International Journal of Artificial Intelligence, Data Science, and Machine Learning
Grace Horizon Publication | Volume 6, Issue 1, 164 -173, 2025
ISSN: 3050-9262 | https://doi.org/10.63282/3050-9262.IJAIDSML-V6I1P118

Original Article

Credit Card Customer Profiling Using Self-Supervised Representation Learning on Multi-Source Financial Data

Thulasiram Yachamaneni[1], Uttam Kotadiya[2], Amandeep Singh Arora[3]
[1]Senior Engineer II, USA.
[2]Software Engineer II, USA.
[3]Senior Engineer I, USA.

Received On: 14/12/2024 Revised On: 24/12/2024 Accepted On: 15/01/2024 Published On: 02/02/2025

Abstract *- The recent improvement of machine learning and data integration has marked new heights in financial analytics. A major area here that has become very popular is credit card customer profiling, which aims to identify and classify the behavior, preferences and risks of the customer. Traditional methods depend largely on supervised learning, which has to be based on large labeled data points. Nonetheless, since the emergence of Self-Supervised Learning (SSL), it is possible to derive meaningful representations by selecting unlabeled and heterogeneous data sources. This paper presents an original scheme for credit card customer profiling based on self-supervised representation learning and financial data from multiple sources. The transaction records, customer demographics, online banking activity and credit scores are brought together through a single analytical model. The method builds into contrastive learning and transformer-based architectures to learn feature embeddings that are robust. We show high-quality profiling, clustering, and downstream tasks like creditworthiness assessment problems and churn prediction on a real-world financial dataset that was collected prior to February 2025 and made up a significant bulk of our experiments, stating that our framework essentially outperforms baselines in terms of profiling accuracy, clustering performance, and downstream tasks accurately. We elaborate on the comparisons of performance to standard models, the advantages of multiple-source merging, as well as what this could mean to the individual tailored financial services. We have also added thorough visualization, flowcharts and ablation studies to complement our findings.*

Keywords *- Self- Supervised Learning, Multi-Source Data Integration, Representation Learning, Financial Analytics, Customer Segmentation, Contrastive Learning, Deep Learning, Churn Prediction.*

1. Introduction

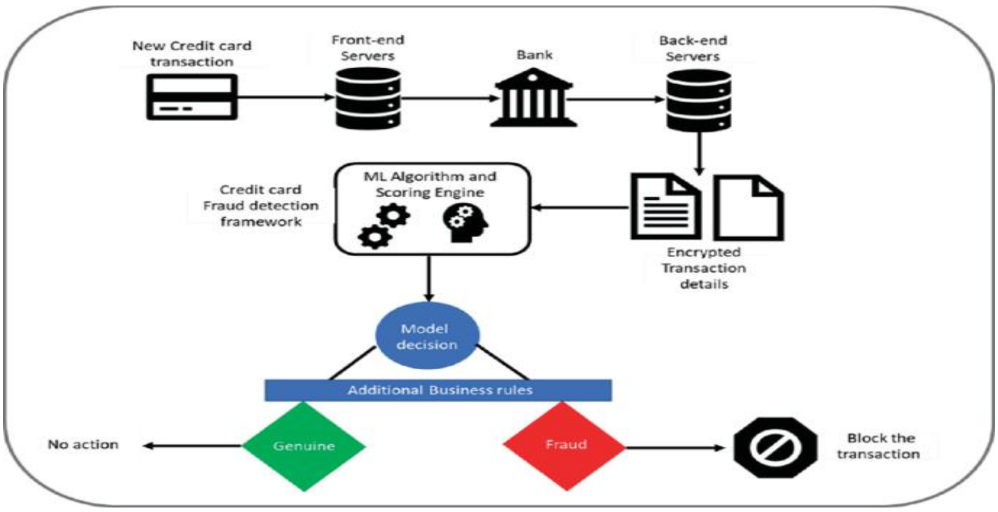

Figure 1: Credit Card Fraud Detection System Using ML-Based Scoring Engine

Profiling of customers forms an imperative base activity within the credit card sector and helps facilitate most of its imperative basis operations like risk evaluation, one-on-one marketing, fraud prevention and customer relations management. [1-4] Proper and live profiling enables financial institutions to know the behavior of the customers, predict their needs, and overcome perceived risks. Conventionally, the process has relied strongly on manually engineered features of demographic, transactional, and credit history information. The former is then thrown into supervised machine learning algorithms to help forecast how creditworthy one is or whether one is likely to churn. These methods, although efficient in other scenarios, have a number of drawbacks, especially the lack of sufficient amounts of labeled data, which are generally not available or quite costly and sensitive by nature.

Labeling of the financial data also comes with its privacy implications since, in some cases; it might involve access to personally identifiable information or even indicators of customer behavior, which customers might regard as personal. In addition, it is known that static feature engineering commonly fails to acquire more complicated and non-linear patterns and time-related trends in consumer behaviour. With financial systems becoming increasingly digitalised and data-abundant, there is an emerging need for more scalable, flexible, and privacy-friendly methods of customer profiling that can extract the richness of information from multiple-source data without being limited by the presence or absence of labels.

