Exploring the Potential of Learning in Credit Scoring Models for Alternative Lending Platforms

Lohith Paripati, Varun Nakra, Pandi Kirupa Gopalakrishna Pandian, Rahul Saoji, Bhanu Devaguptapu

Independent Researcher, USA.

Abstract

Credit scoring is a crucial financial task that has been studied using both statistical and Artificial Intelligence (AI) methods. The biggest risk to any bank or financial institution in terms of impact is credit risk. Since credit risk strategy sets pricing and may even have an impact on apparently unrelated areas like marketing and decision-making, accurate credit risk assessment has an impact on an organization's balance sheet and income statement. Traditional financial organisations have been using credit scoring algorithms extensively for a long time. There are restrictions when using these models in P2P lending. Initially, P2P credit data typically consists of sparse category categories and dense numerical information. Second, it is typically not possible to update the credit scoring algorithms that are now in use online. P2P lending involves a high volume of loan transactions, and the distribution of data changes in response to new information. When a credit scoring model is used without taking data updates into account, significant deviations or even failures in later credit assessments result. We provide a novel Online Integrated Credit Scoring Model (OICSM) for P2P Lending in this research. Gradient boosting decision tree models and neural networks are integrated by OICSM to improve the credit scoring model's ability to handle two different kinds of features and update live. Experiments are carried out both offline and online, utilising authentic and representative credit datasets to confirm the efficacy and superiority of the suggested model. According to experimental data, OICSM's advantage in deep learning over two features may greatly increase performance, and its live dynamic update capacity can further rectify model degradation.

Keywords: Artificial Intelligence (AI), Statistical Techniques, Credit Scoring, P2P Lending, Financial Transaction, Neural Network, (OICSM), Credit Datasets.

I. INTRODUCTION

Financial institutions are paying credit risk an unprecedented amount of attention as a result of the current global financial crisis. Financial organisations now primarily use credit scoring to evaluate credit risk, [1], enhance cash flow, lower potential hazards, and make managerial choices. The goal of credit scoring is to divide applicants into two groups: those with excellent credit and those with poor credit. Candidates with strong credit have a greater chance of repaying debt [1, 2]. Candidates with poor credit are more likely to default. The profitability of financial organisations depends heavily on the accuracy of credit scoring. Financial firms will suffer a significant loss if their credit scoring for applicants with poor credit is even 1% more accurate. Initially, credit scoring was determined subjectively based on individual experiences. Eventually, the 5Cs—the consumer's character, capital, collateral, capacity, and economic circumstances—became the basis for credit rating. However, the sheer volume of applications has made it difficult to handle the labour by hand [2]. Previous studies have examined two groups of automated credit scoring approaches: statistical techniques and Artificial Intelligence (AI) techniques [2, 3].

1.1 Alternative Credit Scoring Workflow Process

The processes involved in the process must be managed by an online lending platform in order to enable the automated management of the workflow for alternative credit analysis. These processes comprise data field organisation and classification, machine learning analysis, decision-making, and ongoing observation [2]. Shorter turnaround times for loan approvals are achievable using an online lending platform, which may be crucial in helping companies survive during difficult economic times [3]. It can also assist lenders in streamlining their procedures so that processing applications for loans costs less money.

Step I – Organising information from API channels: An online lending platform will obtain alternative data about the entities' credit history by contacting third-party data providers (with authorisation) in order to automate and simplifies the credit underwriting procedure. Open API projects are enabling new open banking API interfaces that can provide a straightthrough transfer of that additional data [3, 4]. An online lending system must keep the status of entities' permission up to date in order to comply with Open API policies. For instance, the platform ought to withdraw the entity's consent in the event that the consent's duration has passed or the entity chooses to do so [4]. The updated Payment Service Directive (PSD2), which became operative in the European Union (EU) in January 2018, is a noteworthy illustration of an Open API project. The European Bank Authority's Regulation Technical Regulations and PSD2 must be adhered to by all EU-regulated payment service providers [4, 5]. Financial institutions along with authorised third parties can access consumer financial information under the PSD2 Access to Account (XS2A) regulations. Before the structured data moves on to the next stage, it is imperative that the input data via APIs and other direct access methods be categorised and categorised. The data fields that are received over the API channel are pre-defined and organised.

Step II – Organising bank statement data with XS2A and OCR: The online financing platform allows entities to submit their own financial statement papers and bank records. Next, by using OCR technology, the data fields from the documents that were uploaded may be found, extracted, and captured. As an alternative, [5] entities may grant access to their banking statement data through methods that adhere to XS2A regulations.

Step III – Sorting data variables with natural language processing and machine learning: OCR and other immediate entry processes acquire unstructured and non-pre-defined data fields, which need structuring [5]. The classification of data fields using transactional text analysis is the next stage. To interpret input fields and classify them into the data variables needed by the model of machine learning, Natural Language Processing (NLP) technology is needed [6].

Step IV – Continued observation: Based on any updated alternative data obtained, the online lending platform has the ability to revaluate the entity's creditworthiness. Lenders can reduce their risk exposure and maintain control over their operations by closely monitoring any changes in the entity's creditworthiness. Another significant advantage of using an online lending platform for lenders over traditional credit scoring methods is that it allows them to continuously monitor the continuing financial risk associated with the entities in their loan portfolios [6, 7]. An online lending platform may assess lower loan credit limits more often and identify the following circumstances with continuous surveillance:

- Propensity for misbehaviour;
- A shift in the risk profile;
- Possible fraudulent loan applications;
- Indices of precarious credit situations.

