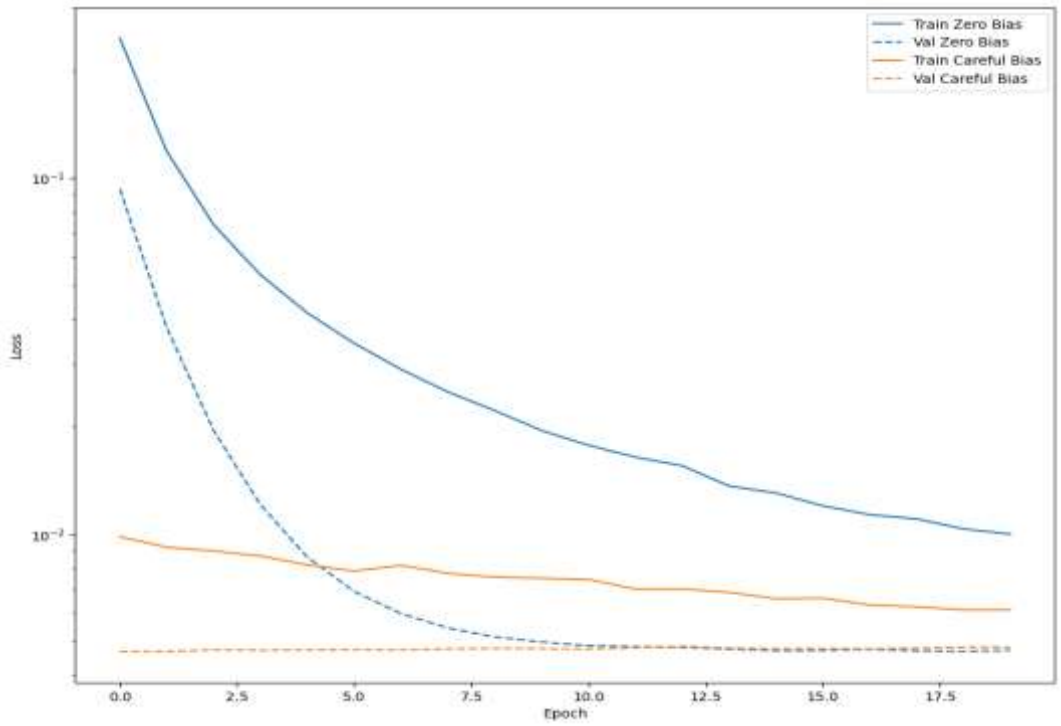

Figure 2 - showing careful initialization.

Looking at the figure above – it was proof that: careful initialization gave us a clear advantage with regards to validation loss.

4.7.Training The Model:

A Finally, we had to train the model. After which, we had to check the training history of our model. This was achieved by producing plots of our model's accuracy and loss on the training and validation set. This was done with the sole purpose of checking for overfitting.

In a nutshell, the below steps were undertaken:

- Data was collected.
- Applied feature engineering techniques on the cleaned data.
- Trained and tested the two deep learning models (ANNs and CNNs) on the preprocessed data, using a variety of different hyper parameter settings to determine which model performed better.
- Model comparison and Evaluation was done using on the same dataset, using metrics like accuracy, precision, and recall.