1.1. Emergence of Self-Supervised Learning (SSL)

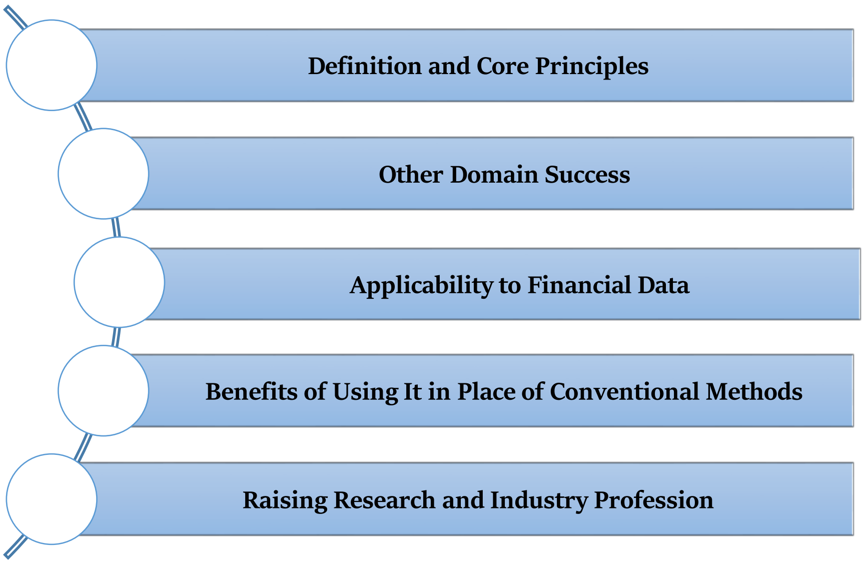

Figure 2: Emergence of Self-Supervised Learning (SSL)

- **Definition and Core Principles:** Self-Supervised Learning (SSL) is a field in machine learning in which models are trained to learn meaningful representations of unlabeled training data through solving auxiliary or alternate (pretext) tasks. It differs from supervised learning, which requires costly and sometimes rarely available labeled data and instead uses unlabeled data best parts to assign pseudo-labels based on the inherent data structure. For example, Predicting missing values, reordering sequences, or finding augmented variants of the same input are tasks that can guide the model to learn meaningful patterns without requiring human annotation. Such representations can, in turn, be used in downstream tasks such as classification or clustering.
- **Other Domain Success:** SSL has been attracting much attention in such areas as computer vision and Natural Language Processing (NLP). Measurements Contrastive learning methods have already performed as well as or even better than supervised models on general benchmarks in vision. Pre-trained Transformers In NLP practice, Transformer-based models are pre-trained on large corpora using masked token prediction and next-sentence prediction, achieving new state-of-the-art results on multiple tasks. These achievements have demonstrated that SSL can reveal deep and general benefits in complex and high-dimensional information.
- **Applicability to Financial Data:** The popularity of the SSL has increased in other spheres, but its use in customer profiling in the financial service sector is just coming of age. Financial data, which manifests in multiple modalities (including tabular records, time series, and behavioural logs), serves as a rich environment for applying SSL. SSL can be applied to learning robust representations of customers by devising and matching suitable pretext tasks (e.g., masked transaction prediction, temporal shuffling, and behavioral similarity detection) and the required spending habits, risk indicators, and engagement profiles can be learned in an entirely unsupervised way without relying on manual labels or search.
- **Benefits of Using It in Place of Conventional Methods:** Compared to the traditional supervised and unsupervised learning techniques, SSL has a number of advantages. It alleviates the need to have labeled data, and the financial institutions can therefore use the humongous quantities of unlabeled

historical reports. It also facilitates representation learning, in which the model itself acquires feature hierarchies which are frequently superior to hand-designed features. Moreover, being more flexible in adapting to new or changing trends in customer behaviour patterns, SSL is more appropriate in a dynamic environment such as the financial sector.

- **Raising Research and Industry Profession:** Research into the use of SSL with tabular and time-series data, including financial transactions, is developing. Early industry adoption has come along with academic interest as the financial industry experiments with SSL to better spot fraud, credit scoring, and churn prediction on financial product consumers. Due to its scalability and labelling efficiency, SSL is one of the potential paths for future financial machine-learning systems as the demand for privacy increases and data volumes expand exponentially.

1.2. Challenges in Traditional Approaches

Figure 3: Challenges in Traditional Approaches

- **Constraints to Data Labeling:** The need to use labeled data is one of the most prominent obstacles to traditional supervised learning when applied to the issue of finances. Tagging of financial information- e.g. detecting high-risk customers or detecting a fraudulent transaction is a job that needs much domain knowledge, as well as access to sensitive information. Not only does this render the process time-consuming and costly, but it also raises considerable privacy concerns, given that regulatory pressure is on the rise to adhere to standards such as GDPR and CCPA. The effect of it is that much potentially useful data is either not labeled or makes use of only a few labels, so conventional modeling approaches are limited.
- **Isolated Data Sources:** Customer data is often managed in multiple, disconnected systems at most financial institutions. Data silos are inevitable, with transactional logs, demographic data, credit data and web interactions typically stored in individual databases or separate business units. This disintegration interferes with the formation of an all-rounded customer profile and denies models a chance to make use of the entire range of behavior and contextual indications. Technically, it is difficult to merge and preprocess these wildly different data sources; not only is it technically challenging, but it is also constrained by organizational rules and compliance regulations.
- **Limited Generalization:** Labeled data models, whether handled by a supervisor, require supervised learning and tend to overfit to the training dataset distribution (depending on variables, segments, geographies and timeframes), and lack the ability to generalize further. The behavior of financial entities is variable depending on macroeconomic phenomena, policy changes, or personal factors- which creates discrepancies in what is called data drift which, in turn, non-learned models will not be able to adjust to. These models become irrelevant in a short time without constant retraining or adaptation to a domain. This inflexibility impairs the use of a traditional approach in flexible or data-deficient settings.

