International Journal of Computer Science and Information Technology Research (IJCSITR)
Vol. 6, No. 4, July - August, 2025, pp. 50-69
Journal ID: 9471-1297
website: www.ijcsitr.com
DOI: https://doi.org/10.63530/IJCSITR_2025_06_04_004

Self-Supervised Learning Approaches for Credit Data Representation and Risk Stratification

Santhosh Kumar Sagar Nagaraj

Staff Software Engineer, Visa Inc, Banking & Finance, 1745 stringer pass, Leander, Texas 78641, USA.

Abstract

The complexity and volume of financial data are also rising and require more powerful methods and more intelligent approaches to successful credit risk evaluation. The weakness of many traditional supervised learning algorithms is that they are sensitive to the availability of labelled data, which is scarce, skewed, and costly to acquire in the credit setting. The paradigm of self-supervised learning (SSL), which relies on supervision by the data in its most basic form, has become an attractive alternative option, particularly where labelled data is scarce. This article examines progressive, self-guided instructions for illustrating credit information and risk classification. This is done by creating a system that can create meaningful embeddings of the credit profile of customers using pretext tasks like masked feature prediction, contrastive learning, and autoencoding. Thereafter, the downstream credit risk prediction tasks are performed using such embeddings. We do substantial experimentation on real-world data sets of credit, and compare our models to classic supervised approaches. We have shown that self-supervised models can show a similar or even better performance in credit risk stratification, especially in the setting where there is a limited number of labelled data available. Moreover, we debate the interpretability, deterioration, and generalization abilities of SSL-based models in financial use cases. We also give an insight into how different Tasks implemented in SSL and architecture options affect the goodness of representations learned. The paper ends with a debate on how self-supervised learning will transform risk management in financial services by helping them create fairer, precise, and efficient credit rating models.

Keywords

Self-supervised learning, credit risk, data representation, contrastive learning, autoencoder, masked feature prediction, credit scoring, risk stratification, deep learning.

How to Cite: Santhosh Kumar Sagar Nagaraj. (2025). Self-Supervised Learning Approaches for Credit Data Representation and Risk Stratification. *International Journal of Computer Science and Information Technology Research (IJCSITR)*, 6(4), 50-.
DOI: https://doi.org/10.63530/IJCSITR_2025_06_04_004

Article ID: IJCSITR_2025_06_04_004

1. Introduction

The role of credit risk assessment in the field of finance is extremely vital, as it enables institutions to determine the likelihood of default if a borrower fails to repay a loan or other financial service. An accurate risk forecast allows a lender to make informed decisions, control risk in the portfolio, and also comply with regulatory requirements. [1-3] In the past, supervised learning models (which demand high amounts of labelled data, i.e. historical data annotated with the performance of the borrower, i.e. default or repayment) were used extensively in this task. Nonetheless, high-quality labelled credit data is, in practice, difficult to acquire in several aspects. This kind of data can be quite scarce, unbalanced, and unavailable to attain, mainly because of privacy restrictions, regulations, and the cost of annotation inherent in it. The defaults are usually much fewer than the non-defaults; hence, class imbalance results in poor model performance. Such limitations may indicate the importance of developing alternative strategies that can learn meaningful and useful representations of credit data without relying heavily on labelled samples. Self-supervised learning (SSL) emerges as a potential remedy, as the structure present in unlabeled data can be leveraged to yield pseudo-supervision in the form of a pretext task. In this manner, we can discover patterns, correlations, and underlying relationships in the credit data, which we can later apply to downstream tasks such as credit scoring and risk stratification using SSL models. The rationale behind the study is related to

the potential ability of SSL to address the shortcomings of conventional method-based approaches and enable a more precise credit risk analysis in data-limited and real-world finance settings.

1.1. The Rise of Self-Supervised Learning (SSL)

Self-supervised learning (SSL) has emerged as a promising paradigm in machine learning, particularly when labelled data is scarce, costly, or time-consuming to obtain. In contrast to conventional supervised learning, where supervision is encoded into manually labelled data, SSL induces supervision on the data itself through the use of intelligently formulated pretext tasks. Through these tasks, models can learn meaningful features without relying on human-labelled inputs, making SSL particularly attractive in real-world scenarios with scarce labelled data, as illustrated in the use cases of finance, healthcare, and cybersecurity.

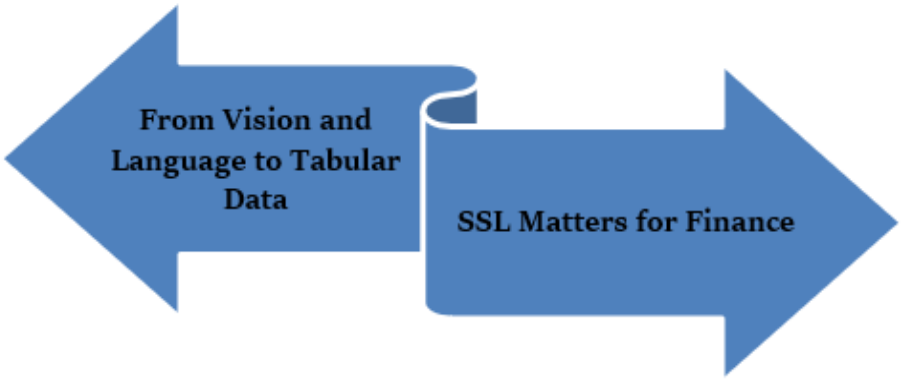

Figure 1: Rise of Self-Supervised Learning (SSL)