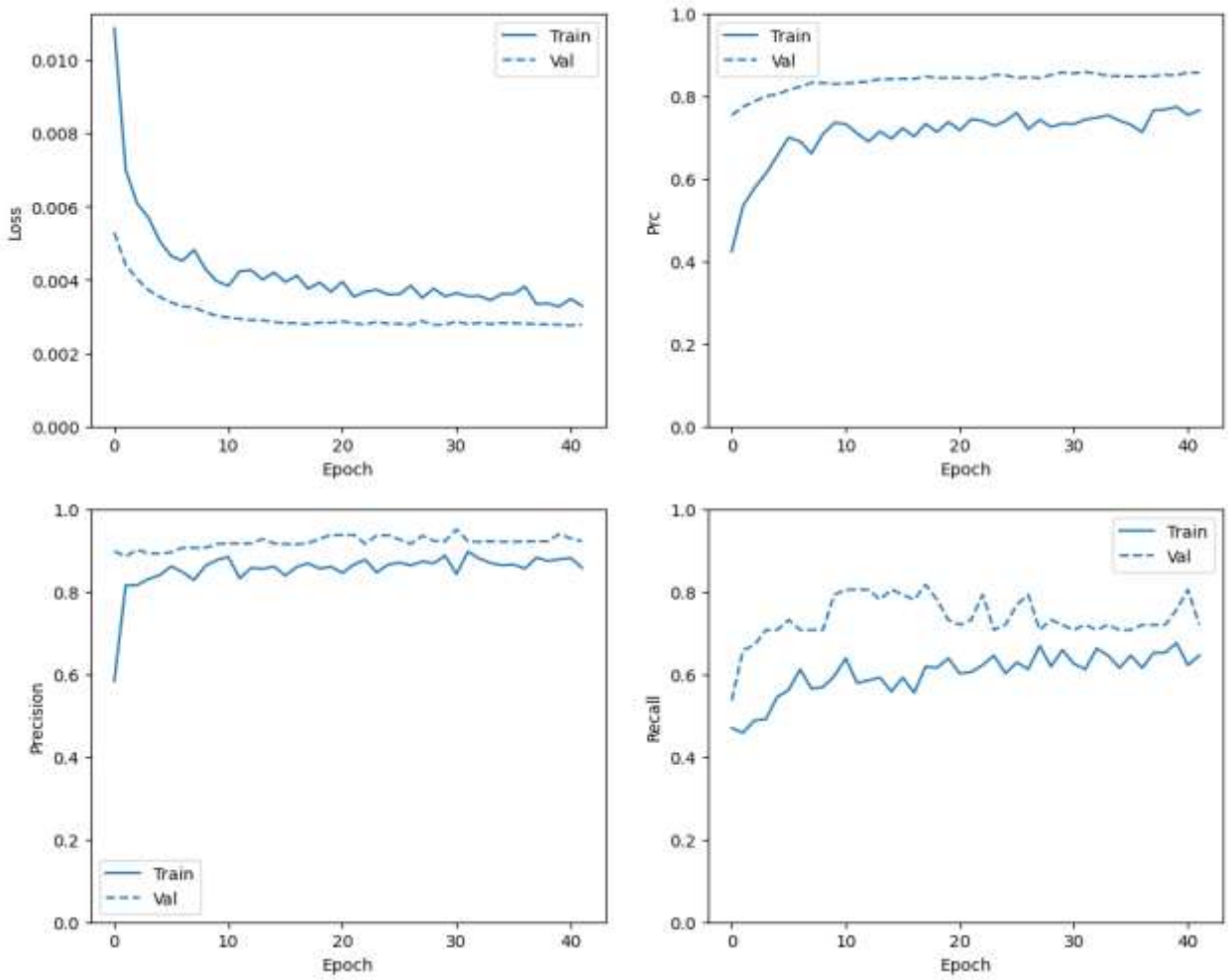

Figure 3 - Training and validation.

5. FINDINGS

In conclusion we can safely say that, on average, the validation curve outperforms the training curve. which is mostly the result of the dropout layer not being active throughout the model's evaluation. In order to determine the trained CNN model's performance metrics—such as accuracy, precision, recall, F1-score, and AUC-ROC—we had to test it on the test set. See below the actual outputs given:

loss: 0.0038855739403516054
cross entropy: 0.0038855739403516054
Brier score: 0.0006162827485240996
tp : 81.0
fp : 11.0

tn : 56840.0
fn : 30.0
accuracy: 0.9992802143096924
precision: 0.8804348111152649
recall: 0.7297297120094299
AUC: 0.9096326231956482
PRC: 0.7863917350769043
Legitimate Transactions Detected (True Negatives): 56840
Legitimate Transactions Incorrectly Detected (False Positives): 11
Defaulted Transactions Missed (False Negatives): 30
Defaulted Transactions Detected (True Positives): 81
Total Defaulted Transactions: 111