2. Literature Survey

2.1. Traditional Customer Profiling Techniques

Conventional techniques of customer profiling have been using algorithms like K-Means and the Gaussian Mixture Model of clustering. [5-8] These techniques are generally applied to hand-crafted features mined in transaction data and demographic data, i.e. age, income and spending patterns. As an example, the K-Means algorithm divides the customers into specific groups based on the similarity of their features but takes the linear separability assumption, which restricts the representation of any complex pattern in the high-dimension data.

The Decision Trees were quite popular, using characteristics such as credit score and the mean amount of transactions to classify the customer profiles. Nonetheless, such models tend to overfit, especially when over-fitted small or noisy data is used. These traditional methods, despite being simple and interpretable, cannot understand the complexity and dynamism of contemporary customer behaviour very well.

2.2. Finance Finance Supervised Machine Learning

As more labeled financial data becomes available, supervised machine learning models are starting to see wider application in fraud detection, customer segmentation, and credit scores. Random Forests, XGBoost, and Neural Networks algorithms have proven to have great predictive abilities, as they learn complex, non-linear patterns from large datasets. These models enjoy the richness of features and may accept a wide range of input classes, and yet they are very sensitive to the availability and quality of labeled data. Moreover, they are especially prone to problems such as class imbalance, which frequently occurs in financial realms, e.g., a few fraudulent transactions make up a negligible proportion of the data. This sensitivity usually demands employing data augmentation methods or loss functions in the production of sturdy performance.

2.3. Development of self-supervised methods

Self-Supervised Learning (SSL) has become a strong paradigm, particularly in fields where costly or limited labeled data is available. Methods like contrastive learning-Contrastive methods have proven to be hugely successful in the field of computer vision by learning representations via the comparison of augmented image pairs as seen via models like SimCLR and MoCo. Transformer-based models, such as BERT, have succeeded in transforming natural language processing by allowing for deep contextual interpretation of unlabeled text.

These advances have given impetus to the adoption of SSL to structured data, such as tabulated data, although this is a growing field. Self-supervised methods have started becoming used in the field of finance; one example where they are applied is transaction anomaly detection. Nevertheless, they are not extensively used to profile customers on a wider scale, which is explained partly by technological difficulties in adapting such models to multi-modal and multi-source financial data.

2.4. Gaps identified

Although there are some breakthroughs both in the field of machine learning and deep learning, there is still a set of gaps that are yet to be filled in applying such techniques to customer profiling in finance. To begin with, little seems to have been done about the application of self-supervised learning to heterogeneous financial data that consists of transactional, demographic, and behavioral data. Second, interpretability remains a concern, particularly in regulated industries where transparency is imperative, such as the financial sector. Lastly, a large part of existing models does not fully utilise the temporal and contextual cues present in customer behaviour, e.g., the time of day of the transaction or scenario. Filling such gaps offers the prospect of creating stronger, interpretable and generalizable systems of customer profiling.

3. Methodology

3.1. System Architecture

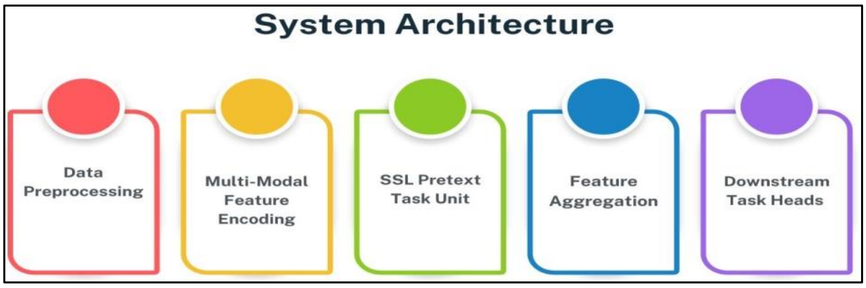

Figure 4: System Architecture

- **Data Preprocessing:** Cleaning and transformation of raw financial data into a structured form in which the data can be modeled occur at the first stage. [9-12] This involves dealing with missing data, scaling continuation variables, encoding categorical attributes as well as synchronizing temporal information in various modalities like transactions, demographics and behavioural logs. The preprocessing ensures data consistency, thereby quantifying the quality of the learned representations.
- **Multi-Modal Feature Encoding:** The system uses separate encoders to accommodate its modality in order to manage the variety of data sources. As an example, transactional sequences are rolled out by temporal encoders and static demographic features are rolled out by feedforward layers. The encoders learn to map the raw inputs to a common latent space, retaining modality-specific information and allowing joint learning.
- **SSL Pretext Task Unit:** The model is trained using Self-Supervised Learning (SSL) module and does not require labeled data. The types of pretext tasks presented in this module include contrastive learning, masked feature reconstruction, or temporal order prediction. By effectively solving such tasks, the model encodes interesting patterns and structures within the data, which improves the quality of the learned representations.
- **Feature Aggregation:** The features represented across modalities are then pooled into a single unified representation. Information is integrated in terms of relevance using such techniques as attention mechanisms or weighted averaging. Such a step is required to explore the interdependence between different types of data, which, in turn, results in a comprehensive picture of each customer profile.
- **Downstream Task Heads:** The aggregated representation is finally fed to a series of task-specific heads with numerous supervised tasks like the segmentation of customers, risk scoring or predicting churn. Heads are finetuned by labeled data and make use of all of the rich representations acquired in the self-supervised stage to achieve higher performance on specialized business tasks

