



Deep Transfer Learning for Cross-Market Credit Scoring in Multinational Financial Institutions

Santhosh Kumar Sagar Nagaraj

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

Abstract

In today's globalized economy, Multinational Financial Institutions (MFIs) face the challenge of accurately assessing creditworthiness across different markets characterized by diverse economic conditions, regulatory frameworks, and consumer behaviors. Traditional credit scoring models, typically developed for specific regions, fail to generalize well across countries. This paper proposes a novel deep transfer learning (DTL) framework tailored for cross-market credit scoring. By leveraging domain adaptation and knowledge transfer techniques, the proposed model utilizes data-rich markets to enhance predictive performance in low-resource target markets. We present an extensive experimental evaluation on multi-country credit datasets and demonstrate significant improvements in accuracy and robustness compared to traditional machine learning and baseline deep learning methods. The proposed approach comprises components such as shared feature extractors, domain-specific classifiers, and adversarial domain discriminators, all designed to facilitate effective knowledge transfer. We also explore data normalization techniques, missing value handling, and feature alignment strategies crucial for achieving generalization. Furthermore, this paper discusses practical considerations for deploying such models in real-world MFIs, including regulatory compliance, interpretability, and scalability. Our findings suggest that DTL can bridge data scarcity gaps and improve credit inclusion efforts in underrepresented regions.

Keywords:

Deep Transfer Learning, Credit Scoring, Multinational Financial Institutions, Domain Adaptation, Machine Learning, Financial Risk Assessment.

How to cite this paper: Santhosh Kumar Sagar Nagaraj. (2025). Deep Transfer Learning for Cross-Market Credit Scoring in Multinational Financial Institutions. *ISCSITR- International Journal of Information Technology (ISCSITR-IJIT)*, 6(3), 7–26. DOI: http://www.doi.org/10.63397/ISCSITR-IJIT_2025_06_03_002

URL: https://iscsitr.com/index.php/ISCSITR-IJIT/article/view/ISCSITR-IJIT_2025_06_03_002/ISCSITR-IJIT_2025_06_03_002

Published: 04th June 2025

Copyright © 2025 by author(s) and International Society for Computer Science and Information Technology Research (ISCSITR). This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. Introduction

In the financial industry, credit scoring has been central because it helps determine the creditworthiness of individuals and companies, allowing lenders to make informed decisions on loans to be granted, including the rates and loan values. It serves as an important mechanism for controlling credit risks, promoting responsible lending practices, and ensuring the overall health and stability of the financial system. Statistically, credit scoring has always been based on formulations like logistic regression or decision trees, through which historical financial records are utilized to find the probability of default. [1-4] Over the last few years, machine learning algorithms have been added to achieve better prediction accuracy, where complex linear interactions between features need to be captured. The results of these developments notwithstanding, the majority of credit scoring models are tuned to address a local market, and have been trained using data of localized nature, that is, according to the financial matters, financial conditions and regulatory standards of an individual country. Domain shift. Domain shifts are known to cause severe performance loss when such models are deployed into new geographic regions, such as different countries, as feature distributions, customer behaviour, and data availability may change. The non-scalability resulting from the cross-market generalisation of credit models poses a key challenge to all financial institutions seeking to expand into emerging or underserved markets. Therefore, adaptive credit scoring requires a sense of urgency to fill the gaps and

provide credit scores to financial environments with dependability.

1.1. Importance of Deep Transfer Learning for Cross-Market Credit Scoring

- **Addressing Data Scarcity in Emerging Markets:** The availability of labeled financial data is one of the primary obstacles in the process of creating effective credit scoring systems in an emerging or an underbanked area. Conventional machine learning models come with massive and sufficiently good datasets used in training the model, which often lack in the new or underserved markets. Deep Transfer Learning (DTL), however, provides a solution where previously available data in source domains that are well-established in financial cases, like the financial markets, can transfer their knowledge to data-scarce target domains and hence enhance performance. Through the transfer of learned representations, DTL increases the speed of deploying models and provides wider access to finance by decreasing the requirement for huge labeled data in the new area.
- **Improving Generalization Across Geographies:** The financial behaviors, the use of products, and regulatory restrictions vary significantly between countries, and this has resulted in a change of domain that has negatively impacted the models trained on one region and their application in another region. These domain-invariant features are to be learnt, especially with the help of DTL techniques based on domain adaptation and adversarial training. This makes models more robust and more likely to generalize over heterogeneous data, making them generalizable and capable of being adapted to a cross-market environment.