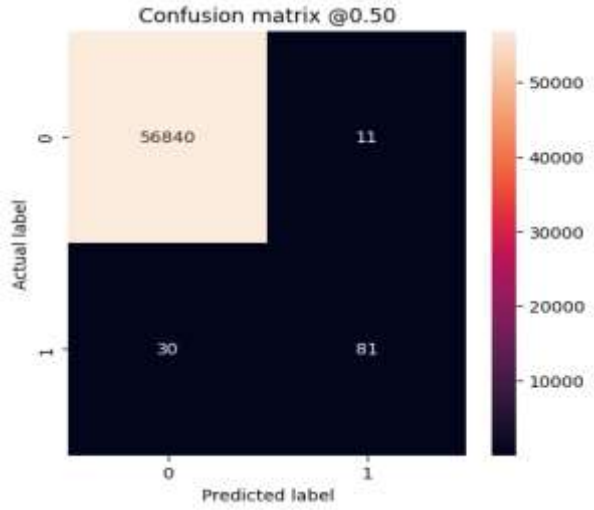

Figure 4 - Confusion Matrix @ 0.50

5.1 Changing The Threshold:

The default threshold of T = 50% corresponds to equal costs of false negatives and false positives. In the case of loan default detection, however, you would likely associate higher costs to false negatives than to false positives. (Si Shi et al, 2022) This trade off may be preferable because false negatives would allow fraudulent transactions to go through, whereas false positives may cause an email to be sent to a customer informing them that their loan has been approved when in fact it's a false positive. By decreasing the threshold, we attribute higher costs to false negatives, thereby increasing missed transactions at the price of more false positives. We test thresholds at 10% and at 1%. See below sample code doing exactly just that:

plot_cm(test_labels,test_predictions_baseline, threshold=0.1)
plot_cm(test_labels,test_predictions_baseline, threshold=0.01)
Legitimate Transactions Detected (True Negatives): 56834
Legitimate Transactions Incorrectly Detected (False Positives): 17
Defaulted Transactions Missed (False Negatives): 23
Defaulted Transactions Detected (True Positives): 88
Total Defaulted Transactions: 111

Legitimate Transactions Detected (True Negatives): 56806
Legitimate Transactions Incorrectly Detected (False Positives): 45
Defaulted Transactions Missed (False Negatives): 22
Defaulted Transactions Detected (True Positives): 89
Total Defaulted Transactions: 111