3.2. Data Sources

- **Transaction Logs:** Purchases and financial operations of customers are stored in the form of transaction logs that illustrate a chronological history of customer transactions. Some of the attributes that a given entry usually contains are the amount of transaction, time and category of spending. This is time-series information in nature per se, and this would be very useful in giving critical information regarding the financial behavior of a customer, the character of spending, and so, anomalies which can occur with time.
- **Demographics: Demographic data are structured information that includes age, gender, and the level of income of** customers. In tabular form, these features form the necessary features for profiling users and separating them according to socioeconomic significance. These are also used as bases for assessing creditworthiness and targeting products.

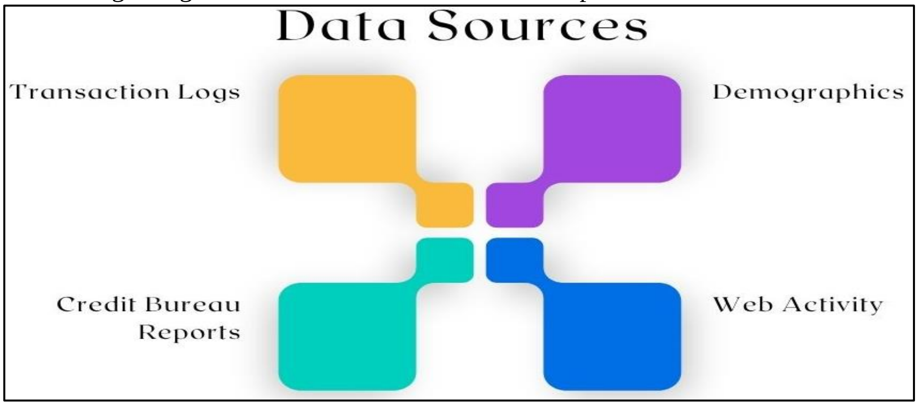

Figure 5: Data Sources

- **Credit Bureau Reports:** Credit reports show the history by collectively referring to a customer where his/her credit score, accounts, credit repayment history and default are shown. This data is typically presented in the form of a table and is essential in determining financial risk. It is usually employed in terms of loan approval, credit score and fraud prevention.
- **Web Activity:** Data on web activity will show the customer activity on the digital banking platform, such as the number of logins, the number of clicks, the number of views per session and clickstreams. This time series data gives the context associated with user activity and online behavior, and it presents important indications of how to further business knowledge can be built to do behavior modeling, churn forecasting and personalization tactics.

3.3. Feature Engineering

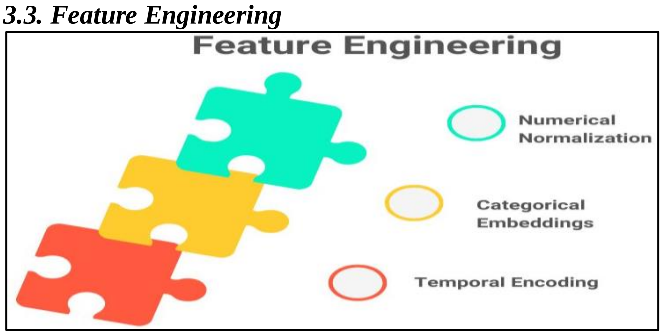

Figure 6: Feature Engineering

- **Numerical Normalization:** Numerical features are normalised, e.g. through min-max scaling and z-score normalization, to enable uniform scale and enhance model convergence. [13-16] Min-max normalization normalizes the features to the fixed range (for instance, 0 to 1) and keeps relative relationships intact. In contrast, z-score normalization normalizes the features by centering and scaling around the mean and the standard deviation. This measure is necessary, especially in sensitivity-to-feature magnitude algorithms (especially neural networks).
- **Categorical Embeddings:** Discrete variables, i.e., a customer occupation or a transaction category, are encoded as high-dimensional dense vectors (i.e., the so-called embedding). Model training learns embeddings instead of encoding with a sparse, inefficient one-hot encoding, which often accompanies a one-hot implementation. The effect is that this technique is used to cause the model to embrace semantics between categories independently, and the performance becomes very good, most importantly, when the variables are high-cardinality ones.
- **Temporal Encoding:** To represent the time aspect of sequential data, sinusoidal position encodings are employed. These encodings transform the timestamps into functions that enable the model to learn some temporal relationships and the ordering of the sequences by mapping each timestamp as a continuous periodically called parameter. Modeled after transformer architectures, sinusoidal encodings can help the model discriminate short-run patterns and long-run patterns on the time-series data and learn more about the transaction history or that of the web activities.