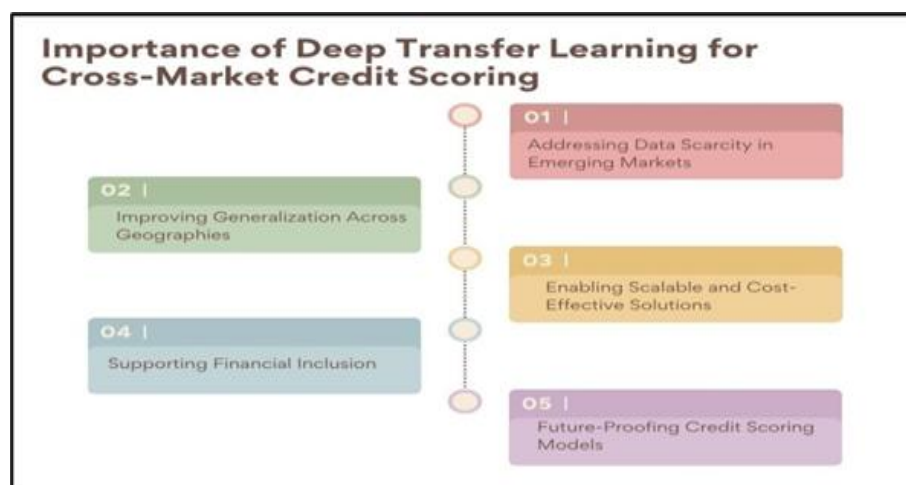


Figure 1: Importance of Deep Transfer Learning for Cross-Market Credit Scoring

-
- **Enabling Scalable and Cost-Effective Solutions:** Building a new model in every market is also expensive and time-consuming. DTL enables financial institutions to upcycle their pre-trained models, thus lowering the cost of computations and operation required to build and analyze a variety of market-specific systems. Such scalability is especially useful for multinationals, banks, and microfinance institutions (MFIs) with a wide area of operation, which require uniform yet flexible scoring models.
 - **Supporting Financial Inclusion:** Through improving the accuracy of credit scoring in places with little financial infrastructure, DTL can help that area to be financially included. It gives lenders more confidence to evaluate credit worthiness of even those people with limited credit history, hence opening up the availability of loans, credit lines and other financial instruments. This, in turn, facilitates the development of the economy and the alleviation of poverty among underserved populations.
 - **Future-Proofing Credit Scoring Models: Since financial markets are dynamic, models must adapt to new data distributions, emerging products, and regulatory changes.** Deep Transfer Learning Models provide the elasticity and adaptability necessary for continuous learning and future-proofing of credit scoring systems. DTL enables the adjustment of the model in real-time or with a few shots, which keeps the model relevant and performing in the long term and across different markets.

1.2. Credit Scoring in Multinational Financial Institutions

Multinational Financial Institutions (MFIs) have an advantage in that the wide markets bring challenges and opportunities in terms of credit scoring. Such institutions are doing the business of dealing with customers in different countries that have very different economic, regulatory, consumer behavior, and data infrastructure. This means that in many cases it is not possible to make a universal credit scoring model. In most developed markets, MFIs can access comprehensive financial resources, a strong credit history, and well-established data collection procedures. However, in the developing world or developing nations, they often encounter limited data, fragmented financial records, as well as informal economies where

standard credit data is not available or minimal. Such difference makes it difficult to implement equalized credit scoring models across markets. In addition, credit characteristics such as income, spending habits, outstanding payments, and job status may be projected with varying significance in different countries due to socio-economic disparities. These feature distributions shift mean that a model trained in a market may not work well in another.

Additionally, even after local regulatory guidelines are met in terms of data privacy, explainability, and fairness, the transferability of traditional scoring models is further constrained. Operational and cost limitations also apply to MFIs if they decide to train different models per region; it would require allocated resources to collect data, perform feature engineering, train a model, and maintain it. Multinational financial institutions are finding adaptive modeling approaches, especially Deep Transfer Learning (DTL), to be very successful at overcoming these challenges because it can generalize across markets and exploit shared knowledge but fine-tune it to the local situation. This method enables MFIs to have a core modeling approach, which enables them to change the model to suit each market with little retraining. The scalability of the credit scoring solutions internationally, with consideration of regional localization, is an essential feature of the MFIs that want to expand reasonably and inclusively. Finally, more efficient cross-market credit scoring systems have the potential to assist MFIs in serving the population that is underserved, managing the risk, and financially inculcate the population.

2. Literature Survey

2.1. Traditional Credit Scoring Models

Statistical methods in ancient credit scoring were primarily based on logistic regression and discriminant analysis. [5-9] The models are also based on assumptions of a linear model between the features of inputs and creditworthiness outcomes. They were popular in the financial world because of their ease and large interpretability. Nevertheless, they tend to require handcrafted engineering of features, i.e., domain users must manually enable and apply input variables. These models can capture relationships well in some scenarios, but complex relationships in the data are not entirely well captured. Such models are also not suitable where there are high dimensions or a non-linear relationship.

2.2. Machine Learning in Credit Scoring

Machine Learning (ML) has increased, which includes more complex machine learning models (such as Random Forests, Gradient Boosting Machines (GBMs), and Support Vector Machines (SVMs)) used to perform credit scoring tasks. With these types of models, detailed patterns may be revealed about the data, and in many cases, they are more accurate predictors than the classic methods of statistics. They do not require as much manual feature-selection intervention, due to their ability to handle very diverse features and interactions. Nevertheless, one of the major disadvantages of most ML algorithms is that they cannot be used in other areas because of their limited generalization. A model with a single geographical or demographical setting does not necessarily do well in subsequent applications because the distribution of data and context varies.