Step V – Models of champions and opponents: The results of this study of the traditional credit scoring model and the alternative credit scoring model utilising a machine learning model must be merged in the last stage. Financial institutions would be wise to use both traditional and innovative credit scoring algorithms. Combining the data facilitates risk management for ongoing observation as well as ultimate decision-making. Alternative credit scoring models need historical data and iterative fine-tuning to increase their performance since machine learning algorithms are inherently flawed [6]. Therefore, while assessing creditworthiness, the insights produced by traditional models should always be considered as a baseline reference. The champion-challenger technique is a popular strategy for handling the coexistence of traditional and alternative scoring systems for credit.

The separate champion and challengers scores on credit should be utilised to create a combined ranking result in order to aid in decision-making [5, 6]. There are three distinct ways to display the overall scoring result of the challenger and champion designs.

- **Score Matrices:** A matrix representation that shows various bands of risk level may be used to compare the risk ratings of the champion and individuals opponents.
- **Decision tree:** Loan applications can be subjected to several evaluation processes use the championship and challengers models. The risk scores of each challenger and the champion can then be evaluated in a certain order, phase by phase [5]. A second opportunity may be given to loan applicants based on the order of evaluation.
- **Composite score:** A combination of scores that is produced from several separate individual scores may also be used to inform decisions. If the scores of the champion and several challengers add up to a composite risk score, artificial intelligence techniques like stacking and Logistic Regression (LR) may be applied.

The primary technique for reporting personal credit, credit scoring, is an automated evaluation tool used to decide whether to approve or deny loan requests. Based on the features of personal data, it divides the borrower into two categories: excellent and poor credit, and then determines whether to provide the loan. The technique of credit scoring has long been employed by conventional financial organisations. The credit scoring models and their applications in P2P lending are still in their infancy, given the many aspects of P2P lending [6]. Current research indicates that data mining and machine learning are the foundations of most credit scoring techniques [8, 9]. Two restrictions remain. First, poor categorisation results from complicated data types [6, 7]. Two types of characteristics are often found in the feature space of P2P credit data: sparse categorical variables (like gender and credit rating) and dense numerical features (like loan amount and asset-liability ratio). Unfortunately, current classifiers are often limited to analysing a single type of data [8]. For instance, a neural network model performs better on sparse category features than a tree classifier while processing dense numerical features [10]. For peer-to-peers credit databases with numerous data kinds, we must simultaneously ensure high performance and create an efficient model.

1.2 Objectives of the study

- • Examine how to create sophisticated credit scoring models using Machine Learning (ML) and Artificial Intelligence (AI) approaches.
- Evaluate the dependability and predicted accuracy of novel credit scoring models that use learning algorithms.

II. LITERATURE REVIEW

(Tigges, M., 2024) [11] In the financial industry, credit rating is a major factor in determining credit accessibility. This in turn has a major effect on the distribution of economic possibilities. Our work applies information asymmetry theory to the analysis of artificial intelligence and alternative data in fintech financing. We collect, analyse, and synthesise findings from a broad group of 26 specialists in fintech borrowing, Artificial Intelligence (AI), data sciences, machine learning, and academic by utilising a qualitative study methodology and the Gioia technique.

(Fu, G., 2020) [12] China's consumer loan market has grown rapidly since the late 1990s, particularly for short-term loans. As a result of advancements in financial technology, non-bank lending has outpaced bank lending in terms of growth. However, China lacks a common credit score and registration system that would help lenders in their credit evaluation and risk management procedures. For instance, internet lenders are unable to view an individual's bank credit records and vice versa. In light of this, this study aims to accomplish three goals.

(Pang, S., 2021) [13] This article suggests a way to assess the credit characteristics of borrowers based on the extreme learning machine, the Fuzzy C-Means (FCM) algorithm, and the computation of a confusion matrix by analysing the weight of evidence method and computing the Information Value (IV). We created a borrower credit score model by filtering credit rating indices. Furthermore, we developed methods to calculate the default loss rate and default likelihood. Additionally, the model categorises borrowers' credit attributes.

(Yan, J. 2024) [14] Examining the use of alternative information in bank financing to Small and Medium-Sized Enterprises (SMEs) is the aim of this study. Since alternative information is becoming a more significant component of SME financing in the future, it is crucial to comprehend its application in bank lending to SMEs. The study's results and conclusions contribute to the finance literature while also addressing a novel and intricate issue that is unique to the information technology and finance sectors: the application of Fintech in SME loan risk management.

III. METHODOLOGY

We propose the architecture of the P2P lending Online Integrated Credit Scoring Model (OICSM), which is built on the Deep-GBM architecture. OICSM combines the benefits of NN and GBDT. It may be updated live in addition to learning from two distinct P2P lending data sources. The OICSM architecture that this study proposes is seen in Figure 1. According to this strategy, pre-processing is done on the completed loan data in the data warehouse initially [11]. Using numerical and categorical characteristics, we partition the data into two categories and encode each one independently. In order to create an initial credit scoring model offline, the two types of information are then put into the "learning over two features" module [11, 12].