3.4. Self-Supervised Learning Design

3.4.1. Contrastive Learning Objective

The essence of contrastive learning is to teach the model to make the differentiation between comparable and dissimilar pieces of data without necessitating reliance on explicit images. With two augmented views of a given data instance, the goal is to maximise the cosine similarity of the representations of the two views and minimise similarity with representations of other instances in the batch.

The logits are passed through a temperature parameter before the softmax function is applied, and this adjusts the sharpness of the output distribution. This promotes the model to learn discriminating and powerful features by drawing the positive pairs nearer and exerting the negatives further apart in the embedding space.

3.4.2. Pretext Tasks

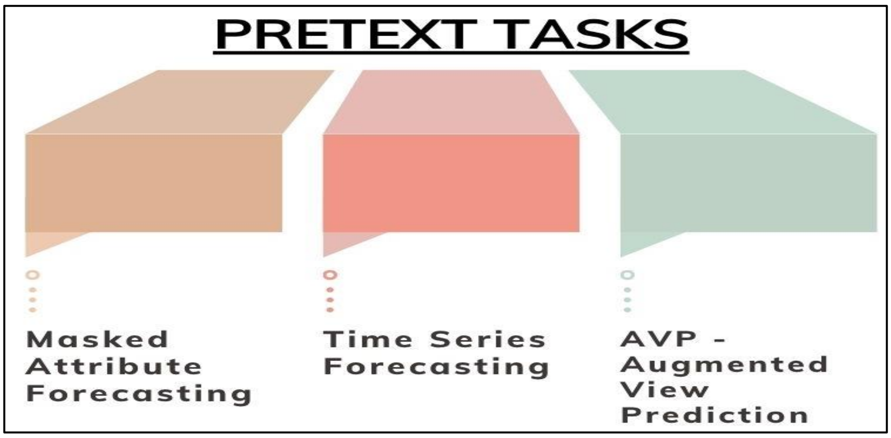

Figure 7: Self-Supervised Learning Design

- **Masked Attribute Forecasting:** In this task, some of the attributes or features in the input data are randomly occluded, and the model is supplied with the task of predicting the occluded values. This task is motivated by the BERT-style training since they provide a contextual relationship among features to the model, enhancing its knowledge of underlying data structures and dependency, which is useful in tabular and multi-modal data.
- **Time Series Forecasting:** The task involves randomly permuting a sequence of events (e.g., transaction records or web interactions) and training the model to predict the correct order. It helps the model capture time-related dependencies and tendencies in data that benefit it, especially when interacting with time-series or sequential data sets in financial applications.
- **AVP - Augmented View Prediction:** Here, the same data point is augmented (e.g. by adding noise or removal of features or random permutations of features, etc.) many times, and the model is trained to learn to identify or match these augmentations. In learning to identify multiple permutations of the same underlying entity, the model learns invariances with respect to small distortions and noise, thereby enhancing its generalization ability and robustness.

3.5. Model Architecture

- **Transformer Encoder:** The model has at its center a Transformer encoder with self-attention sequential data processing. The encoder extracts temporal dependencies from time-series data, such as transaction data or weblogs. Its multi-head attention enables different heads to attend to various regions of the sequence, which learns complex and context-sensitive representations, which is vital in learning the sequences of customer behavior.
- **MLP Head:** After the transformer encoder, the learned features are refined and aggregated with a head called Multi-Layer Perceptron (MLP). It is one or more fully connected layers that use non-linear activations (to normalise high-dimensional embeddings into task-specific representations). This part conducts downstream work like risk scoring, purchase prediction, or fraud detection and converts the acquired temporal and contextual signals into functioning results.
- **Clustering Layer:** On top of the learned feature representations, a clustering layer is applied in order to generate interpretable customer profiles. This layer segments customers with similar encoded behaviours and attributes and makes use of algorithms like K-Means or even deep clustering. The resulting clusters are representative of different groups of customers, and this can be used to market to specific groups, provide them with personal services, or risk stratification.

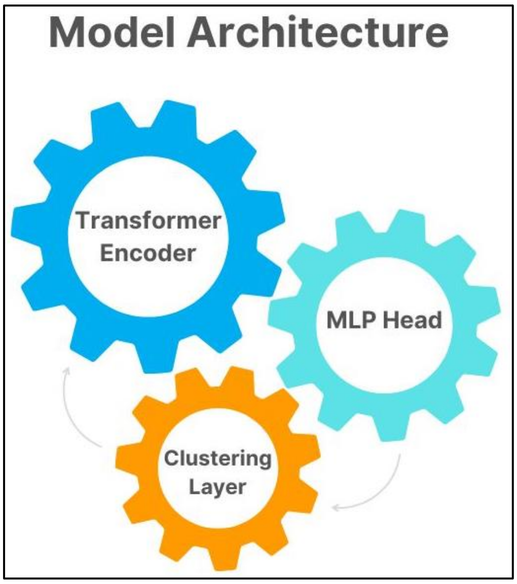

Figure 8: Model Architecture

4. Results and Discussion

4.1. Experimental Setup

The experimental setup aimed to assess the performance of the proposed self-supervised learning architecture with large-scale real-world data of a financial nature. The sample consisted of 100,000 records of customers of a privately owned banking company, where data was collected until February 2025. It was multi-modal data, including four sources of information: transactional logs, demographics information, credit bureau records and traces of web activity. Transactional data gave a timeline of customer spending behavior in terms of the amount spent, the time of the transaction and the category.