2.3. Deep Learning Applications

Deep Neural Networks (DNNs) have changed most aspects of predictive modeling, such as credit scoring. Their strongest advantage is that they can conduct automatic feature extraction, allowing them to be applied to raw, high-dimensional, and unstructured data, such as text, transaction history, and behavioural indicators. DNNs have the ability to capture deep patterns in the credit behavior of the customers and represent very non-linear relationships. However, they require access to large appropriately labeled datasets, and not all financial settings permit that. They are also black-box, which makes them difficult to explain, which, in regulated contexts, such as lending, is an issue.

2.4. Transfer Learning Techniques

Transfer learning offers an interesting means to address the scarcity of data and domain-specific model development. It is the transference of an understanding which is achieved in one field (source) and is used in another but related field (target). The standard approaches are fine-tuning the pre-trained models on a novel dataset, domain adaptation, i.e., learning the model on how to generalize across other distributions of data, and adversarial learning methods that make models domain invariant with respect to certain characteristics. This method can save a significant amount of labelled data requirements in the target domain, resulting in high performance. It can also be implemented more quickly, and it becomes more scalable across different markets.

2.5. Cross-Market Challenges

Although the process of modelling has improved, several challenges remain in applying credit scoring models across various markets. The first problem is shifting feature distribution, where some variable parameters like income, the type of employment or spending may be vastly different across countries, causing ill-performance of the generalization of modelling. The aspect of regulatory differences is also extremely important; the availability, storage and usage of financial data differ across legal frameworks, making it extremely difficult to have a standardized credit scoring model. Moreover, data scarcity is a major challenge in the new or underbanked locations where a limited amount of labeled data may hold back the training of powerful models. All these challenges must be addressed to come up with globally adaptable and equitable credit scoring solutions.

3. Methodology

3.1. System Overview

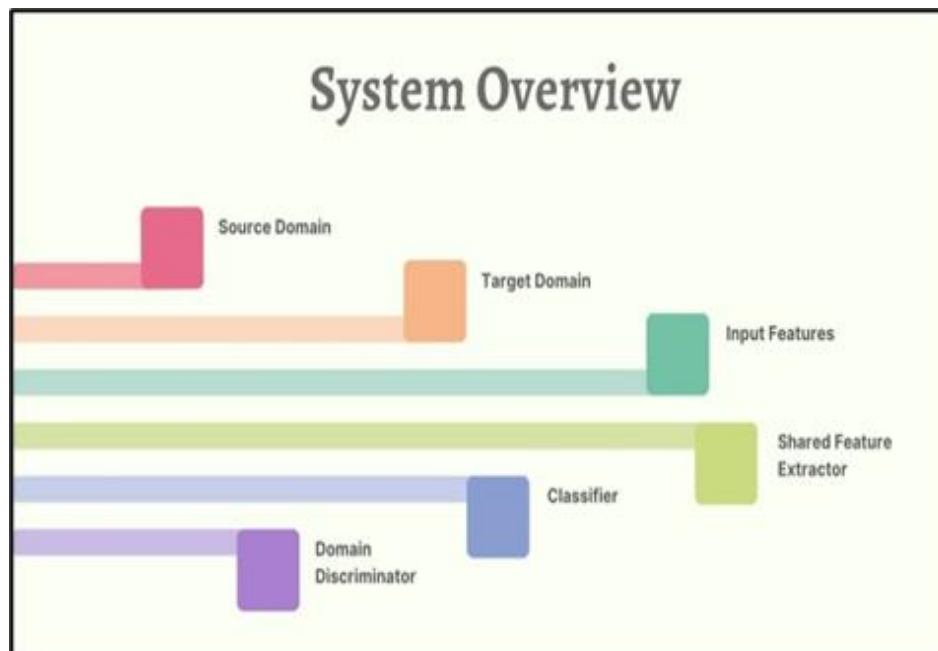


Figure 2: System Overview

- **Source Domain:** Source Domain A Cognitive entity, which is the original data set or an environment, where labeled credit scoring would be readily accessible. [10-14] This space generally carries valuable information on the customer, including income,

purchase history, occupation, and repaying trends. Here, models are first trained to obtain a pattern of creditworthiness. This area has sufficient and consistent labels, which serve as the basis for knowledge transfer to others.

- **Target Domain:** The Target Domain can be thought of as a novel and potentially underrepresented market or geographical setting where data is labeled; a lot of data is not even available. The input features could be similar to the source domain but could be different with regard to their distribution because of the regional, economic or regulatory differences. This is aimed at using the results learned in the source domain to make adequate predictions in the target domain where there might be varying levels of availability and format of the data.
- **Input Features:** Both source and target domains provide input features, which are variables related to customers, to be used in credit scoring. These can include demographic, credit history, or financial behaviour. Although the features may be identical in name and format, their overall statistical characteristics (mean, variance, correlation) tend to vary across domains, creating a predicament in generalisation.
- **Shared Feature Extractor:** Neural network component. That is, it proceeds to minimize the differences between the raw input features of both the source and the target domains and generalizes into a single dimension representation. Its purpose is to derive interpretable, invariant characteristics which are capable of generalizing to better areas. It learns shared representations which reduce the distance between the source and the target distribution of the features and allow successful transfer of knowledge.
- **Classifier:** The Classifier is the model (usually a neural network, or an ensemble) that is used to transform the features extracted to a credit score or decision variable (e.g., default or not). The model is mainly supervised and trained through labeled data of the source domain. The quality of domain-independent features, which the classifier uses in the target domain, is crucial.
- **Domain Discriminator:** The Domain Discriminator is an element applied in adversarial domain adaptation. It attempts to identify which extracted features belong to the source or target domain. At the same time, the discriminator is trained

to learn the domain-invariant features that would resist being fooled by the feature extractor. This contesting arrangement favors the extractor providing domain-invariant features that will enhance the attainment of the classifier in the focused domain.