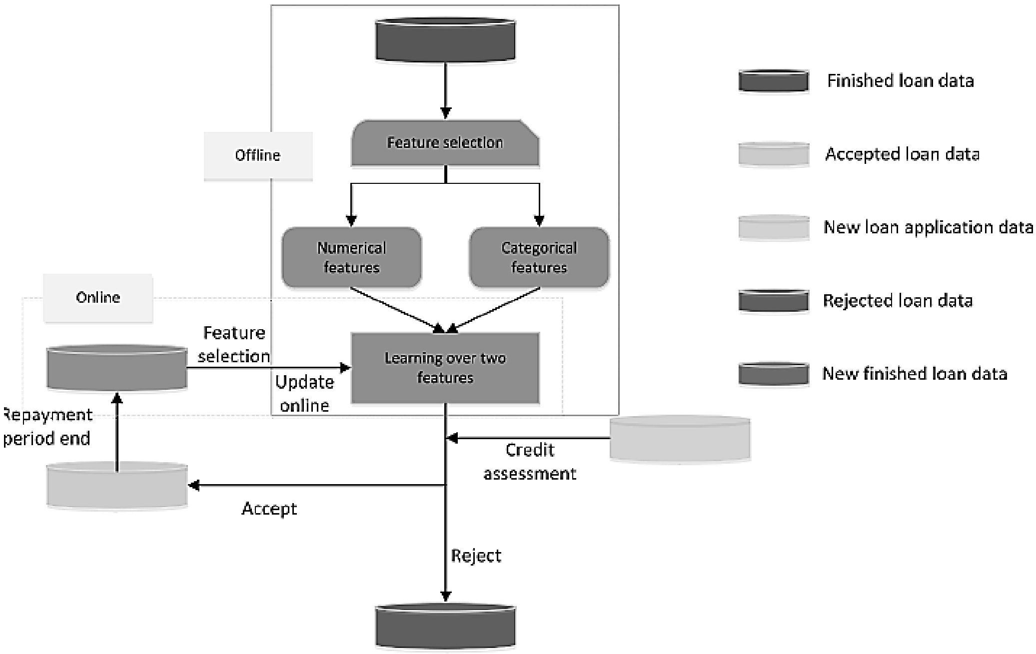

Fig. 1 The OICSM Structure. [12]

$E_{vi}(x_i) = embedding\_lookup(V_i, X_i)$, …….1

$yFM(x) = w_0(w,x) + \sum_{i=1}^{d}\sum_{i=1}^{d}\langle E_{vi}(x_i), E_{Vj}(x_j)\rangle x_j x_{j,}$ ……2

$yDeep(x) = N\left(\left[E_{v_1}(E_{V2})^T, E_{V_2}(x_2)^T \dots, E_{Vd}(x_d)^T\right]^T; \theta\right)$, ……3

$ycat(x) = yFM(x) + yDeep\ (x)$. ………4

$\underset{w, wo, w^t}{min} \frac{1}{i=1} \sum_{i=1}^{n} \mathcal{L}\ (W^T H(L^{t,i}; w^t) + w0, p^{t,i})$, …….5

$\underset{\theta}{min} \frac{1}{n} \mathcal{L}'(N(x'[\text{II}^t]; \theta), H^{t,i})$, …..6

$\underset{w, w_0, w^T}{min} \frac{1}{n} \sum_{i=1}^{n} \mathcal{L}(W^T H(\|_{t\in T}\ (L^{t,i}); w^T) + w_0, \sum_{t\in T} p^{t,i})$, …….7

$\mathcal{L}^T = \underset{\theta^T}{min} \frac{1}{n} \sum_{i=1}^{n} \mathcal{L}'(N([\|^T]; \theta^T), G^{T,i})$, …..8

$yT(x) = w^T \times N(x[\|^T]; \theta^T) + w_{0.}$ …….9

$yGBDT2NN(x) = \sum_{j=1}^{k} yT_j(x)$. …………10

IV. EXPERIMENTAL RESULT

We proceed to great depth on the experimental setups, models that are compared, data descriptions, along with specific experimental designs in this section [12].

We do pre-processing procedures such as eliminating post-loan features, deleting characteristics with smaller variations, deleting issues and features with a lot of missing values, and disregarding incomplete loan items since the original datasets contain a lot of post-loan characteristics and noisy data [12, 13]. Following that, Table 1 displays the specifics of the two datasets that we employed in the course of our study.

Table 1 Specifics of the datasets utilised in the research. Number of samples is called a sample. The number of numerical and categorical characteristics is denoted by Num and Cat, respectively. [13]

Dataset	Public dates	Sampling	N_{um}	C_{at}
LC	2015-17 Q1	0.9M	18	9
PPD	2013-11 2014-11	50k	209	18

The data from 2015 is utilised as the training set for the LC credit dataset, while the data from 2016Q1–2016Q2 (where Qn is the nth quarter) is used as the test set. The information collected in 2013.11-2014.08 is utilised as a training set and the remaining data is used as a test set for the PPD credit dataset [13]. Table 2 displays the specifics of the split datasets.

Table 2 Information on the credit datasets utilised in the offline trials.

Dataset	Training		Test	
	Public Date	Sample	Public Date	Sample
LC	2016	36909	2018Q12016Q2	14960
PPD	2013-11-2019-09	9695	2015.09-2018.15	23690

First, we partition each credit sample into six consecutive batches (Batches 0 to 5) based on the time slice in order to perform batch data division. In particular, we utilise the 2015 data as Batch 0 for the LC credit the data set and the 2016Q1–2017Q1 data are split into 5 successive Batches 1–5, with quarter (Q) serving as the period slice. The data from 2013.11–2014.05 is utilised as Batch 0 for the PPD credit dataset. The remaining data is split into 5 batches that use monthly (M) as the duration slice. Table 3 [14] displays the specifics of the split datasets.