- **From Vision and Language to Tabular Data:** The first application of SSL had a head start in computer vision and natural language processing (NLP). Models such as SimCLR, MoCo, and BERT have demonstrated that deep neural networks can learn powerful, general-purpose representations by simply solving tasks, including predicting masked words or distinguishing between augmented pairs of images. The progress made the use of labelled data much less crucial and gave opportunities to transfer learning between tasks and job fields. Encouraged by these successes, researchers have recently begun to investigate the idea of SSL in structured and tabular data, which is a typical case in the financial setting. This shift will be an influential step towards training state-of-the-art learning methods for other domains,

such as credit scoring, fraud detection, and customer segmentation.

- **SSL Matters for Finance:** SSL has the following benefits in the financial arena. It first enables models to utilise large quantities of unlabeled transactional or credit data that can be gathered more easily than labelled examples of defaults or delinquencies. Second, it facilitates data efficiency, as we do not need to carry out large and costly human labelling. Third, SSL frameworks can be more efficient in increasing model robustness and generalization, allowing systems to be reliable in dynamic market environments or when used in other institutions. Lastly, SSL facilitates the transfer and reuse of models, since with very limited labelled data, one can fine-tune the pre-trained encoder on a particular credit risk task.

1.2. Approaches for Credit Data Representation and Risk Stratification

The age of credit data representation and risk stratification is a key building block of modern credit risk assessment systems. Such processes entail converting raw financial and demographic data of borrowers into useful features and subsequently applying these features to classify applicants based on their likelihood of default. [4,5] Over the decades, different methods have been suggested, starting with more classic statistical models, to more modern machine learning and, more recently, deep learning and self-supervised learning techniques. All of these methods have their strengths and weaknesses based on what type of data is available, how complex credit behavior might be, and what financial institutions require when it comes to operation. Logistic regression and decision trees are traditional tools used in assessing credit risks and have remained the norm due to their simplicity, interpretability, and regulatory compliance. Such models work on handcrafted features designed by domain experts, which can usually be associated with variables such as income, debt-to-income ratio, length of credit history, and repayment behaviour. Although these models are practical in most scenarios, they have a high dependence on labelled data and feature engineering by experts, which may fail to capture complicated, non-linear trends in borrower behaviour.

Additionally, they are prone to poor performance in large and high-dimensional datasets, which are often present in modern financial systems. To overcome these weaknesses, the machine learning community has considered a vast number of representation learning approaches. These tend to automatically acquire useful feature representations in the data, minimizing or obviating the need to manually engineer features. For example, autoencoders

reduce the dimension of their input into a low-dimensional latent space that encodes the most relevant characteristics of the original features. Further improvements to the ability to learn complex relationships in tabular credit data include deep neural networks, such as fully connected feedforward networks and more recent architectures, including TabNet and TabTransformer. Nonetheless, training these methods typically requires large amounts of labelled information. This restriction is not always feasible due to privacy concerns or data sparsity (particularly in small banks or niche markets). Self-supervised learning (SSL) presents an important alternative because the labelled data is not necessary to teach the model rich and high-level representations. Rather, it is the pretext tasks in SSL frameworks that result in generating well-designed pseudo-labels. These pretext tasks include masked feature prediction, contrastive learning and autoencoding. Such work will stimulate the model to learn about the intermediate relationships and the inner structure of the data. For example, in masked prediction of features, the model learns to predict unobserved features based on observed features, thereby learning about feature correlations. In contrastive learning, it learns a discriminative feature space that separates similar and dissimilar samples. After pre-training is complete, the representations are retained and can be fine-tuned to perform specific downstream tasks, such as credit risk stratification, using limited labelled data. The effort to sort borrowers into specific risk categories, known as risk stratification, is facilitated by the existence of robust data representations. Well-formed embeddings enable downstream classifiers (e.g., logistic regression, random forests or neural networks) to better discriminate between high-risk and low-risk borrowers.

Furthermore, the ability to replace representations and decouple them with classification provides flexibility to SSL-based models. Once trained on a pretext task (data), the same network encoder can be adapted to other institutions, products, or geographies, leading to model transferability and scalability. Moreover, good credit representation will also help address issues of fairness and interpretability, which are of great importance in any financial modelling. More equitable risk assessment can be based on representations taking into account informative behavioral patterns, but suppressing noise and bias. Also, in embeddings-based models, one can afford interpretability using post-hoc methods (such as SHAP or LIME), especially with transparent downstream models.

2. Literature Survey

2.1. Traditional Credit Risk Models

The standard techniques of credit risk assessment have relied on statistics and machine learning models, including logistic regression, decision trees, and support vector machines (SVMs), in recent times. These models have been valued because they are easily interpreted and they have sound methodologies. Nevertheless, they typically require a lot of manual work in feature [6-9] engineering and an abundance of quality labelled data. Additionally, traditional methods may not readily identify intricate, non-linear associations present in financial information. These shortcomings have been addressed with the increased interest in deep learning methods, which have the capability of automatically learning hierarchically structured feature representations that can result in potentially more precise and robust credit scoring systems in recent years.