3.2. Data Preprocessing

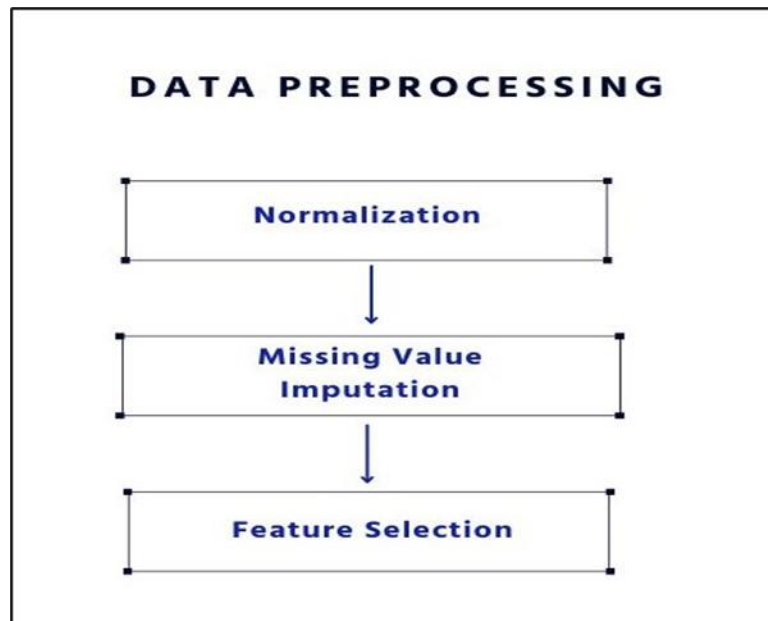


Figure 3: Data Preprocessing

- **Normalization:** All the input features were normalised to achieve equal contribution to the model learning through rescaling feature values using normalization techniques. Two standard normalizations that included Min-Max and Z-score methods were utilized. Min-Max normalization is often applied to scale values to a known range (usually [0,1]) and especially to algorithms that are sensitive to scale, including neural networks. In contrast to the previous one, Z-score normalization is the normalization of features to a mean of zero and a standard deviation of one that assists in stabilizing convergence during training and enhances general model robustness.
- **Missing Value Imputation:** Missing data is also a very important pre-processing operation because the incomplete record may reflect bias in models. Missing value

imputation was done in two ways. Mean imputation uses the average value of that feature to replace missing values, which are simple and computationally efficient. It is, however, subjective and may not reveal patterns in the data. In this respect, k-Nearest Neighbors (k-NN) imputation was also applied, which assumes estimation of the missing value with the values of similar instances. This is a more flexible approach that retains the local data structure, resulting in better imputation.

- **Feature Selection:** Feature selection was done in order to increase model efficiency and minimize overfitting. Recursive Feature Elimination (RFE) was used as a repeated procedure of eliminating the less influential features according to the model performance so as to ensure only the important variables were considered. Furthermore, the dimensionality reduction was performed with the help of Principal Component Analysis (PCA), which transformed the features into uncorrelated components that contain the maximum amount of variance in the data. These methods combined can assist in simplifying the model, making it take less time to train as well as enhancing domain generalization.