Table 3 Specifics of the PPD and LC credit databases' batch data division. Number of samples is called a sample. Qn represents the nth quarter. [14]

Dataset	LC		PPD	
	Public Date	Sample	Public Date	Sample
Batch 0	2015	369600	2013.11-2016.09	21699
Batch 1	2016Q1	60616	2015.09	5964
Batch 2	2016Q2	21596	2015.09	9689
Batch 3	2016Q3	21591	2015.08	9648
Batch 4	2016Q4	21560	2015.08	5169
Batch 5	2017Q1	25932	2015.08	2966

Table 4 displays the offline experimental results for all models on two credit datasets. The findings indicate that:

- **LR performs the poorest:** Since LR finds it challenging to match the actual distribution of extensive and intricate credit data.
- **On both datasets, GBDT outperforms Deep FM and Wide & Deep:** In both the LC and PPD credit datasets, we can observe that there are noticeably more numerical features than category variables [15]. In learning tasks with additional numerical information, GBDT outperforms NN models (Wide & Deep and Deep FM) [14, 15].
- **As a crucial component of OICSM, GBDT2NN is extracted by GBDT:** The experimental findings demonstrate that in both credit datasets, GBDT2NN performs better than GBDT. This suggests that GBDT2NN can enhance GBDT performances through distillation of information for credit databases with both numerical and qualitative variables [15].
- **OICSM is the best Baseline Model Available:** According to the findings, OICSM improves AUC more than LR and by 1%–7% when compared to the other four baseline models [15].

European Economic Letters
ISSN 2323-5233
Vol 13, Issue 4 (2023)
http://eelet.org.uk

Table 4 The area under the curve for the LC and PPD databases from an offline experiment.

Model	LC	PPD
LR	0.595 ± 1*e*-8	0.0150 ± 1*e*-8
GBDT	0.590 ± 2*e*-5	0.0492 ± 4*e*-8
Wide & Deep	0.764 ± 3*e*-6	0.0590 ± 9*e*-8
GBDT2NN	0.742 ± 9*e*-2	0.0596 ± 4*e*-8
O1CSM	0.934 ± 2*e*-0	0.0542 ± 7*e*-8

V. DISCUSSION

Tables 5 to 6 display the outcomes of the models' online experiments using the LC and PPD datasets. The effectiveness of each model varies in accordance with the addition of new batch data, as demonstrated by the AUC score findings [16, 17]. The AUC values in Tables 5 and 6 are shown as figures in Figures 4 and 5, respectively, in order to further show these changes [17, 18].

Table 5: The online experiment's AUC scores for the LC sample. [18]

	Batch 1	Batch 2	Batch 3	Batch 4	Batch 4
GBDT	0.2670 ± 1*e*-8	0.4797 ± 1*e*-8	0.7279 ± 1*e*-8	0.6767 ±1*e*-8	0.6493 ± 1*e*-8
Wide & Deep	0.1530 ± 2*e*-5	0.7943 ± 4*e*-8	0.7473 ± 2*e*-5	0.0674 ± 2*e*-5	0.9674 ± 4*e*-8
Deep FM	0.7630 ± 1*e*-8	0.7457 ± 1*e*-8	0.7378 ± 1*e*-8	0.9764 ± 1*e*-8	0.4973 ± 1*e*-8
GBDT2NN	0.7996 ± 2*e*-5	0.8189 ± 4*e*-8	0.2974 ± 2*e*-5	0.6494 ± 2*e*-5	0.2967 ± 4*e*-8
OICSM-off	0.7659 ± 3*e*-6	0.4673 ± 9*e*-8	0.7944 ± 3*e*-6	0.3194 ± 3*e*-6	0.4697 ± 9*e*-8
OICSM	0.5793 ± 9*e*-2	0.7797 ± 4*e*-8	0.3469 ± 9*e*-2	0.4967 ± 9*e*-2	0.3264 ± 4*e*-8

Table 6 The online experiment's AUC values using the PPD dataset. [19]

	Batch 1	Batch 2	Batch 3	Batch 4	Batch 4
GBDT	0.2670 ± 1*e*-3	0.4797 ± 2*e*-4	0.7279 ± 1*e*-5	0.6767 ±2*e*-8	0.6493 ± 7*e*-1
Wide & Deep	0.1530 ± 4*e*-2	0.7943 ± 5*e*-9	0.7473 ± 6*e*-2	0.0674 ± 3*e*-5	0.9674 ± 4*e*-3
Deep FM	0.7630 ± 8*e*-0	0.7457 ± 5*e*-3	0.7378 ± 2*e*-6	0.9764 ± 9*e*-8	0.4973 ± 6*e*-3
GBDT2NN	0.7996 ± 5*e*-7	0.8189 ± 3*e*-6	0.2974 ± 1*e*-4	0.6494 ± 8*e*-5	0.2967 ± 2*e*-6
OICSM-off	0.7659 ± 7*e*-3	0.4673 ± 7*e*-4	0.7944 ± 6*e*-6	0.3194 ± 6*e*-6	0.4697 ± 0*e*-7
OICSM	0.5793 ± 8*e*-7	0.7797 ± 8*e*-2	0.3469 ± 3*e*-5	0.4967 ± 7*e*-2	0.3264 ± 1*e*-9

In result, the experiment's findings allow for the following deductions:

- OICSM outperforms all previous models in an offline simulation [20], demonstrating the necessity for simultaneous learning across both numerical and categorical data in a model for credit rating [21, 22].
- Updatable methods perform better in online experiments than non-updatable models, demonstrating the need to update the simulation model online dynamically with freshly generated [23] data in order to rectify the model deviations brought on by variations in the data distribution [24].
- OICSM outperforms all baseline models when combining offline and online tests, demonstrating the effectiveness of our model [25, 26] and its ability to address both issues in existing models at the same time [27].

VI. CONCLUSION

We provide in this research a novel OICSM credit assessment model for peer-to-peer lending. OICSM is divided into two sections. This integration may use its NN structure's batch processing capacity to update online dynamically in addition to learning over two features at once. We design both offline and online trials using two genuine and a good representation P2P lending credit datasets in order to confirm the efficacy and superiority of the suggested OICSM. The experimental findings show that OICSM performs better than any other baseline models. There is a cold start issue with this strategy. In the future, we'll attempt to apply the transfer learning approach to address this issue. With its larger user base and high transaction volume, OICSM is particularly well-suited for P2P lending scenarios where loan applicants' credit may be evaluated more accurately.