2.2. Self-Supervised Learning Paradigms

Self-supervised learning (SSL) has become a strong alternative in contexts where labelled data are sparse or costly to acquire. SSL methods utilise the inherent structures of the data to create pseudo-labels through pretext tasks, thereby performing representation learning without the need for manual data labelling. Central paradigms include contrastive learning, which trains models to discriminate between similar (positive) and different (negative) data pairs to understand discriminative features. Masked modelling, which is based on NLP progress, entails covering sections of the input and training the model to deduce the enclosed elements. Applications to tabular data have also occurred in other current models, such as TabNet and TabTransformer. Another common SSL algorithm used in the structured data setting involves reconstructing the initial data input via a latent code in the form of autoencoders.

2.3. SSL in Financial Domains

The use of SSL on financial data, especially time-series data and tabular data, is a recent topic of research. An independent contrastive learning model for fraud detection, presented by Liu et al., showed that self-supervised methods performed better than their supervised baselines when the data under the label is scarce. On the same note, Cheng et al. employed masked pre-training with a BERT-style model on financial transaction sequences, demonstrating the possibility of transfer learning for financial tasks. With these upbeat developments, the

application of SSL in credit scoring lacks sufficient research. There are considerable gaps in current research that correspond to the specificities of credit data, including imbalanced classes, time dependencies, and regulations, which present a significant opportunity for innovation.

2.4. Gaps Identified

Although SSL has been increasingly applied in a broader context of financial analytics, several research gaps remain regarding its use in credit scoring. To begin with, one can find obvious evidence of the absence of cross-paradigmatic comparisons between such SSL paradigm opposites as contrastive, masked modelling, and reconstruction-based techniques on structured credit data. In the absence of such comparisons, one cannot determine which techniques of SSL prove to be most effective in this sphere. Second, model interpretability remains one of the most important factors, as financial institutions must use transparent risk models that comply with regulations and ensure trust among stakeholders. Little research has been conducted to study the effect of SSL-derived features on model predictions in a meaningful manner. Lastly, there has been limited research on pretraining-fine-tuning in credit risk stratification. Knowledge of how pre-training a large, unlabeled set can be effectively transferred to targeted credit scoring tasks would provide useful insights into building large-scale, low-data risk models.

3. Methodology

3.1. Data Description

The present research paper is based on two well-established standard datasets used to assess credit risk prediction models, namely the German Credit Data and the Lending Club Loan Data. [10-13] These datasets are very diverse in size and complexity, and the set of features in each dataset is quite different, providing a broad foundation to evaluate the performance and generalizability of self-supervised learning (SSL) methods applied to credit scoring tasks. The UCI Machine Learning Repository originally provided the German Credit Data set, as it consists of 1,000 instances with 20 features, including both numerical and categorical variables. Each row represents a loan applicant, and the features included refer to the elements relevant to the loan applicant, including loan purpose, credit history, employment status, age, and housing status. The target variable is 0/1, indicating whether the applicant is a good or bad credit risk. Being relatively small, this dataset can be applied in many studies to

check the model's performance in situations where an insufficient amount of data is available, as well as to test methods that are not prone to overfitting. It can also be used as a baseline for those models that require little preprocessing and computational resources. In comparison, the Lending Club Loan Data is a vast, practical credit dataset derived from one of the largest peer-to-peer lending platforms in the United States. This dataset contains rich information on borrower characteristics, loan parameters and terms, and repayment performance or patterns, as well as credit experiences and habits, based on more than 100,000 loan records and featuring over 50 features. Numerical indicators include the amount and interest rate of the loan, whereas the categorical ones include the loan grade, employment title and use. The Lending Club dataset presents a more realistic and challenging setting for testing current machine learning and SSL models due to its size and scope. It allows investigations of scalability, robustness, and capacity and models to learn useful representations.

3.2. Preprocessing

Preprocessing of credit data is an important part of machine learning model preparation, particularly in the setting of self-supervised learning, where data quality plays a substantial role in the representations acquired. In the current study, the same preprocessing pipeline was used for the German Credit dataset and the Lending Club Loan dataset through the application of standardization to provide consistency and compatibility of the model in subsequent experiments. To begin with, the one-hot encoding technique was applied to the categorical variables, which is one of the most popular methods for converting categorical features into a binary matrix form. A new binary variable is created for each distinct category under a feature, and then models are induced to read categorical information without assuming any levels of categories. The dimensionality of the dataset is increased by the one-hot encoding technique, especially for high-cardinality features, yet it ensures that the categories of data contribute accordingly in models that accept numerical input. Secondly, similar to the categorical features, numerical features were converted through min-max normalization scale, in particular, within a prescribed range that is usually [0, 1]. This is a significant change to distance-based and gradient-based methods, because without it, features with wider numerical ranges might end up dominating the learning process. Normalization also helps in a rapid convergence of the training and greater numerical stability, which proves useful to the models of deep learning and SSL algorithms. Thirdly, K-Nearest Neighbor (KNN) imputation was applied to missing values,

which is a typical issue in real-world finance data. Here, missing values are estimated using the feature values of the most similar records in the dataset. KNN imputation is non-parametric, adaptive, and should be used with datasets in which the missingness is not random or bias may be introduced to the dataset by a simple imputation method (such as mean or median imputation). The local structure in the data is aided by KNN, which provides a better imputation, thereby preserving the underlying data distribution. This is essential for representation learning.