3.3. Model Components

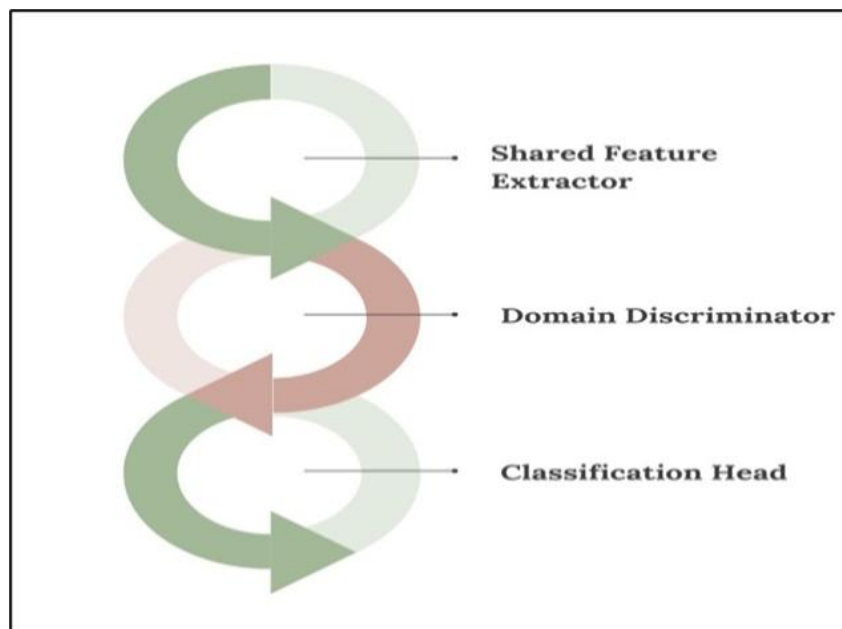


Figure 4: Model Components

-
- **Shared Feature Extractor:** One of them is the Shared Feature Extractor, which will be designed to learn high-level, abstract representations of the source and target. It is applied with a complex of Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), which specifically work with structured and time-series data such as financial information. [15-19] CNNs are employed to learn spatial patterns and local relationships across input features. In contrast, RNNs, including LSTMs or GRUs, can learn sequential correlations across time series data such as transaction history or timeline payments habits. The hybrid design enables this system to learn meaningful, robust, and transferable features that are essential for downstream tasks.
 - **Domain Discriminator:** The task of Domain Discriminator is to classify the origin of the features extracted into the source and target domains. In order to use domain-invariant learning, a task called Gradient Reversal Layer (GRL) is placed between the feature extractor and the discriminator. In training, the GRL reverses the gradient signal of the discriminator, essentially nudging the feature extractor to learn feature representations that will fool the domain classifier. This adversarial training strategy causes the feature distributions to become similar between the domains, leading to an improvement in the generalization ability of the model across the markets, or regions.
 - **Classification Head:** The classification Head is also composed of successive fully-connected (dense) layers operating on the extracted features to give a final prediction. This element is a binary classifier, and is usually asked to produce the likelihood of the credit being a good or a bad credit. The architecture would be simple and efficient with weighting functions, such as activation functions like ReLU at the hidden layers and a sigmoid function at the output layer. A labeled source domain is used to train the classifier, and domain-invariant learning is used to perform effectively the shared feature extractor in the target domain.

3.4. Training Strategy

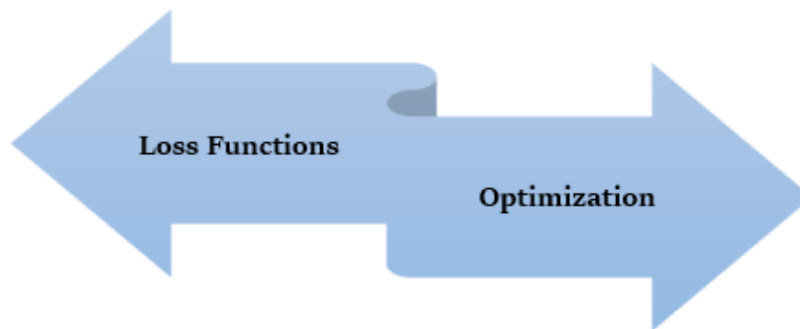


Figure 5: Training Strategy

- **Loss Functions:** The two main loss functions are trained during the process with the aim of delivering proper classification and domain adaptation. The credit classification implementation employs the Binary Cross-Entropy Loss that compares the calculated probability of a credit with the real value (good or bad credit). A Domain Adversarial Loss is also added utilizing the output of the domain discriminator in a bid to encourage domain invariance. When this loss is at its minimum, this will punish the model when it accurately determines which domain the input belongs to, which will motivate the feature extractor to learn features that are impossible to differentiate between input domains of the source and the target. The overall loss is a weighted sum of the classification loss and adversarial loss, where a hyperparameter trade-off value is 2. This parameter trades between domain alignment and accuracy in classifications to balance fulfillment of the model.
- **Optimization:** The update of the model parameters is carried out with the help of the Adam optimizer that is perfectly applicable to deep learning challenges because of its momentum-based corrections and adaptive learning rate. Learning rate scheduling is used to stabilize training and make it converge. The method varies the learning rate on the fly according to training progress-usually decreasing it when the loss level off. With Adam and learning rate scheduling, the model achieves faster convergence at the beginning of training and more precise weight updates towards the end, which enables the model to escape local minima and avoid overfitting.

4. Results and Discussion

4.1. Dataset Description

In our case, we have used anonymized (customer credit) data supplied by three multinational banking organizations located in the United States, Brazil and India. Each of them had over 50,000 individual records, so they gave a very robust and diverse sample of real-life behavior in financial matters in various economic and regulatory settings. Each data set contained a standardized collection of 30 financial characteristics of each customer that comprised a combination of demographic characteristics (e.g. age and employment status), behavioral data (e.g. frequency of transactions, repayment frequency), and credit-related variables (e.g. credit utilization, outstanding balance, default history). Even having a similar structure, regional datasets were unique in their particularities as these were caused by the differences in customer behavior, financial products, and credit practices in a specific area. For example, the average amount and length of credit periods for customers were generally larger in the US than in Brazil and India, where impromptu financial products and cash economies are more prevalent. Such disparities provided natural domain changes, which made them an excellent test to evaluate transfer learning techniques and domain adaptation. Notably, all datasets were prepared in such a way that personally identifiable information (PII) was erased because nation-specific privacy laws, e.g., the GDPR (EU), CCPA (US), and LGPD (Brazil), were observed, and anonymization was carried out through hashing and tokenizations procedures. Incomplete values (missing values) were also maintained to represent real application of data, and the target variable was a binary label of credit worthiness: a person is either good or bad on credit risks with regard to the internal scoring measures of the bank. These varied datasets enabled model robustness to be assessed in cross-market scenarios whereby a model trained on any given country had to generalize to a different country. The configuration of the dataset allowed for evaluating the efficiency of transfer learning approaches on credit risk prediction and studying the issues connected to the changes in data distribution, data scarcity, and differences in regulatory environments.