REFERENCES

[1] Hung, C., & Chen, J. H. (2009). A selective ensemble based on expected probabilities for bankruptcy prediction. Expert Systems with Applications, 36(3), 5297–5303.

[2] Karels, G., & Prakash, A. (1987). Multivariate normality and forecasting of business bankruptcy. Journal of Business Finance Accounting, 14(4), 573–593.

[3] Makowski, P. (1985). Credit scoring branches out. Credit World, 74(2), 30–37

[4] Li K, Niskanen J, Kolehmainen M, Niskanen M. Financial innovation: credit default hybrid model for SME lending. Expert Syst Appl. 2016; 61:343–355.

[5] Guegan D, Hassani B. Regulatory learning: how to supervise machine learning models? An application to credit scoring. J Financ Data Sci. 2018; 4(3):157–171.

[6] Fonseca DP, Wanke PF, Correa HL. A two-stage fuzzy neural approach for credit risk assessment in a Brazilian credit card company. Appl Soft Comput J. 2020; 92.

[7] Kvamme H, Sellereite N, Aas K, Sjursen S. Predicting mortgage default using convolutional neural networks. Expert Syst Appl. 2018; 102:207–217.

[8] Abdou HA, Tsafack MDD, Ntim CG, Baker RD. Predicting creditworthiness in retail banking with limited scoring data. Knowl Base Syst. 2016; 103:89–103.

[9] Zhang T, Zhang W, Xu W, Hao H. Multiple instance learning for credit risk assessment with transaction data. Knowl Base Syst. 2018; 161:65–77.

[10] Li Y, Wang X, Djehiche B, Hu X. Credit scoring by incorporating dynamic networked information. Eur J Oper Res. 2020; 286 (3):1103–1112.

[11] Tigges, M., Mestwerdt, S., Tschirner, S., & Mauer, R. (2024). Who gets the money? A qualitative analysis of fintech lending and credit scoring through the adoption of AI and alternative data. Technological Forecasting and Social Change, 205, 123491.

[12] Fu, G., Sun, M., & Xu, Q. (2020). An Alternative Credit Scoring System in China's Consumer Lending Market: A System Based on Digital Footprint Data. Available at SSRN 3638710.

[13] Pang, S., Hou, X., & Xia, L. (2021). Borrowers' credit quality scoring model and applications, with default discriminant analysis based on the extreme learning machine. Technological Forecasting and Social Change, 165, 120462.

[14] Yan, J. (2024). How the use of alternative information in risk management fintech platforms influences SME lending: a qualitative case study. Qualitative Research in Financial Markets.

[15] Silva DMB, Pereira GHA, Magalh~aes TM. A class of categorization methods for credit scoring models. Eur J Oper Res. 2021.

[16] Roa L, Correa-Bahnsen A, Suarez G, Cortes-Tejada F, Luque MA, Bravo C. Super-app behavioural patterns in credit risk models: financial, statistical and regulatory implications. Expert Syst Appl. December 2020; 2021:169.

[17] Zhou J, Wang C, Ren F, Chen G. Inferring multi-stage risk for online consumer credit services: an integrated scheme using data augmentation and model enhancement. Decis Support Syst. 2021, 113611.

[18] Jiang J, Liao L, Lu X, Wang Z, Xiang H. Deciphering big data in consumer credit evaluation. J Empir Finance. 2021; 62:28–45.

[19] Y. Xia, C. Liu, B. Da, and F. Xie, ''A novel heterogeneous ensemble credit scoring model based on bstacking approach,'' Expert Syst. Appl., vol. 93, pp. 182–199, Mar. 2018.

[20] X. Qiu, Y. Zuo, and G. Liu, ''ETCF: An ensemble model for CTR prediction,'' in Proc. 15th Int. Conf. Service Syst. Service Manage. (ICSSSM), Jul. 2018, pp. 1–5.

[21] C. J. Burges, ''from ranknet to lambdarank to lambdamart: An overview,'' Microsoft Res., Redmond, WA, USA, Tech. Rep. MSR-TR-2010-82, 2010.

[22] T. Chen and C. Guestrin, ''XGBoost: A scalable tree boosting system,'' in Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2016, pp. 785–794.

[23] V. Sugumaran, V. Muralidharan, and K. I. Ramachandran, ''Feature selection using decision tree and classification through proximal support vector machine for fault diagnostics of roller bearing,'' Mech. Syst. Signal Process., vol. 21, no. 2, pp. 930–942, Feb. 2007.

[24] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, ''LightGBM: A highly efficient gradient boosting decision tree,'' in Proc. 31st Int. Conf. Neural Inf. Process. Syst., 2017, pp. 3149–3157.

[25] T. Mikolov, K. Chen, G. S. Corrado, and J. Dean, ''Efficient estimation of word representations in vector space,'' in Proc. ICLR (Workshop Poster), 2013, pp. 1–12.

[26] S. Rendle, ''Factorization machines,'' in Proc. IEEE Int. Conf. Data Mining, Dec. 2010, pp. 995–1000.

[27] M. Fernández-Delgado, E. Cernadas, S. Barro, and D. Amorim, ''Do we need hundreds of classifiers to solve real world classification problems,'' J. Mach. Learn. Res., vol. 15, no. 1, pp. 3133–3181, 2014.