3.3. SSL Framework Overview

Their suggested self-supervised learning (SSL) model for credit risk modelling features a modular pipeline that trains raw credit data into useful embeddings, which are then utilised in the downstream classification. All the procedures in this pipeline help to learn strong and broad representations, particularly in circumstances where the scarcity of labelled data is at its peak.

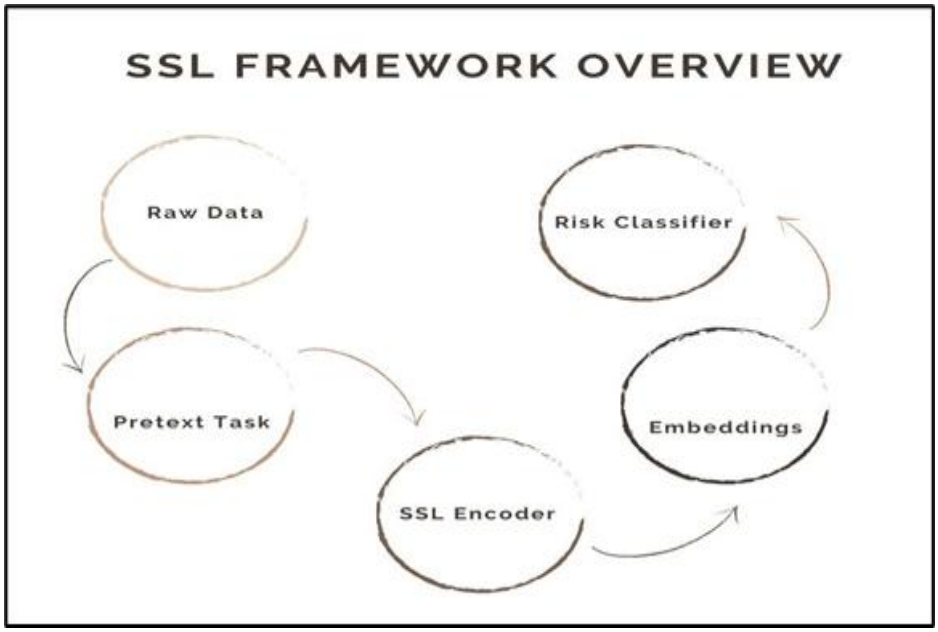

Figure 2: SSL Framework Overview

- **Raw Data:** The procedure begins with preprocessed raw table records, which include both numerical and categorical variables for credit samples, such as German Credit and Lending Club. This information has become the start point of the SSL framework. At this step, increasing data quality to such points is necessary, normalization, encoding, imputation, since the efficiency of SSL models is highly

predetermined by the character of the underlying data distribution utilized to provide powerful pseudo-labels to the training process.

- **Pretext Task:** This is then followed by a pretext task that helps create self-supervised signals using unlabelled data. Examples of common pretext tasks include contrastive learning (separating similar and dissimilar records), masked feature modelling (predicting hidden features), and reconstruction (e.g., as in autoencoders). The idea is to make the model usefully learn transferable, informative feature representations by addressing a corresponding auxiliary task in line with the organization of the data.
- **SSL Encoder:** The SSL encoder is a neural network (e.g. Transformer, MLP, or TabNet), trained on the solution of the pretext task. With this training, the encoder is trained to map a given input data into a concise feature space owing to its high dimensions. The encoder is designed to extract semantic relationships between records, which allows this encoder to generalize across distributions and tasks.
- **Embeddings:** The encoder produces embeddings, which are the compressed features learned for each credit record. These embeddings capture crucial consistency and form in the information, becoming the inputs to downstream tasks. Under a given learning strategy, they can be stored, reused, or polished up.
- **Risk Classifier:** The embeddings are, ultimately, passed into a risk classifier, which is typically a supervised network trained on labelled credit risk results. In this model, the representations are learned to correspond to discrete risk categories (e.g. good vs. bad credit). Owing to the SSL, the classifier does not need to be fed with as many labels, as the embeddings are learned through SSL, with the added benefit that the classifier can be very accurate in low-resource settings with relative ease.

3.4. Pretext Tasks Implemented

This paper applied three commonly used self-supervised learning (SSL) pretext tasks (masked feature prediction, contrastive learning and autoencoder-based pre-training) to test their usefulness in credit risk modelling. Every task trains the model to learn informative representations [14-17] from unlabeled credit data, utilising intrinsic data structures and patterns.