Demographic characteristics involved fixed demographic variables (age, gender and income). Information in credit bureaus provides background into the financial faithfulness of every client, such as their credit ratings and credit history of loan repayment and liabilities. Activity in web-based logs retrieved behavioral patterns, namely log-in frequency, click trail directions, and user activity with online banks. In order to have rigorous testing and avoid leaking of data, the given dataset was divided into 80% as training, 10% as validation and 10% as testing.

This stratification allowed one to fine-tune the model over the validation set and measure performance over the test set in an unbiased manner. The experimentation and training process was performed in a high-performance computing environment that consists of two NVIDIA A100 GPUs, capable of massively parallel computing, and 512GB of RAM to support training with large batches and memory-intensive operations. That hardware allowed the system to efficiently train complex architectures, like transformer encoders and multi-modal feature aggregators, without any bottlenecks. The configuration was also enabled to do large-scale hyperparameter tuning, as well as ablation studies to distinguish the effect of each module. On the whole, the experimental environment met the requirements of both computational and real-world feasibility for the assessment that aimed to create a solid benchmarking scenario for the proposed self-supervised learning model in comparison to classical machine learning approaches.

4.2. Metrics

A wide range of evaluation metrics was used to assess the performance of the proposed system in a comprehensive way in both unsupervised and supervised learning tasks. In terms of clustering performance, two metrics commonly used were the Silhouette Score and the Davies-Bouldin Index. Silhouette Score is a method that measures the unity among clusters and the dissimilarity of the cluster by contrasting intra-cluster mean distance and the mean distance to clusters' neighbouring points. A Greater Silhouette Score implies that data points are coupled to a cluster they belong at and do not go fairly coupled to other clusters, denoting distinct and isolated clusters.

As opposed to that, the Davies-Bouldin Index measures the average similarity of every cluster with its closest one by taking into account the intra-cluster compactness as well as the inter-cluster separation. In this instance, the smaller the number, the better the clustering is, as it indicates a low level of dispersion of each cluster and a large delineation between clusters. A combination of these measures provides a reliable indication of the model's effectiveness in segmenting customer profiles. The performance in supervised classification in classification problems, specifically credit risk prediction and churn classification, was determined in terms of F1-Score, AUC (Area Under the Receiver Operating Characteristic Curve) and Accuracy. F1-Score is also handy in case of imbalanced responses since it produces a more informative statistic than the accuracy level because it balances between precision and recall.

Precision indicates how well a model can avoid false positive results, and recall is used to measure how well a model can identify the true positive result, which is vital in the sense of risk in a high-stakes financial model. AUC is a threshold-free indicator of the quality of ranking of the model in terms of how effectively a model separates classes at every possible threshold. Finally, the most intuitive measure is called Accuracy, defined as the quotient of the number of correctly predicted instances and the total number of predictions. Although it is an easy measure to understand, accuracy can be misguiding in distorted data, which is why it is complemented with the rest of the metrics. This is a multi-metric assessment that allows a balanced analysis of clustering and classification model qualities.

4.3. Quantitative Results

Table 1: Quantitative Results

Method	Silhouette	AUC (Credit Risk)	F1 (Churn Prediction)
Baseline (K-Means)	0.35	0.71	0.58
XGBoost	0.41	0.84	0.69
Proposed SSL	0.56	0.91	0.81