[28] Kaur, Jagbir. "Streaming Data Analytics: Challenges and Opportunities." International Journal of Applied Engineering & Technology, vol. 5, no. S4, July-August 2023, pp. 10-16.https://romanpub.com/resources/ijaetv5-s4-july-aug-2023-2.pdf

[29] Pandi Kirupa Kumari Gopalakrishna Pandian, Satyanarayan kanungo, J. K. A. C. P. K. C. (2022). Ethical Considerations in Ai and Ml: Bias Detection and Mitigation Strategies. International Journal on Recent and Innovation Trends in Computing and Communication, 10(12), 248–253. Retrieved from https://ijritcc.org/index.php/ijritcc/article/view/10511

[30] Ashok : "Ashok Choppadandi, Jagbir Kaur, Pradeep Kumar Chenchala, Akshay Agarwal, Varun Nakra, Pandi Kirupa Gopalakrishna Pandian, 2021. "Anomaly Detection in Cybersecurity: Leveraging Machine Learning Algorithms" ESP Journal of Engineering & Technology Advancements 1(2): 34-41.")

[31] Kaur, J. (2021). Big Data Visualization Techniques for Decision Support Systems. Jishu/Journal of Propulsion Technology, 42(4). https://propulsiontechjournal.com/index.php/journal/article/view/5701

[32] Ashok : "Choppadandi, A., Kaur, J.,Chenchala, P. K., Nakra, V., & Pandian, P. K. K. G. (2020). Automating ERP Applications for Taxation Compliance using Machine Learning at SAP Labs. International Journal of Computer Science and Mobile Computing, 9(12), 103-112. https://doi.org/10.47760/ijcsmc.2020.v09i12.014

[33] Chenchala, P. K., Choppadandi, A., Kaur, J., Nakra, V., & Pandian, P. K. G. (2020). Predictive Maintenance and Resource Optimization in Inventory Identification Tool Using ML. International Journal of Open Publication and Exploration, 8(2), 43-50. https://ijope.com/index.php/home/article/view/127

[34] Kaur, J., Choppadandi, A., Chenchala, P. K., Nakra, V., & Pandian, P. K. G. (2019). AI Applications in Smart Cities: Experiences from Deploying ML Algorithms for Urban Planning and Resource Optimization. Tuijin Jishu/Journal of Propulsion Technology, 40(4), 50-56.

[35] Case Studies on Improving User Interaction and Satisfaction using AI-Enabled Chatbots for Customer Service . (2019). International Journal of Transcontinental Discoveries, ISSN: 3006-628X, 6(1), 29-34. https://internatioaljournals.org/index.php/ijtd/article/view/98

[36] Kaur, J., Choppadandi, A., Chenchala, P. K., Nakra, V., & Pandian, P. K. G. (2019). Case Studies on Improving User Interaction and Satisfaction using AI-Enabled Chatbots for Customer Service. International Journal

[37] of Transcontinental Discoveries, 6(1), 29-34. https://internationaljournals.org/index.php/ijtd/article/view/98

[38] Choppadandi, A., Kaur, J., Chenchala, P. K., Kanungo, S., & Pandian, P. K. K. G. (2019). AI-Driven Customer Relationship Management in PK Salon Management System. International Journal of Open Publication and Exploration, 7(2), 28-35. https://ijope.com/index.php/home/article/view/128

[39] Ashok Choppadandi, Jagbir Kaur, Pradeep Kumar Chenchala, Akshay Agarwal, Varun Nakra, Pandi Kirupa Gopalakrishna Pandian, 2021. "Anomaly Detection in Cybersecurity: Leveraging Machine Learning Algorithms" ESP Journal of Engineering & Technology Advancements 1(2): 34-41.

[40] Ashok Choppadandi et al, International Journal of Computer Science and Mobile Computing, Vol.9 Issue.12, December- 2020, pg. 103-112. (Google scholar indexed)

[41] Choppadandi, A., Kaur, J., Chenchala, P. K., Nakra, V., & Pandian, P. K. K. G. (2020). Automating ERP Applications for Taxation Compliance using Machine Learning at SAP Labs. International Journal of Computer Science and Mobile Computing, 9(12), 103-112. https://doi.org/10.47760/ijcsmc.2020.v09i12.014

[42] Chenchala, P. K., Choppadandi, A., Kaur, J., Nakra, V., & Pandian, P. K. G. (2020). Predictive Maintenance and Resource Optimization in Inventory Identification Tool Using ML. International Journal of Open Publication and Exploration, 8(2), 43-50. https://ijope.com/index.php/home/article/view/127