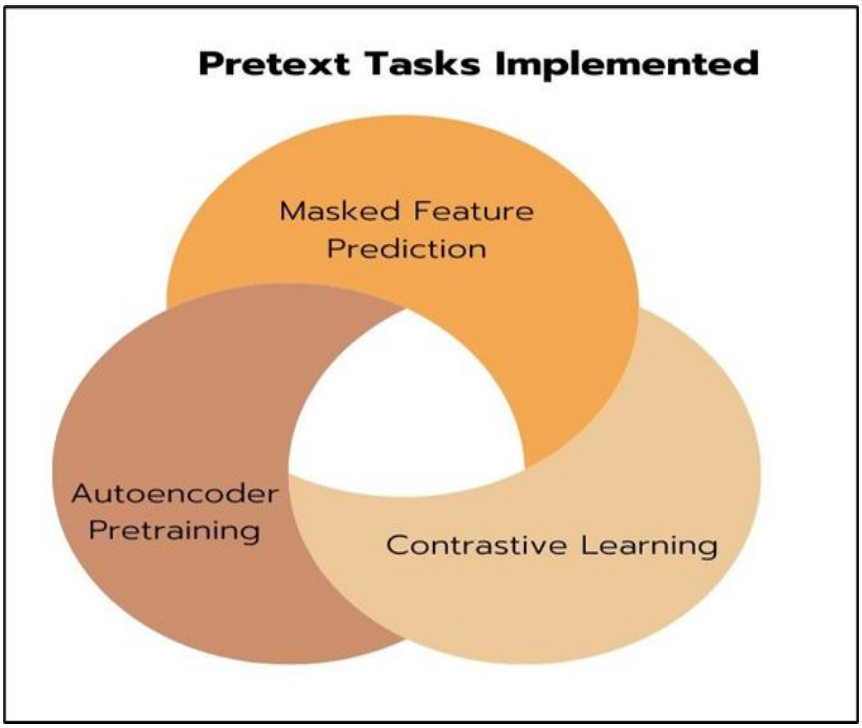

Figure 3: Pretext Tasks Implemented

- **Masked Feature Prediction:** The masked feature prediction was based on the concept of masked language modelling, a popular approach in models such as BERT. In this method, 15 per cent of the input features are chosen randomly and masked, and the model is trained to reconstruct the masked values using the information from the remaining unmasked inputs. The approach enables the model to identify cross-dependencies among features and correlations within the data. Depending on the type of data being optimized, the objective is usually represented by the mean squared error (MSE) and/or cross-entropy loss in the cases of numerical and categorical features, respectively. This assignment is particularly specific in the case of tabulated data, as the task of forecasting missing values is closely tied to the essence of structured characteristics.
- **Contrastive Learning:** In contrastive learning, one learns representations by trying to drive similar (positive) examples closer together and dissimilar (negative) examples far apart in the representation space. Positive pairs involve augmented versions of the same data instance (typically as a result of a process such as feature masking, noise injection or dropout). In contrast, other data instances are used as

negative pairs. This training of the model differentiates such relationships with the normalized Temperature-scaled Cross Entropy Loss (NT-Xent). This makes the model learn invariant and discriminative features that are generalized with or without labelled data.

- **Autoencoder Pre-training:** This approach to pre-training involves learning a compressed representation (encoding) of the input by training a neural network that encodes the input into a reduced representation, and then reconstructs the original input through decoding. The architecture typically consists of an encoder, which takes the input and maps it into a low-dimensional bottleneck vector, and a decoder, which attempts to reconstruct the input using the bottleneck vector. The training goal is to minimize the reconstruction loss, which is typically an MSE. It is a good task to initialize downstream classifiers with: the task enables the model to learn the most salient features in the data.

3.5. Downstream Risk Classification

After the self-supervised learning (SSL) models have created the embeddings of the original credit data, they become input to downstream credit risk classification jobs. The idea is to determine whether the learned embeddings reflect some meaningful patterns which can enhance the prediction of creditworthiness. To guarantee the overall estimation, the three most popular classifiers were utilized: Logistic Regression, Random Forest, and Multi-Layer Perceptron.

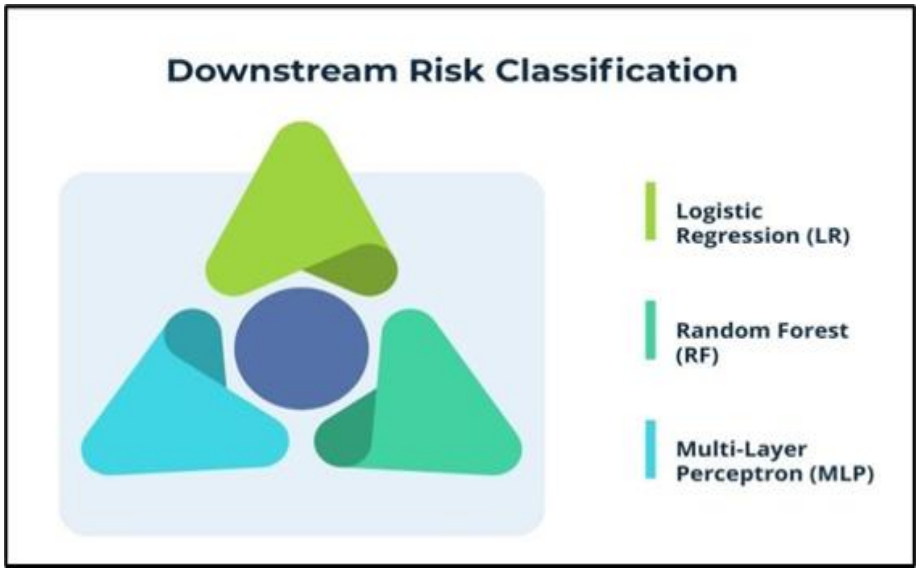

Figure 4: Downstream Risk Classification

- **Logistic Regression (LR):** Logistic Regression is a linear model that is kept simple (powerful) and is widely applied in credit scoring because it is easy to deploy and interpret. In the context of the SSL embeddings, LR can be considered a baseline classification model that can be used to evaluate the possibility of focusing on the linearly separable features. Its coefficients can also provide information on which aspects of the embedding space are considered most helpful in differentiating among high- and low-risk borrowers.
- **Random Forest (RF):** Random Forest is an ensemble learning technique which constructs many decision trees and compiles their estimates. It is also particularly suitable for dealing with complicated, non-linear relationships in the data. With RF above the SSL embeddings, the model allows for the exploitation of feature interactions that simpler models, such as LR, may not grasp. Furthermore, feature importance scores are available in RF, and they can be used to explain the significance of various embedding dimensions when predicting credit risk.
- **Multi-Layer Perceptron (MLP):** A Multi-Layer Perceptron is a feed-forward neural network that has the capability of modelling highly nonlinear functions. It is composed of one or more concealed levels, each with activation functions, which enable it to learn complex configurations in the SSL embeddings. MLPs perform especially well in cases where a complex relationship exists between latent features and risk labels. Incorporating MLPs enables the downstream classification task to have all the upstream representational power of the self-supervised encoders.