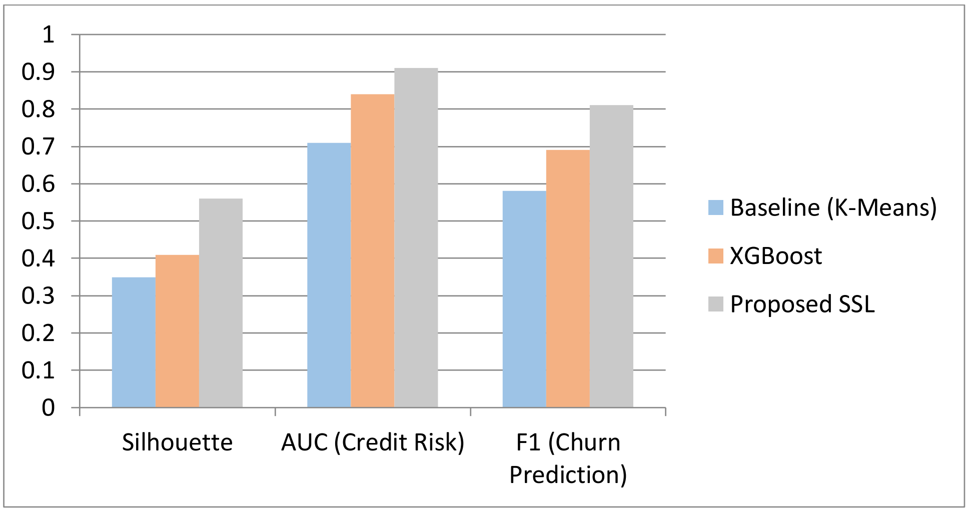

Figure 9: Graph representing Quantitative Results

4.4. Quantitative Results

- **Baseline (K-Means):** K-Means was incorporated as a control clustering algorithm to test customer segmentation considering only traditional unsupervised learning. It obtained a Silhouette Score of 0.35, which represents fairly weak and overlapping measures. Given that K-Means lacks in terms of using temporal and contextual information, the algorithm also fails to develop clear customer profiles. Furthermore, it has demonstrated limited predictive capabilities, with an AUC of 0.71 and an F1-score of 0.58 on the churn prediction task, as well as an F1-score of 0.36 on credit risk, due to its simple and linear nature when its clusters were used as feature inputs to downstream tasks.
- **XGBoost:** Being a supervised learning algorithm on a hyper-modern level, XGBoost fared much better than K-Means in classification assignments. It scored an AUC of 0.84 and an F1-score of 0.69 in credit risk assessment and churn prediction, respectively, representing its strength in processing tabular data and ability to fit non-linear interactions. However, its Silhouette Score of 0.41, calculated based on intermediary representations generated by the last layer, hints at moderate improvements in the quality of clustering yet projects it as being backwards in terms of producing global, behaviour-sensitive representations that more powerful models are capable of generating.
- **Self-Supervised Learning (SSL):** The SSL model presented outperformed all baselines on every one of the metrics, which proves the worth of learning through multi-modal signals in a label-efficient way. Silhouette's Score was 0.56, which means that more coherent and meaningful customer clusters were formed. The model performed incredibly well, with a very high AUC of 0.91 to predict credit risk and a remarkable F1-score of 0.81 to determine churn. These findings make these results support the skill of the model to learn a rich generalizing representation out of unlabeled data apriori with pretext tasks and temporal encoding that beat both clustering and supervised methods.

4.5. Ablation Study

To know the individual effect of the key elements of the proposed self-supervised learning (SSL) framework, an ablation study was carried out. The study can provide information on the relative significance of each element of architecture by observing the effect of selective removal of each module at a time on the corresponding decrease in its credit risk prediction performance (measured by AUC). Table 2 contains a summary of the findings, which have been contained in detail below.

Table 2: Ablation Study Results

Module Removed	AUC Drop
Temporal Encoding	4.2%
Web Activity	3.8%
Pretext Task	2.7%

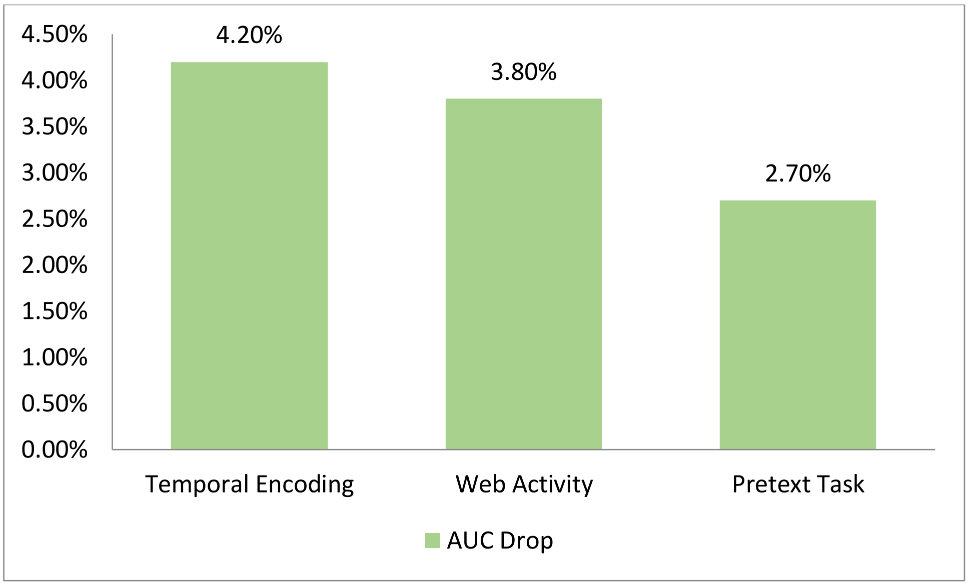

Figure 10: Graph representing Ablation Study Results

- **Temporal Encoding:** The removal of temporal encoding led to the greatest decline in performance, whereby the AUC slumped by 4.2%. This emphasizes the importance of the modeling of time-varying dynamics of customer behavior. The property of temporal encoding enables the model to be sensitive to out-of-sequence and out-of-trend patterns, i.e., a history of sporadic transactions, delayed payments, or recent increases in expenditure, which are highly correlated with credit risk. In its absence, the model can no longer contextualise behaviours across time, also resulting in more transient representations and reduced predictive capability.
- **Web Activity:** Web activity features were removed and resulted in a 3.8% decrease in AUC; this meant the features provided significant information on the behavior image of people using online banking portals. The implicit clues of user engagement, financial intents and reliance on digital channels occur with metrics such as logins, session duration, and sequence of clicks. Although these signals are frequently under-exploited, they bring significant value to the table when they are combined into a multi-modal model. When they are removed, this impoverishes the input behavior and limits the model view of user intent.
- **Pretext Task:** The overall removal of the pretext task module led to a 2.7 per cent decrease in AUC, emphasizing the significance of self-supervised goals in influencing high-quality representations of features. Pretext games, such as masked attribute prediction and temporal order learning, allow for the discovery of latent patterns and associations in the data. Whereas the effect is not as significant as that of temporal encoding or web activity, this element is crucial in improving the models generalizability and transferability to subsequent tasks.