[43] AI-Driven Customer Relationship Management in PK Salon Management System. (2019). International Journal of Open Publication and Exploration, ISSN: 3006-2853, 7(2), 28-35. https://ijope.com/index.php/home/article/view/128
[44] Pradeep Kumar Chenchala. (2023). Social Media Sentiment Analysis for Enhancing Demand Forecasting Models Using Machine Learning Models. International Journal on Recent and Innovation Trends in Computing and Communication, 11(6), 595–601. Retrieved from https://www.ijritcc.org/index.php/ijritcc/article/view/10762
[45] Tilala, Mitul, Saigurudatta Pamulaparthyvenkata, Abhip Dilip Chawda, and Abhishek Pandurang Benke. "Explore the Technologies and Architectures Enabling Real-Time Data Processing within Healthcare Data Lakes, and How They Facilitate Immediate Clinical Decision-Making and Patient Care Interventions." European Chemical Bulletin 11, no. 12 (2022): 4537-4542. https://doi.org/10.53555/ecb/2022.11.12.425.
[46] Mitul Tilala, Abhip Dilip Chawda, Abhishek Pandurang Benke, Akshay Agarwal. (2022). Regulatory Intelligence: Leveraging Data Analytics for Regulatory Decision-Making. International Journal of Multidisciplinary Innovation and Research Methodology, ISSN: 2960-2068, 1(1), 78–83. Retrieved from https://ijmirm.com/index.php/ijmirm/article/view/77
[47] Mitul Tilala. (2023). Real-Time Data Processing in Healthcare: Architectures and Applications for Immediate Clinical Insights. International Journal on Recent and Innovation Trends in Computing and Communication, 11(11), 1119–1125. Retrieved from https://www.ijritcc.org/index.php/ijritcc/article/view/10629
[48] Tilala, Mitul, and Abhip Dilip Chawda. "Evaluation of Compliance Requirements for Annual Reports in Pharmaceutical Industries." NeuroQuantology 18, no. 11 (November 2020): 138-145. https://doi.org/10.48047/nq.2020.18.11.NQ20244.
[49] Dodda, Suresh, Navin Kamuni, Venkata Sai Mahesh Vuppalapati, Jyothi Swaroop Arlagadda Narasimharaju, and Preetham Vemasani. "AI-driven Personalized Recommendations: Algorithms and Evaluation." Propulsion Tech Journal 44, no. 6 (December 1, 2023). https://propulsiontechjournal.com/index.php/journal/article/view/5587
[50] Kamuni, Navin, Suresh Dodda, Venkata Sai Mahesh Vuppalapati, Jyothi Swaroop Arlagadda, and Preetham Vemasani. "Advancements in Reinforcement Learning Techniques for Robotics." Journal of Basic Science and Engineering 19, no. 1 (2022): 101-111. ISSN: 1005-0930.
[51] Dodda, Suresh, Navin Kamuni, Jyothi Swaroop Arlagadda, Venkata Sai Mahesh Vuppalapati, and Preetham Vemasani. "A Survey of Deep Learning Approaches for Natural Language Processing Tasks." International Journal on Recent and Innovation Trends in Computing and Communication 9, no. 12 (December 2021): 27-36. ISSN: 2321-8169. http://www.ijritcc.org
[52] Jigar Shah , Joel lopes , Nitin Prasad , Narendra Narukulla , Venudhar Rao Hajari , Lohith Paripati. (2023). Optimizing Resource Allocation And Scalability In Cloud-Based Machine Learning Models. Migration Letters, 20(S12), 1823–1832. Retrieved from https://migrationletters.com/index.php/ml/article/view/10652
[53] Joel lopes, Arth Dave, Hemanth Swamy, Varun Nakra, & Akshay Agarwal. (2023). Machine Learning Techniques And Predictive Modeling For Retail Inventory Management Systems. Educational Administration: Theory and Practice, 29(4), 698–706. https://doi.org/10.53555/kuey.v29i4.5645
[54] Narukulla, Narendra, Joel Lopes, Venudhar Rao Hajari, Nitin Prasad, and Hemanth Swamy. "Real-Time Data Processing and Predictive Analytics Using Cloud-Based Machine Learning." Tuijin Jishu/Journal of Propulsion Technology 42, no. 4 (2021): 91-102.
[55] Nitin Prasad. (2022). Security Challenges and Solutions in Cloud-Based Artificial Intelligence and Machine Learning Systems. International Journal on Recent and Innovation Trends in Computing and Communication, 10(12), 286–292. Retrieved from https://www.ijritcc.org/index.php/ijritcc/article/view/10750
[56] Varun Nakra, Arth Dave, Savitha Nuguri, Pradeep Kumar Chenchala, Akshay Agarwal. (2023). Robo-Advisors in Wealth Management: Exploring the Role of AI and ML in Financial Planning. European Economic Letters (EEL), 13(5), 2028–2039. Retrieved from https://www.eelet.org.uk/index.php/journal/article/view/1514
[57] Varun Nakra. (2023). Enhancing Software Project Management and Task Allocation with AI and Machine Learning. International Journal on Recent and Innovation Trends in Computing and Communication, 11(11), 1171–1178. Retrieved from https://www.ijritcc.org/index.php/ijritcc/article/view/10684
[58] Joel lopes, Arth Dave, Hemanth Swamy, Varun Nakra, & Akshay Agarwal. (2023). Machine Learning Techniques And Predictive Modeling For Retail Inventory Management Systems. Educational Administration: Theory and Practice, 29(4), 698–706. https://doi.org/10.53555/kuey.v29i4.5645

[59] Big Data Analytics using Machine Learning Techniques on Cloud Platforms. (2019). International Journal of Business Management and Visuals, ISSN: 3006-2705, 2(2), 54-58. https://ijbmv.com/index.php/home/article/view/76

[60] Shah, J., Prasad, N., Narukulla, N., Hajari, V. R., & Paripati, L. (2019). Big Data Analytics using Machine Learning Techniques on Cloud Platforms. International Journal of Business Management and Visuals, 2(2), 54-58. https://ijbmv.com/index.php/home/article/view/76

[61] Cygan, Kamil J., Ehdieh Khaledian, Lili Blumenberg, Robert R. Salzler, Darshit Shah, William Olson, Lynn E. Macdonald, Andrew J. Murphy, and Ankur Dhanik. "Rigorous Estimation of Post-Translational Proteasomal Splicing in the Immunopeptidome." bioRxiv (2021): 1-24. https://doi.org/10.1101/2021.05.26.445792

[62] Shah, Darshit, Ankur Dhanik, Kamil Cygan, Olav Olsen, William Olson, and Robert Salzler. "Proteogenomics and de novo Sequencing Based Approach for Neoantigen Discovery from the Immunopeptidomes of Patient CRC Liver Metastases Using Mass Spectrometry." The Journal of Immunology 204, no. 1_Supplement (2020): 217.16-217.16. American Association of Immunologists.