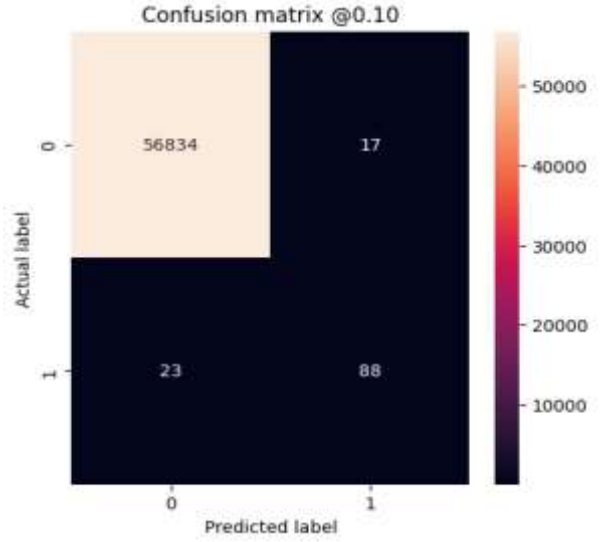

Figure 5 - Confusion Matrix @ 0.10

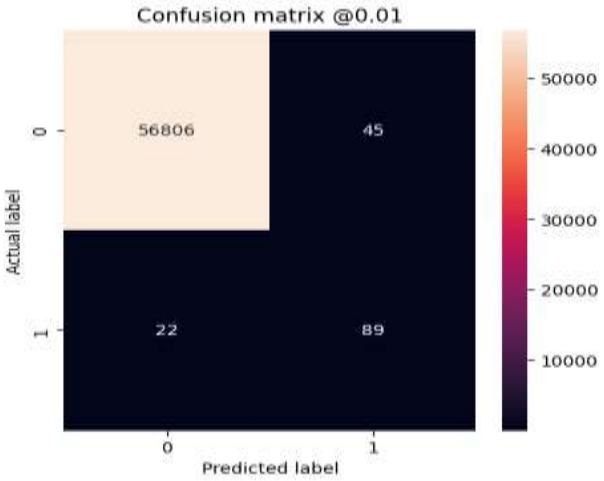

Figure 6 - Confusion Matrix @ 0.01

5.2 Plot The ROC:

To plot the Receiver Operating Characteristic (ROC) curve in order to assess a classification model's performance. Because it illustrates the model's reach when the output threshold is adjusted across the entire range of 0 to 1, this figure is helpful. See below:

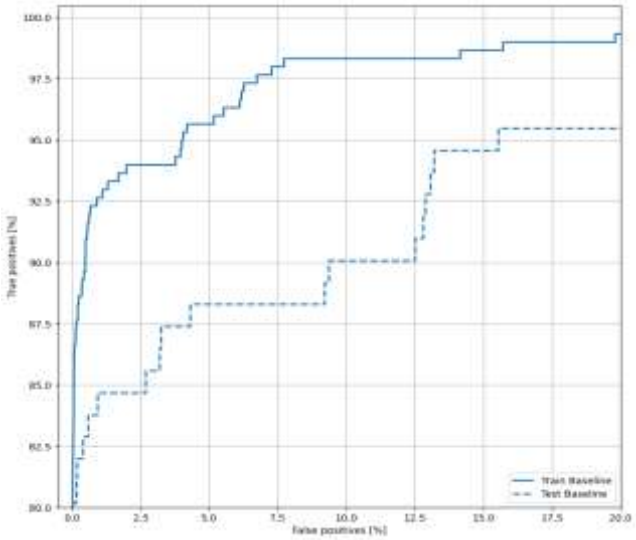

Figure 7 - Tuning Output threshold over its full range (0 to 1)

5.3 Plotting The PRC:

To plot the Area Under the Precision-Recall Curve (AUPRC), we followed a similar process as plotting the ROC curve. However, instead of using the ROC curve metrics, we instead used the Precision-Recall curve metrics and the average_precision_score function from Scikit-Learn to compute the AUPRC score. The Precision-Recall curve is particularly useful for evaluating models on imbalanced datasets or tasks where the focus is on positive class prediction accuracy and minimizing false positives.