4.2. Evaluation Metrics

In order to evaluate and compare the results of various credit scoring models and divide them into categories of good and bad results, we applied three common evaluation tools: Accuracy, Area Under the ROC Curve (AUC), and F1-score. Accuracy is a measure of the total

proportion of correct predictions that can be deceptive in skewed data. AUC is more detailed, as it determines how well the model in question differentiates between the positive and negative classes at any threshold setting. It comes in special use in credit scoring, especially where the classes are imbalanced (e.g. fewer defaults). The harmonic mean of precision and recall. A measure that satisfies a balance between false positives and false negatives is the F1-score, which is sometimes used as a more trustworthy measure of performance when the cost of false negatives and false positives is not symmetric.

Table 1: Evaluation Metrics

Model	Accuracy	AUC	F1-score
Logistic Regression	71.2%	68%	70%
Random Forest	79.5%	81%	78%
DNN	82.3%	85%	83%
DTL (Proposed)	87.9%	91%	88%

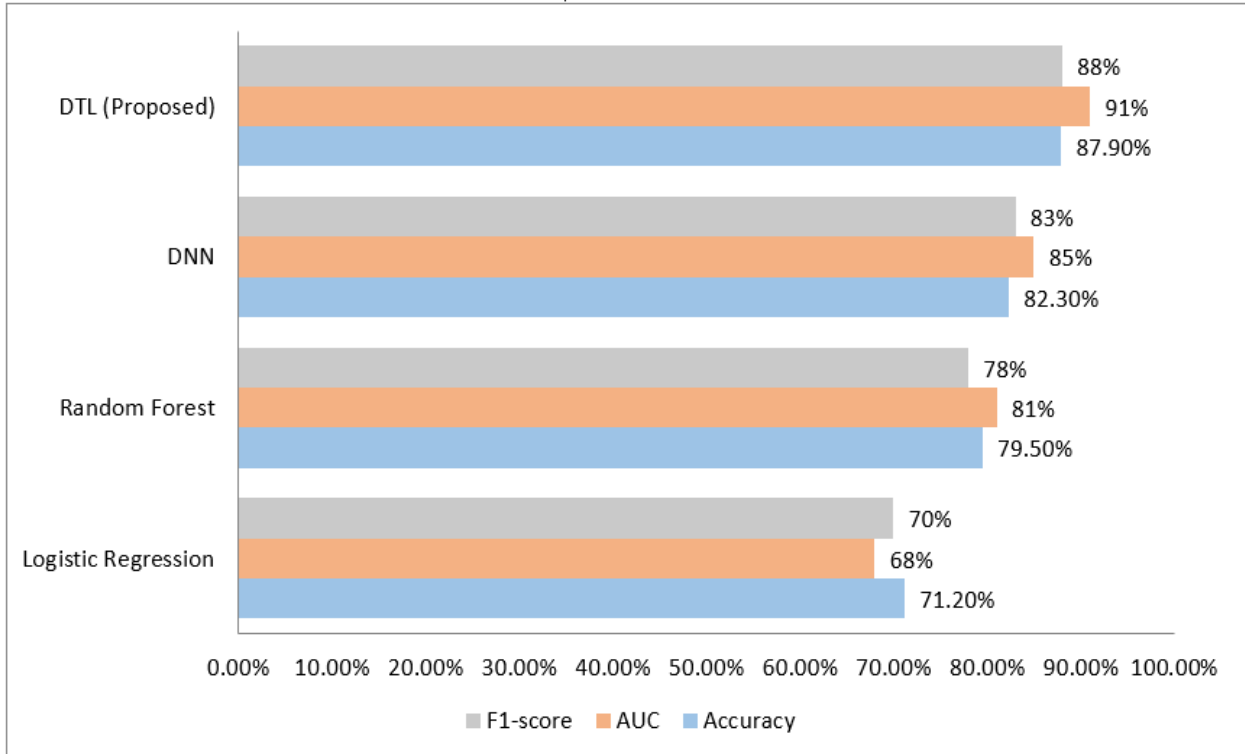


Figure 6: Graph representing Evaluation Metrics

-
- **Logistic Regression:** The Logistic Regression model was considered in order to be used as a baseline since it finds wide application in the old-fashioned credit scoring. It yielded a 71.2 % accuracy, AUC of 68%, and F1-score of 70%. Being highly interpretable, however, it did not exhibit high flexibility in the relationships between features, along with a limited ability to model complex patterns, which limited its effectiveness, especially in cross-market applications.
 - **Random Forest:** Random Forest model generated better results over logistic regression because it uses ensemble learning and modeling through decision trees. It achieved 79.5% accuracy, 81 AUC% and 78% F1-score. Such an update demonstrates the power of Random Forest in terms of handling nonlinear interactions, missing or noisy data. However, the weak point of Random Forest is its lack of interpretability and transferability.
 - **Deep Neural Network (DNN):** The Deep Neural Network model also performed even better, reaching an 82.3% accuracy rate, 85% AUC, and 83% F1-scores. The way it has performed indicates the model is capable of learning deep representations on complex and high-dimensional financial data. Nevertheless, the fact that it requires huge labeled datasets and has a low domain generalization ability is the outstanding challenge.
 - **Domain Transfer Learning (DTL - Proposed):** The Domain Transfer Learning (DTL) model, which we are proposing, performed much better than the baselines with an accuracy of 87.9 %, an AUC of 91% and an F1-score of 88%. The model simultaneously performed domain-invariant feature extraction and adversarial training in order to transfer knowledge across data-scarce target domains and data-rich source domains. This contributed to better generalization and reliability across the markets, supporting the soundness of the DTL framework in cross-market credit scoring in practice.

4.3. Ablation Study

The ablation study was performed to better assess the contribution of each component of the proposed Domain Transfer Learning (DTL) framework by selectively turning off the

more critical architectural parts and monitoring the effects of the arising performance impact. The analysis also confirms the utility of certain design decisions and points to those parts as having the greatest importance to effective cross-domain generalization. To begin with, we tested the model in the absence of the Domain Discriminator, which performs adversarial alignment of the distributions of the source and target domains. The deletion of this component caused the decrease in accuracy by 4 percent, which means that the model failed to learn domain-invariant features. The feature extractor trained in the absence of adversarial training based on the Gradient Reversal Layer (GRL) generated representations that were more narrow to the specific source domain and did not work well with target data. Such an outcome establishes that the Domain Discriminator is an essential tool for bridging domain disparity and enabling sturdy performance across markets with varied statistical characteristics. Then, we examined the results of dropping the Shared Feature Extractor and substituting it with domain-specific encoders. The resulting change led to accuracy being reduced by 6 percent, which is the largest difference in the research. The absence of shared layers made