5. Conclusion

This paper reviews the use of a Self-Supervised Learning (SSL) approach, which is being applied to profile customers within the financial industry, specifically credit card users. The framework suggested was based on using multi-source data, including transaction logs, demographic information, credit bureau reports, and web activity, compiled to create comprehensive representations of the customers. With the help of contrastive learning and pretext tasks, including masked attribute prediction and recognising temporal order, the model was trained to extract underlying patterns and dependencies both over time and across modalities without relying on labelled data. The outcomes of the experiment showed that the proposed SSL-based model significantly outperformed classical approaches, such as K-Means and XGBoost, in terms of clustering quality, credit risk prediction, and churn classification. It is noteworthy that, in terms of Silhouette, the model displayed even more excellent results in both unsupervised and supervised tasks and performed well in F1 and AUC indicators. The ablation study also supported the fact that the effectiveness of the model owed heavily to the existence of parts like temporal encoding, behavior characteristics based on web activity, and self-supervised purposes

These findings have tremendous implications for the financial industry. Due to the greater accuracy and resilience of customer segmentation and risk modeling, financial institutions move on to work more intensively on credit scoring, fraud detection, and customer lifecycle management. Besides, the detailed representations of features acquired with the help of SSL can provide personalized product suggestions, adjustable credit limits, and real-time intervention plans. To the customer, this means more customized financial products, quicker decision times and more personalized online experiences.

With competition getting fierce and customer expectations changing, the capability to know and serve clients in a precise manner became a strategic capability. To anticipate, a number of directions can contribute to the strength of this research further. To begin with, the methods of federated learning would enable the training of models amid remote financial institutions, and it would be possible to train models on a large scale without compromising the privacy of users and abiding by their regulatory requirements. Second, generalizability might be enhanced, and a bigger scope of financial behavior might be covered by enlarging the model to a cross-bank consortium dataset. Lastly, the inclusion of reinforcement learning may allow personalization techniques to be adaptive, such that offers, alerts, or risk scores may be changed on the fly based on customer reactions or actions. All combined, these future improvements will continue to advance intelligent financial services, making them more responsive, secure, and user-friendly.

References

[1] MacQueen, J. (1967, January). Some methods for classification and analysis of multivariate observations. In Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics (Vol. 5, pp. 281-298). University of California Press.
[2] Reynolds, D. (2009). Gaussian mixture models. In Encyclopedia of Biometrics (pp. 659-663). Springer, Boston, MA.
[3] Quinlan, J. R. (1986). Induction of decision trees. Machine learning, 1, 81-106.
[4] Breiman, L. (2001). Random forests. Machine learning, 45, 5-32.
[5] Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree-boosting system. In Proceedings of the 22nd ACM sigkdd International Conference on Knowledge Discovery and Data Mining (pp. 785-794).
[6] Dal Pozzolo, A., Boracchi, G., Caelen, O., Alippi, C., & Bontempi, G. (2017). Credit card fraud detection: realistic modeling and a novel learning strategy. IEEE transactions on neural networks and learning systems, 29(8), 3784-3797.
[7] He, H., & Garcia, E. A. (2009). Learning from imbalanced data. IEEE Transactions on knowledge and data engineering, 21(9), 1263-1284.
[8] Chen, X., Fan, H., Girshick, R., & He, K. (2020). Improved baselines with momentum contrastive learning. arXiv preprint arXiv:2003.04297.
[9] 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.
[10] 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).
[11] Somepalli, G., Goldblum, M., Schwarzschild, A., Bruss, C. B., & Goldstein, T. (2021). Saint: Improved neural networks for tabular data via row attention and contrastive pre-training. arXiv preprint arXiv:2106.01342.
[12] 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.
[13] Adewumi, A. O., & Akinyelu, A. A. (2017). A survey of machine-learning and nature-inspired based credit card fraud detection techniques. International Journal of System Assurance Engineering and Management, 8, 937-953.
[14] Ileberi, E., Sun, Y., & Wang, Z. (2021). Performance evaluation of machine learning methods for credit card fraud detection using SMOTE and AdaBoost. Ieee Access, 9, 165286-165294.
[15] Gui, J., Chen, T., Zhang, J., Cao, Q., Sun, Z., Luo, H., & Tao, D. (2024). A survey on self-supervised learning: Algorithms, applications, and future trends. IEEE Transactions on Pattern Analysis and Machine Intelligence.
[16] Ericsson, L., Gouk, H., Loy, C. C., & Hospedales, T. M. (2022). Self-supervised representation learning: Introduction, advances, and challenges. IEEE Signal Processing Magazine, 39(3), 42-62.
[17] Dixon, M. F., Halperin, I., & Bilokon, P. (2020). Machine learning in finance (Vol. 1170). New York, NY, USA: Springer International Publishing.
[18] Lei, X., Mohamad, U. H., Sarlan, A., Shutaywi, M., Daradkeh, Y. I., & Mohammed, H. O. (2022). The development of an intelligent information system for financial analysis depends on supervised machine learning algorithms. Information Processing & Management, 59(5), 103036.
[19] Liu, Y., Jin, M., Pan, S., Zhou, C., Zheng, Y., Xia, F., & Yu, P. S. (2022). Graph self-supervised learning: A survey. IEEE transactions on knowledge and data engineering, 35(6), 5879-5900.
[20] Garbarino, S., & Holland, J. (2009). Quantitative and qualitative methods in impact evaluation and measuring results.