[63] Mahesula, Swetha, Itay Raphael, Rekha Raghunathan, Karan Kalsaria, Venkat Kotagiri, Anjali B. Purkar, Manjushree Anjanappa, Darshit Shah, Vidya Pericherla, Yeshwant Lal Avinash Jadhav, Jonathan A.L. Gelfond, Thomas G. Forsthuber, and William E. Haskins. "Immunoenrichment Microwave & Magnetic (IM2) Proteomics for Quantifying CD47 in the EAE Model of Multiple Sclerosis." Electrophoresis 33, no. 24 (2012): 3820-3829. https://doi.org/10.1002/elps.201200515.

[64] Big Data Analytics using Machine Learning Techniques on Cloud Platforms. (2019). International Journal of Business Management and Visuals, ISSN: 3006-2705, 2(2), 54-58. https://ijbmv.com/index.php/home/article/view/76

[65] Cygan, K. J., Khaledian, E., Blumenberg, L., Salzler, R. R., Shah, D., Olson, W., & ... (2021). Rigorous estimation of post-translational proteasomal splicing in the immunopeptidome. bioRxiv, 2021.05.26.445792.

[66] Mahesula, S., Raphael, I., Raghunathan, R., Kalsaria, K., Kotagiri, V., Purkar, A. B., & ... (2012). Immunoenrichment microwave and magnetic proteomics for quantifying CD 47 in the experimental autoimmune encephalomyelitis model of multiple sclerosis. Electrophoresis, 33(24), 3820-3829.

[67] Mahesula, S., Raphael, I., Raghunathan, R., Kalsaria, K., Kotagiri, V., Purkar, A. B., & ... (2012). Immunoenrichment Microwave & Magnetic (IM2) Proteomics for Quantifying CD47 in the EAE Model of Multiple Sclerosis. Electrophoresis, 33(24), 3820.

[68] Raphael, I., Mahesula, S., Kalsaria, K., Kotagiri, V., Purkar, A. B., Anjanappa, M., & ... (2012). Microwave and magnetic (M2) proteomics of the experimental autoimmune encephalomyelitis animal model of multiple sclerosis. Electrophoresis, 33(24), 3810-3819.

[69] Salzler, R. R., Shah, D., Doré, A., Bauerlein, R., Miloscio, L., Latres, E., & ... (2016). Myostatin deficiency but not anti-myostatin blockade induces marked proteomic changes in mouse skeletal muscle. Proteomics, 16(14), 2019-2027.

[70] Shah, D., Anjanappa, M., Kumara, B. S., & Indiresh, K. M. (2012). Effect of post-harvest treatments and packaging on shelf life of cherry tomato cv. Marilee Cherry Red. Mysore Journal of Agricultural Sciences.

[71] Shah, D., Dhanik, A., Cygan, K., Olsen, O., Olson, W., & Salzler, R. (2020). Proteogenomics and de novo sequencing based approach for neoantigen discovery from the immunopeptidomes of patient CRC liver metastases using Mass Spectrometry. The Journal of Immunology, 204(1_Supplement), 217.16-217.16.

[72] Shah, D., Salzler, R., Chen, L., Olsen, O., & Olson, W. (2019). High-Throughput Discovery of Tumor-Specific HLA-Presented Peptides with Post-Translational Modifications. MSACL 2019 US.

[73] Srivastava, M., Copin, R., Choy, A., Zhou, A., Olsen, O., Wolf, S., Shah, D., & ... (2022). Proteogenomic identification of Hepatitis B virus (HBV) genotype-specific HLA-I restricted peptides from HBV-positive patient liver tissues. Frontiers in Immunology, 13, 1032716.

[74] Big Data Analytics using Machine Learning Techniques on Cloud Platforms. (2019). International Journal of Business Management and Visuals, ISSN: 3006-2705, 2(2), 54-58. https://ijbmv.com/index.php/home/article/view/76

[75] Pavan Ogeti, Narendra Sharad Fadnavis, Gireesh Bhaulal Patil, Uday Krishna Padyana, Hitesh Premshankar Rai. (2022). Blockchain Technology for Secure and Transparent Financial Transactions. European Economic Letters (EEL), 12(2), 180–188. Retrieved from https://www.eelet.org.uk/index.php/journal/article/view/1283

[76] Challa, S. S. S., Chawda, A. D., Benke, A. P., & Tilala, M. (2023). Regulatory intelligence: Leveraging data analytics for regulatory decision-making. International Journal on Recent and Innovation Trends in Computing and Communication, 11(11), 1426-1434. Retrieved from http://www.ijritcc.org

[77] Fadnavis, N. S., Patil, G. B., Padyana, U. K., Rai, H. P., & Ogeti, P. (2021). Optimizing scalability and performance in cloud services: Strategies and solutions. International Journal on Recent and Innovation Trends in Computing and Communication, 9(2), 14-23. Retrieved from http://www.ijritcc.org

[78] Challa, S. S. S., Tilala, M., Chawda, A. D., & Benke, A. P. (2021). Navigating regulatory requirements for complex dosage forms: Insights from topical, parenteral, and ophthalmic products. NeuroQuantology, 19(12), 971-994. https://doi.org/10.48047/nq.2021.19.12.NQ21307

[79] Fadnavis, N. S., Patil, G. B., Padyana, U. K., Rai, H. P., & Ogeti, P. (2020). Machine learning applications in climate modeling and weather forecasting. NeuroQuantology, 18(6), 135-145. https://doi.org/10.48047/nq.2020.18.6.NQ20194