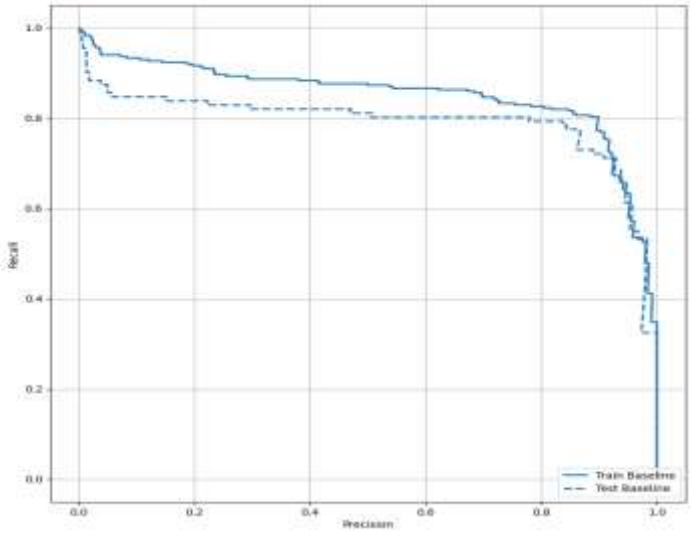

Figure 8 - The Precision-Recall curve

5.4 Class Weights on a Model:

Training a deep learning model with class weights is essential, especially when dealing with imbalanced datasets where one class (e.g., positive class, minority class) is significantly underrepresented compared to the other class (e.g., negative class, majority class) (Ampountolas et al, 2021). Class weights help to address this imbalance by assigning higher weights to the minority class during training, thereby giving the model more emphasis on learning from the minority class examples (Reddy et al,2020). We concluded

that with class weights the accuracy and precision were lower because there were more false positives, but conversely the recall and AUC were higher because the model also found more true positives. Despite having lower accuracy, this model had higher recall (and identifies more fraudulent transactions than the baseline model at threshold 50%).

By training our deep learning model with class weights, we gave more importance to the minority class examples, which can lead to better performance and improved generalization on imbalanced datasets (Sahem et al, 2017). Based on the unique distribution of classes in our dataset, we modified the class weights. To get the best outcomes, we also experimented with various model architectures and hyper parameters.

6. DISCUSSION

Improving the model's interpretability can also involve adding significant features. It is simpler to justify the model's predictions to stakeholders and regulatory agencies when features are engineered to be simply understood and to exhibit strong correlations with the target variable. Effective feature engineering can assist in removing pertinent information from unprocessed data and producing new features that improve the deep learning model's capacity for prediction. More precise credit risk assessment and defaults may result from this.

Deep learning models such CNNs have shown effectiveness in improving credit risk default prediction accuracy in P2P lending platforms. These deep learning models are suitable for capturing temporal dependencies and sequential patterns in borrower data, such as changes in income, spending habits, and repayment behavior over time. CNNs excel at extracting spatial features and patterns from structured data, images, or documents related to credit applications and borrower profiles.

The proposed models facilitate automatic feature learning, eliminating the need for manual feature engineering and enhancing the ability to handle diverse and complex borrower attributes. Compared to traditional statistical models, the proposed deep learning techniques have demonstrated improved accuracy in credit scoring and risk assessment in P2P lending environments. Data augmentation techniques have been investigated to enhance the performance and robustness of deep learning models by increasing the diversity and quality of the training dataset.

The research findings indicate that deep learning techniques, particularly CNNs with feature engineering, hold promise for enhancing credit risk defaulting prediction accuracy and aiding in risk assessment in P2P lending platforms. These models can capture complex patterns, temporal dependencies, and spatial features in borrower data, leading to more informed credit decisions. Therefore, we can deduce that our proposed model shows its robustness in modeling the default rate of P2P lending market, whether certain macroeconomic features were present or not. Some of these macroeconomic features include the monthly unemployment rate, GDP etc. our inference provides a good reference for both investors and potential borrowers to understand the entire system of P2P lending, especially the default rate on an aggregative level. This is very critical in making their future investment plans.

7. CONCLUSION

Despite their effectiveness, challenges such as model interpretability, explainability, data privacy, and regulatory compliance remain important considerations in deploying deep learning-based credit scoring systems. Continued research and advancements in model development, data quality, and interpretability are essential to address these challenges and further improve the effectiveness of deep learning techniques

in real-world P2P lending applications. Overall, the research in predicting credit risk defaulting in P2P lending using deep learning showcases significant potential for enhancing credit risk assessment, but ongoing efforts are needed to overcome challenges and ensure the responsible and effective deployment of these advanced models in practice. As a future enhancement, researchers should focus on establishing mechanisms for continuous learning and model updates based on new data, changing borrower behavior patterns, regulatory updates, and market dynamics to ensure ongoing model relevance and accuracy in credit score prediction. These research methods demonstrate the application of deep learning techniques in predicting credit scores for P2P lending platforms, emphasizing the importance of data preprocessing, model selection, optimization, and interpretability for effective credit risk assessment and decision-making.