it impossible to learn an integrated representation of financial features in different domains. In its place, it had to depend on autonomous pipelines between source and target data, and this did not facilitate the transfer of valuable knowledge between the source domain, which had a lot of data and the target domain, which had very little data. This finding highlights the significance of common layers in facilitating effective knowledge transfer and learning across domain characteristics. In general, the ablation study shows without any doubt that the Shared Feature Extractor and Domain Discriminator are crucial parts of the DTL architecture. Their composite effect helps the model to be highly accurate even in difficult situations when there are distribution shifts and the target domain data is limited, proving the performance and feasibility of the proposed method.

4.4. Practical Implications

- **Credit Inclusion:** One of the most important practical applications of the proposed Domain Transfer Learning (DTL) framework is its ability to increase credit inclusion, particularly for underbanked or credit-invisible groups. In a number of developing areas, there is a lack of access to formal credit as a result of the realization of past financial records. By using the labeled data in well-established markets and transferring the acquired knowledge to markets that possess little data, the model

can execute credible credit decisions even where little data is available in the local environment. This opens up the space to allow financial organizations to provide loans and credit products even to those who could not be included in the system, opening up the environment of financial equality and economic prosperity.

- **Regulatory Readiness:** Regulatory flexibility is also achieved through the modular structure of the DTL system, which allows institutions to adjust and comply with the regulations of their respective countries. Because the feature extractor and domain discriminator are independent of the classification head, the banks will be able to switch or adapt blocks to meet local policies on data use, privacy legislation, and reporting requirements. For example, when some features are banned in one country, retraining the extracted feature can be performed by updating the input space without retraining the entire model. This flexibility in the design of the structure will avoid the possibility of global deployment having a negative impact on legal or ethical standards, which is vital in financial systems with high levels of regulation.
- **Operational Scalability:** The given model is strategically addressed to be scalable in terms of operations, which allows the financial institutions to apply it in their wide geographical areas with little retraining involved. The model is also domain-adaptive, so it does not need a scratch when entering a new market. Rather, it is adaptable in new fields by doing fine-tuning using a minimal amount of unlabeled or semi-labeled information. This saves a significant amount of time and deployment cost, making the solution feasible and economical to use in real-life scenarios at scale. Consequently, although international institutions may continue to have a uniform framework for credit scoring, they will be able to adjust to local market characteristics.

5. Conclusion

To address the above problem, this paper introduces a new Deep Transfer Learning (DTL) model that considers the specifications of cross-market data, including limited data and distributional differences across different regions, to enhance the quality of cross-market credit scoring. This essentially mitigates the problem of availability of labeled data in emerging economies and enables the model to effectively tackle one of the most recurring

issues in global credit modeling, i.e., the paucity of sufficient and, by extension, high-quality labeled data in emerging markets. Our architecture consists of a shared feature extractor, a domain discriminator learned with adversarial learning and a classification head, letting the model not only learn domain-invariant representations that can be used across domains but also perform well in its primary task, classification. The performance of the DTL approach was compared with traditional machine learning and deep learning baselines on three different datasets, which showed full potential to use alternative financial ecologies as training sets and achieve overall success (in terms of accuracy, AUC and F1-score) in trading-related tasks.