4. Results and Discussion

4.1. Evaluation Metrics

To evaluate the effectiveness of Self-Supervised Learning (SSL) in credit risk prediction, several standard performance metrics were employed: Accuracy, F1-Score, and Area Under the Receiver Operating Characteristic Curve (AUC). These metrics provide additional insights into modelling performance, particularly when using imbalanced or noisy financial data.

Table 1: Evaluation Metrics

Model	Accuracy	F1-Score	AUC
Supervised MLP	0.75	0.73	0.78
SSL + LR	0.77	0.75	0.81
SSL + MLP	0.80	0.78	0.84

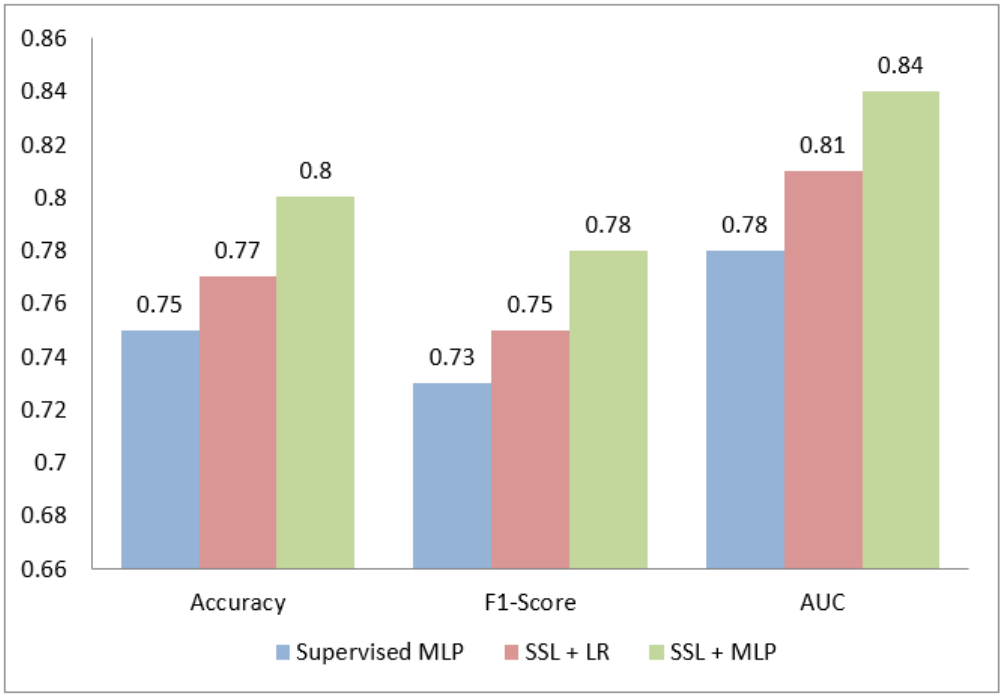

Figure 5: Graph representing Evaluation Metrics

- **Accuracy:** Accuracy is a percentage of the number of correctly classified examples divided by the total number of all samples. It is a simple metric for measuring the accuracy of models. The supervised baseline, using MLP, achieved an accuracy of 0.75, but the SSL-enhanced models were found to be better, with SSL + LR performing at 0.77 and SSL + MLP performing at 0.80. This demonstrates that pretext tasks, such as SSL embedding, can uncover helpful patterns that enhance classification accuracy, even when combined with simple linear models.
- **F1-Score:** The F1-Score is simply the harmonic mean of both precision and recall, and is particularly useful when facing class imbalance, as frequently found in credit risk data sets (e.g. having lots and lots of good loans and not so many bad ones). An

increase in the F1-score is an indication of the aptitude of a model to classify positive or negative classes accurately. The F1-score is 0.73 in the case of supervised MLP, increasing to 0.75 in the case of SSL + LR and further to 0.78 in the case of SSL + MLP. This development demonstrates the capability of SSL representations to enhance the sensitivity and accuracy of models, particularly in identifying cases that are at high risk.

- **AUC (Area Under the ROC Curve):** The AUC ranks the model's capability to distinguish between classes under various threshold conditions. The greater AUC indicates a superior overall ranking of positive versus negative classes, irrespective of the classification threshold. Supervised MLP yielded an AUC of 0.78, compared to 0.81 for SSL + LR and 0.84 for SSL + MLP. Such a boosted competency demonstrates that SSL-based embedding significantly enhances the discriminative capabilities of models and increases their secure usage in high-risk decision-making practices.