8. REFERENCES

1. D. Wang, Y. Yu, Q. Yang, and X. Niu, ‘‘Study on a prediction of P2P network loan default based on the machine learning LightGBM and XGboost algorithms according to different high dimensional data cleaning,’’ Electron. Commerce Res. Appl., vol. 31, pp. 24 Sep. 2018.
2. Marcos Roberto Machado, Salma Karry (2022) “Assessing credit risk of commercial customers using hybrid machine learning algorithms”. Volume 200, Expert systems with Applications Goodfellow I, Bengio Y, Courville A (2016) Deep Learn. MIT press, Cambridge
3. Kavesh Bastani, Elham Asgari, Hamed Asgari ‘‘Wide and deep learning for P2P lending prediction of P2P network loan default based on the machine learning Light-GBM and XGboost algorithms according to different high dimensional data cleaning,’’ Expert Systems. Appl., vol. 134, pp. 209 - 224 Jan. 2019.
4. Van-Sang Ha*, Dang-Nhac Lu*, Gyoo Seok Choi, Ha-Nam Nguyen, Byeongnam Yoon, “Improving Credit Risk Prediction in Online Peer-to Peer (P2P) Lending Using Feature selection with Deep learning” pp. 511 - 515 Feb. 2017. Department of Economic Information System, Academy of Finance, Hanoi, Vietnam Academy of Journalism and Communication, Vietnam
5. Xolani Dastile and Turgay Celik, ‘‘Making Deep Learning-Based Predictions for Credit Scoring Explainable,’’ School of Computer Science and Applied Mathematics, Vol 9 , March. 2021.
6. Rahim Khan (2023) “Online Loan Default Prediction Model Based On Deep Learning Neural Networks”. School of Statistics and Big Data, Henan University Of Economics and Law, Zhengzhou 450046 Henan China.
7. Chong Pei Swee, Farid Meziane, and Jane Labadin (2022) “Credit Risk Prediction for Peer-To-Peer Lending Platforms: An Explainable Machine Learning Approach”. Faculty of Computer Science and Information Technology, Universiti Malaysia Sarawak, 94300 Kota Samarahan, Sarawak, Malaysia.
8. Shuhui Chen, Qing Wang and Shuan Liu (2019) “Credit Risk Prediction in Peer-to-Peer Lending with Ensemble Learning Framework”. College of Information Science and Engineering, Northeastern University, Shenyang 110819, China
9. James Heckman (1977) Sample Selection Bias As a Specification Error (with an Application to the Estimation of Labor Supply Functions): Econometrica, Vol 47 No 1
10. Vinod Kumar L, Natarajan S and Keerthana S (2016) Credit Risk Analysis in Peer-to-Peer Lending System. IEEE International Conference on Knowledge Engineering and Applications
11. Apostolos Ampountolas, Titus Nyarko Nde, Paresh Date and Corina Constantines (2021) “A Machine Learning Approach for Micro-Credit Scoring”. Department of Mathematics, College of Engineering,