Though the existing framework provides encouraging outcomes, some directions remain open to improvement. Among the major options is the incorporation of the few-shot learning methodology, which would enable the model to better fit target domains with an extremely small amount of labeled data. This may enhance applicability further in areas where labeled credit data is limited or nonexistent. The other notable expansion consists of integrating Explainable AI (XAI) modules. With the increasing regulatory pressure to introduce some degree of transparency in automated decision-making systems, adding an aspect of interpretability to the model will support the establishment of trust with stakeholders and facilitate compliance. Additionally, by adopting real-time adaptation through online learning, the implemented model will constantly adapt to new data streams, enabling it to dynamically respond to changing market situations, seasonal trends, or other economic shocks —much-needed features in the dynamic financial world.

Finally, the deep transfer learning provides an innovative solution involving the enhancement of financial inclusion, especially with Microfinance Institutions (MFIs) and banks who want to extend their service to the underbanked or credit-invisible population. This proposal closes the data-model mismatch between geographies, limiting the importance of having large labeled data and retraining expenses, and retaining high levels of predictive accuracy. With greater connectivity in financial systems internationally, however, there is a need to implement highly scalable, flexible, and interpretable tools to model credit across multiple markets. The study provides a viable and efficient solution to the current development of inclusive, data-driven financial ecosystems that will enable addressing the challenges of sustainable development and economic equality in the global community.

References

- [1] Thomas, L. C. (2000). A survey of credit and behavioural scoring: forecasting financial risk of lending to consumers. *International journal of forecasting*, 16(2), 149-172.
- [2] 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.
- [3] Baesens, B., Van Gestel, T., Viaene, S., Stepanova, M., Suykens, J., & Vanthienen, J. (2003). Benchmarking state-of-the-art classification algorithms for credit scoring. *Journal of the Operational Research Society*, 54(6), 627-635.
- [4] Crook, J. N., Edelman, D. B., & Thomas, L. C. (2007). Recent Developments in Consumer Credit Risk Assessment *European Journal of Operational Research*, 183(3), 1447-1465.
- [5] Zhou, Z. H. (2025). *Ensemble methods: foundations and algorithms*. CRC Press.
- [6] Sadhwani, A., Giesecke, K., & Sirignano, J. (2021). Deep learning for mortgage risk. *Journal of Financial Econometrics*, 19(2), 313-368.
- [7] Bengio, Y. (2012, June). Deep learning of representations for unsupervised and transfer learning. In *Proceedings of ICML workshop on unsupervised and transfer learning* (pp. 17-36). *JMLR Workshop and Conference Proceedings*.
- [8] Pan, S. J., & Yang, Q. (2009). A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10), 1345-1359.
- [9] Weiss, K., Khoshgoftaar, T. M., & Wang, D. (2016). A survey of transfer learning. *Journal of Big Data*, 3(1), 9.
- [10] Zeng, Z., Zhang, Z., Song, D., et al. (2021). Adversarial domain adaptation for cross-market credit risk prediction. *Knowledge-Based Systems*, 221, 106916.
- [11] Xiao, Y., Wang, H., & Liu, B. (2012). Feature selection based on selective ensemble for credit scoring. *Expert Systems with Applications*, 39(3), 2237-2247.
- [12] Galindo, J., & Tamayo, P. (2000). Credit risk assessment using statistical and machine learning: Basic methodology and risk modeling applications. *Computational Economics*, 15(1-2), 107-143.
- [13] Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, 16,

321–357.

- [14] Ahmed, M. I., & Rajaleximi, P. R. (2019). An empirical study on credit scoring and credit scorecard for financial institutions. *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, 8(7), 2278-1323.
- [15] Sadatrasoul, S. M., Gholamian, M., Siami, M., & Hajimohammadi, Z. (2013). Credit scoring in banks and financial institutions via data mining techniques: A literature review. *Journal of AI and Data Mining*, 1(2), 119-129.
- [16] Temelkov, Z., & Georgieva Svrtinov, V. (2024). The impact of AI on traditional credit scoring models. *Journal of Economics*, 9(1), 1-9.
- [17] Dastile, X., Celik, T., & Potsane, M. (2020). Statistical and machine learning models in credit scoring: A systematic literature survey. *Applied Soft Computing*, 91, 106263.
- [18] Huang, J., Chai, J., & Cho, S. (2020). Deep learning in finance and banking: A literature review and classification. *Frontiers of Business Research in China*, 14(1), 13.
- [19] Dumitrescu, E., Hué, S., Hurlin, C., & Tokpavi, S. (2022). Machine learning for credit scoring: Improving logistic regression with non-linear decision-tree effects. *European Journal of Operational Research*, 297(3), 1178-1192.
- [20] Hjelkrem, L. O., & Lange, P. E. D. (2023). Explaining deep learning models for credit scoring with SHAP: A case study using Open Banking Data. *Journal of Risk and Financial Management*, 16(4), 221.
- [21] Suryanto, H., Mahidadia, A., Bain, M., Guan, C., & Guan, A. (2022). Credit risk modeling using transfer learning and domain adaptation. *Frontiers in Artificial Intelligence*, 5, 868232.