4.2. Ablation Studies

To gain more insights into how the individual self-supervised pretext task contributes to the overall model performance, ablation studies were undertaken to remove components one by one from the SSL framework and assess the results on the downstream credit risk classification. Such experiments are crucial in determining the most effective tasks for acquiring high-quality, transferable embeddings based on raw tabular data. The first important insight was that the removal of the contrastive learning task resulted in a significant decrease in performance level by almost 5 points in terms of all evaluation metrics, such as F1-Score and AUC. This suggests that contrastive learning is particularly useful in enabling the model to recognise subtle differences in borrower profiles. With the model learning a discriminatory representation by comparing positive pairings with negative ones, contrastive learning helps build a more organized and informative embedding map. Its lack will lead to the diminishment of the capability of the model to distinguish between risky and non-risky borrowers, which will cause a decrease in the classification accuracy and generalization. Conversely, the masked feature prediction task proved to be a highly effective element in adding robustness and stability. Omitting this task resulted in the model gaining elevated variance in training and diminished noise resistance of the input information. Masked prediction helps the model learn about inter-

feature relationships, which is crucial in cases like tabular financial data, where feature relationships may be subtle but extremely important (such as between income, loan amount, and credit history). It requires the model to fill in the gaps contextually, and therefore provides more comprehensive and transferable representations of features. Overall, the ablation study highlights the complementary effect of SSL pretext tasks, where contrastive learning enhances discriminative capacity and masked prediction yields improved contextual knowledge. The results thus serve as concrete recommendations for designing the SSL framework specifically for structured data in finance applications, where one seeks to optimise both performance and interpretability downstream in a credit scoring application.

4.3. Interpretability and Explainability

Interpretation is a fundamental attribute in the modelling of credit risk, and most importantly, it significantly impacts financial judgments, where compliance with regulatory structures and stakeholder trust are vital. Although self-supervised learning (SSL) models are often associated with intricate structures and abstract representations, the present research highlights the idea that they can be equally explainable if analysed correctly. SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-Agnostic Explanations) were used to evaluate the interpretability of SSL-based classifiers by placing them on downstream models that utilised SSL-produced embeddings. The use of SHAP and LIME has proven to have transparent and regular feature importance attribution, which identifies the mode of input variables that have the most effect on model predictions. In contrast to conventional end-to-end deep learning models, where the feature attributions may be confounded by degrees of abstraction, recent studies suggest that SSL-based models distinguish between representation learning and classification. Such a split allows for the interpretation of the classifier independently of the encoder, a feature that can be leveraged to derive meaningful results using post-hoc explainability techniques. For example, SHAP values revealed that loan purpose, credit history, and processed debt-to-income ratio were among the features that had a significant impact on decisions regarding credit risk when calculated based on SSL embeddings. Similarly, LIME provided local explanations that conformed to domain knowledge, making it easier to countercheck the predicted results on a per-case basis. In addition, embeddings produced by tasks such as masked prediction and contrastive learning semantically transferred the features, allowing downstream classifiers to retain certain interpretability despite the dense

vectors. This is in contrast to deep neural networks trained end-to-end, where the transformations of features are typically obscure to the point of being uninterpretable. Based on this, the models that were based on SSL not only had better predictive performance but were also more transparent, which is why they are better suited for use in a real-life financial setting. This interpretability strength also favours the wider adoption of SSL within credit scoring, where there is a need to balance performance commitments with the desire to explain them.

4.4. Scalability and Generalization

Among the main strengths of self-supervised learning (SSL) is that it can scale well with large, unlabeled data and still have good generalization to other domains. Saliability and generalization were tested in this work on the dataset of the lending club (over 100,000 examples and several features greater than 50), and it is much more extensive (and diverse) than the German Credit dataset. Its SSL pre-training framework performed very well in terms of scalability, learning good representations on this large, high-dimensional, and noiseless dataset without requiring any type of label supervision during pre-training. The large size of data to be processed and learned is representative of the computational effectiveness and relevance of SSL in real-world financial settings that involve big data. Also, the pretrained SSL encoders achieved outstanding cross-dataset generalization. Tuning in this case refers to when an encoder trained on the Lending Club data was fine-tuned on the smaller German Credit dataset, achieving very good performance with only minimal fine-tuning. This transferability demonstrates that the SSL model can learn universal credit-related representations that are not overfit to the distribution of a particular dataset. It is also highly effective in alleviating the need to feed huge masses of labelled information in target areas, which can come in handy in organizations that have limited access to annotated financial history. That a single encoder can generalize across datasets of varying feature distributions and sample sizes implies that SSL is scalable enough to become a basis of large-scale credit scoring systems. Large-scale public and privately held datasets can be pre-trained in these systems, and the systems can be fine-tuned more quickly for new environments or smaller institutions. This direction in itself makes SSL an exciting prospect for international applications of machine learning in credit risk prediction, not only to increase accuracy and robustness but also to enhance the general efficiency and reusability of machine learning pipelines in the financial industry.