Design and Physical Sciences
12. Bhatore S, Mohan L, Reddy YR (2020) Machine learning techniques for credit risk evaluation: a systematic literature review. J Bank Financ Technol 4(1):111–138.
13. Sihem Khemakhem, Younes Boujelbene, Fatma Ben Said (2017) Credit risk assessment for unbalanced datasets based on data mining, artificial neural network, and support vector.
14. Si Shi, Rita Tse, Wuman Luo, Stefano D'Addona and Giovanni Pau (2022) Machine learning- driven credit risk: a systemic review. Neural Computing and Applications Springer.
15. Ligang Zhou a , Hamido Fujita, Hao Ding, Rui Ma "Credit risk modeling on data with two timestamps in peer-to-peer lending by gradient boosting" Vol 110 June. 2021. School of Business, Macau University of Science and Technology, Taipa, Macau
16. Chong Pei Swee, Farid Meziane, and Jane Labadin "Credit Risk Prediction for Peer-To-Peer Lending Platforms: An Explainable Machine Learning Approach" Vol 1 number 2, September, 2022. 3Faculty of Computer Science and Information Technology, Universiti Malaysia Sarawak, 94300 Kota Samarahan, Sarawak, Malaysia School of Computing and Engineering, University
17. machines. J Faculty of Economics and Management of Sfax, University, Tunisia P 932 – 951
18. Miller Janny Ariza-Garzón, Javier Arroyo, Antonio and Maria-Jesus Segovia-Vargas (2020) "Explainability of a Machine Learning Granting Scoring Model in Peer-to-Peer Lending". Santander-UCM Research Project, Call 2019, under Grant PR87/19-22586
19. Linn´ea Machado and David Holmer (2022) "Credit risk modelling and prediction: Logistic regression versus machine learning boosting algorithms". (Advisor Tatjana Pavlenko) Björn Rafn Gunnarsson, Seppe vanden Broucke, Bart Baesens, María Óskarsdóttir and Wilfried Lemahieu (2021) "Deep learning for credit scoring: Do or don't". Volume 200, Expert systems with Applications. Department of Computer Science, Reykjavík University, Menntavegi 1, Reykjavík 101, Iceland.
20. Nikita Kozodoi, Johannes Jacob and Stefan Lessmann (2022) "FAIRNESS IN CREDIT SCORING: ASSESSMENT, IMPLEMENTATION AND PROFIT IMPLICATIONS". Source https://www.federalreserve.gov/releases/g19/current. Source Code: https://github.com/kozodoi/Fair_Credit_Scoring
21. Yujia Chen , Raffaella Calabrese, Belen Martin-Barragan (2024) "Interpretable machine learning for imbalanced credit scoring datasets". European Journal of Operational Research Volume 312, Issue 1, 1 January 2024, Pages 357-372
22. Bertrand K. Hassani (2020) "Societal bias reinforcement through machine learning: a credit scoring perspective". AI and Ethics (2021) 1:239–247 https://doi.org/10.1007/s43681-020-00026-z
23. Simon Williams (2021) "Credit Card Score Prediction Using Machine Learning". M. Sc IT (Artificial Intelligence) Department of Information Technology University of Mumbai, India. Volume 6, Issue 5, May – 2021 International Journal of Innovative Science and Research Technology.
24. David Mhlanga (2021) "Financial Inclusion in Emerging Economies: The Application of Machine Learning and Artificial Intelligence in Credit Risk Assessment". Department of Accountancy, The University of Johannesburg, Johannesburg 2006, South Africa; dmhlanga@uj.ac.za or dmhlanga67@gmail.com. International Journal of Financial Studies 9: 39. https:// doi.org/10.3390/ijfs9030039
25. Marc Schmitt (2022) "Deep Learning vs. Gradient Boosting: Benchmarking state of-the-art machine learning algorithms for credit score". Department of Computer Science, University of Oxford, UK. Department of Computer & Information Sciences, University of Strathclyde, UK.