5. Conclusion

This paper presents a comprehensive analysis of self-supervised learning (SSL) methods for credit risk assessment, specifically those applicable to tabular financial data. The fundamental principle of SSL is the utilization of large quantities of unlabeled data using the pretext job that can empower the design to realize the meaningful mindfulness that can subsequently be embraced in the targeted supervised jobs like credit risk segregation. The results indicate that SSL-based models surpass traditional credit scoring methods, such as logistic regression, decision trees, and even supervised deep neural networks, especially when scarce labelled data is available. SSL is more flexible and data-efficient because it decouples feature representation from the classification step. Among the performed pretext tasks, the most useful ones included contrastive learning and masked feature prediction. The presence of contrastive learning enhanced the discriminative abilities of the embeddings, whereas the masked prediction made them more robust and provided better contextual comprehension of the feature space. Together, all these tasks formed a synergy, making high-quality representations and, consequently, improving risk stratification. Moreover, the learned embeddings were competitive in terms of accuracy and transferred reasonably to other datasets. In combination with interpretable classifiers like logistic regression or decision trees, they provided insights into the most important risk-causing factors, which makes the method both accurate and interpretable.

This work has been instrumental in the application of SSL to credit risk modelling, even though there are still other avenues in this effort that can be explored. An important aspect is the multimodal extension of financial information, where tabular data with defined data types is supplemented with unstructured data, such as textual descriptions of transactions, customer reviews, or documents containing financial information. By encoding text as well as tabular data in the same scheme as SSL, features can be richer, and predictive performance can be enhanced. Transformer and attention mechanisms are also another possible direction that has turned out to be very successful in both NLP and vision. The potential limitation of the study is that applying transformer-based architectures, such as TabTransformer or Financial-BERT, to SSL tasks would allow further increasing the extent to which the model learns complex relationships existing in high-dimensional financial data. Finally, it is essential to mention the analysis of fairness and mitigation of bias in SSL tools. Financial decision-making should be fair, and as much as SSL can decrease reliance on prejudiced labels, it is essential to ensure that

representations learned do not reproduce or exert biases in the information. Further steps should be taken to study fairness-aware SSL frameworks and establish their performance in terms of the quality of the fair, unbiased credit risk biases used.

References

[1] Hand, D. J., & Henley, W. E. (1997). Statistical classification methods in consumer credit scoring: a review. Journal of the royal statistical society: series a (statistics in society), 160(3), 523-541.

[2] Baesens, B., Setiono, R., Mues, C., & Vanthienen, J. (2003). Using neural network rule extraction and decision tables for credit-risk evaluation. Management science, 49(3), 312-329.

[3] Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020, November). A simple framework for contrastive learning of visual representations. In International Conference on machine learning (pp. 1597-1607). PmLR.

[4] Huang, X., Khetan, A., Cvitkovic, M., & Karnin, Z. (2020). Tabtransformer: Tabular data modelling using contextual embeddings. arXiv preprint arXiv:2012.06678.

[5] Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing data dimensionality with neural networks. science, 313(5786), 504-507.

[6] Liu, X., Zhang, F., Hou, Z., Mian, L., Wang, Z., Zhang, J., & Tang, J. (2021). Self-supervised learning: Generative or contrastive. IEEE Transactions on Knowledge and Data Engineering, 35(1), 857-876.

[7] Lessmann, S., Baesens, B., Seow, H. V., & Thomas, L. C. (2015). Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. European Journal of Operational Research, 247(1), 124-136.

[8] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019, June). Bert: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers) (pp. 4171-4186).

[9] Yoon, J., Jarrett, D., & Van der Schaar, M. (2019). Time-series generative adversarial networks. Advances in Neural Information Processing Systems, 32.

[10] Geiping, J., Bauermeister, H., Dröge, H., & Moeller, M. (2020). Inverting Gradients: How Easy Is It to Break Privacy in Federated Learning? Advances in Neural Information Processing Systems, 33, 16937-16947.

[11] Yao, Y. (2024). Self-Supervised Credit Scoring with Masked Autoencoders: Addressing Data Gaps and Noise Robustly. Journal of Computer Technology and Software, 3(8).

[12] Yang, S., Chen, H., Huang, J., Yan, Y., Chen, J., & Xiong, A. (2022, October). Split learning based on self-supervised learning. In International Conference on Computer Engineering and Networks (pp. 95-104). Singapore: Springer Nature Singapore.

[13] Rani, V., Nabi, S. T., Kumar, M., Mittal, A., & Kumar, K. (2023). Self-supervised learning: A succinct review. Archives of Computational Methods in Engineering, 30(4), 2761-2775.

[14] Tian, Y., Yu, L., Chen, X., & Ganguli, S. (2020). Understanding self-supervised learning with dual deep networks. arXiv preprint arXiv:2010.00578.

[15] Li, T., Kou, G., & Peng, Y. (2023). A new representation learning approach for credit data analysis. Information Sciences, 627, 115-131.

[16] Bhatore, S., Mohan, L., & Reddy, Y. R. (2020). Machine learning techniques for credit risk evaluation: a systematic literature review. Journal of Banking and Financial Technology, 4(1), 111-138.

[17] Sui, Y., Wu, T., Cresswell, J. C., Wu, G., Stein, G., Huang, X. S., ... & Volkovs, M. (2023). Self-supervised representation learning from random data projectors. arXiv preprint arXiv:2310.07756.

[18] Self-Supervised Learning Explained, Encord, Online. https://encord.com/blog/self-supervised-learning/

[19] Taherdoost, H. (2024). Beyond supervised: the rise of self-supervised learning in autonomous systems. Information, 15(8), 491.

[20] Yao, T., Yi, X., Cheng, D. Z., Yu, F., Chen, T., Menon, A., ... & Ettinger, E. (2021, October). Self-supervised learning for large-scale item recommendations. In Proceedings of the 30th ACM International Conference on Information & Knowledge Management (pp. 4321-4330).