26. Peter Martey Addo, Dominique Guegan and Bertrand Hassani (2018) "Credit Risk Analysis Using Machine and Deep Learning Models". Risks 2018, 6, 38; doi:10.3390/risks6020038 www.mdpi.com/journal/risks. Direction du Numérique, AFD—Agence Française de Développement, Paris 75012, France Laboratory of Excellence for Financial Regulation (LabEx ReFi), Paris 75011, France; dominique.guegan@univ-paris1.fr (D.G.); bertrand.hassani@gmail.com (B.H.)
27. Vincenzo Moscato, Antonio Picariello and Giancarlo Sperlí (2021) "A benchmark of machine learning approaches for credit score prediction". Department of Electrical Engineering and Information Technology (DIETI), University of Naples ''Federico II", Via Claudio 21, Naples, Italy. Expert Systems with Applications 165 (2021)
28. Feng Shen, Xingchao Zhao, Gang Kou, Fawaz E. Alsaadi (2020) "A new deep learning ensemble credit risk evaluation model with an improved synthetic minority oversampling technique". Applied Soft Computing Journal (2020), doi: https://doi.org/10.1016/j.asoc.2020.106852. School of Finance, Southwestern University of Finance and Economics, Chengdu 611130, PR China.
29. Cheng-Feng Wu, Shian-Chang Huang, Chei-Chang Chiou, Yu-Min Wang (2021) "A predictive intelligence system of credit scoring based on deep multiple kernel learning". School of Business Administration, Hubei University of Economics, Wuhan, Hubei Province, China. School of Business, Wuchang University of Technology, Wuhan, Hubei Province, China
30. Lu Wang, Wenyao Zhang (2023) "A qualitatively analyzable two-stage ensemble model based on machine learning for credit risk early warning: Evidence from Chinese manufacturing companies". Information Processing & Management Volume 60, Issue 3, May 2023, 103267
31. Siddharth Bhatore, Lalit Mohan· Y. Raghu Reddy (2020) "Machine learning techniques for credit risk evaluation: a systematic literature review". Institute for Development and Research in Banking Technology 2020.
32. Yuelin Wanga, Yihan Zhanga, Yan Lua, Xinran Yua (2020) "A Comparative Assessment of Credit Risk Model Based on Machine Learning - A case study of bank loan data". Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/) Peer-review under responsibility of the scientific committee of the 2019 International Conference on Identification, Information and Knowledge in the Internet of Things.
33. Niklas Bussmann, Paolo Giudici, Dimitri Marinelli and Jochen Papen Brock "Explainable Machine Learning in Credit Risk Management" Computational Economics (2021) 57:203–216 https://doi.org/10.1007/s10614-020-10042.
34. Diwakar Tripathi, Damodar Reddy Edla, Venkatanareshbabu Kuppili and Annushree Bablani (2020) "Evolutionary Extreme Learning Machine with novel activation function for credit scoring" SRM University AP-Amaravati, Andhra Pradesh 522502, India. National Institute of Technology Goa, Ponda, Goa 403401, India
35. Ahmed Almustfa Hussin Adam Khatir and Marco Bee (2022)"Machine Learning Models and Data-Balancing Techniques for Credit Scoring: What Is the Best Combination?" Department of Economics and Management, University of Trento, Via Inama 5, 38122 Trento, Italy Correspondence: marco.bee@unitn.it
36. Johannes Kriebel and Lennart Stitz (2022)"Credit default prediction from user-generated text in peer-to-peer lending using deep learning" European Journal of Operational Research, Volume 302, Issue 1, 1 October 2022, Pages 309-323. https://doi.org/10.1016/j.ejor.2021.12.024

37. Wei Li, Shuai Ding, Hao Wang, Yi Chen and Shanlin Yang (2019)"Heterogeneous ensemble learning with feature engineering for default prediction in peer-to-peer lending in China" Springer Science+Business Media, LLC, part of Springer Nature 2019.
38. Cunjie Lin, Nan Qiao, Wenli Zhang, Yang Li and Shuangge Ma (2022)"Default risk prediction and feature extraction using a penalized deep neural network" Under exclusive license to Springer Science Business Media, LLC, part of Springer Nature 2022.
39. ZAIMEI ZHANG, KUN NIU AND YAN LIU (2020)"A Deep Learning Based Online Credit Scoring Model for P2P Lending" College of Economics and Management, Changsha University of Science and Technology, Changsha 410114, China.