Machine Learning and Transfer Learning Applications in Finance Sector

Zaid Altukhi
zaltukhi@gmu.edu

***Abstract*—The American Rescue Plan of 2021 cause consumers to spend and put money back into the economy in the COVID19 pandemic. As we begin to see light at the end of the tunnel, this is good news for the banking industry and especially beneficial to the working-class people.Here is the reality; in able for the banks to grow, the banks need people, and the people need banks for its money. It's a very simple complex but yet simple relationship. Think of it as one hand washes the other. Institutions such as banks can only become bigger if they are able to profit and capitalize from the interest it gains from its borrowers.**

***Index Terms*—creditworthiness, institutional, discrimination,finance**

I. Introduction

Based on the applicant's current credit score, the lender takes into consideration how likely the applicant will repay the obligations of the debt on time. Once the lender determines if the applicant is deemed not a risk and is worthy of credit, the decision will be made if eligible to get the loan. [1], [2] There are some very important factors other than credit score that determine the approval of a loan application status. In other words, proving to the creditors and lenders that timely payments will be made will establish trust with any applicant. Throughout this project, we will attempt to examine other factors and find some patterns using datasets to determine who can establish creditworthiness, without the use of institutional discrimination [1].Machine learning models have different application in finance and business section [3] [4] [5]–[16]

[17]and Health [6], [18] [19] [19] and Social Media [18], [20]–[31].

II. Literature review

The goal is to incorporate new data and harness AI to expand credit to consumers who need it on better terms than are currently provided. AI can easily go in the other direction to exacerbate existing bias, creating cycles that reinforce biased credit allocation while making discrimination in lending even harder to find. Protection against discrimination in a risk-based pricing system layered on top of a society with centuries of institutional discrimination. [32]

AI will build on our existing system's dual goals of pricing financial services based on the true risk the individual consumer poses while aiming to prevent discrimination (e.g., race, gender, DNA, marital status, etc.). Currently, there are not enough sources of standardized information to base decisions and too little credit being made available. Those conditions allowed rampant discrimination by loan officers who could simply deny people because they "didn't look credit worthy." [32], [33]

III. Datasets

Getting reliable, relevant and clean dataset from the finance industry is a tedious task as the industry is governed by the strict federal rules. Cleaning the data to remove Personally Identifiable Information and share with general audience itself is a major task. After performing data exploration extensively, we were able to get hold of three different datasets which are from reliable sources and are relevant to our research topic.

A. BigML.com. The dataset title: Loan Risk Data [34]

The dataset is taken from BigML.com. The dataset title: Loan Risk Data [34] link: *https://bigml.com/user/bigml/gallery/dataset/4f89c38f1552686 459000033#info* The dataset is about loan risk data. It is having around 1000 records which shows the creditworthiness of applications and contains 21 attributes.

B. Lending Club Loan Dataset 2007_2011

The other dataset we use in our project is "Lending Club Loan Dataset 2007_2011" which is a big data set; it contains around 39,000 rows and 111 columns [35]. *Link: https://www.kaggle.com/imsparsh/lending-club-loan- dataset-2007-2011?select=loan.csv*

C. Data on loan delinquency

The dataset is about loan delinquency, data has around 50,000 loans data and 19 attributes. The size of the dataset is 4.3 MB. [36]

Link: *https://bigml.com/user/bigml/gallery/dataset/4f8b5eae155268783e*

IV. Purposed approach

Like any other data science project, the approach we are planning to use includes multiple steps. The steps we are planning to follow are described in Figure **??**.

The goal is already defined for this project, which is to find the impact of social and economic factors on Creditworthiness. While we already found one dataset [37], [38] to start with the research, we will continue the research for more data sources, which could help us investigate the problem. Subsequently a data clean up activity is planned and then normalization and grouping of data is also planned. In the next stage, we will be looking for patterns and derive the required knowledge to address the topic under consideration. Finding insights and

Fig. 1. ATTRIBUTES FROM LOAN RISK DATA

Sl No	Attribute Name	Data type	Description
1	checking_status	Categorical	Status of the loan (status can be in process, grace, repayment, forbearance, etc.,)
2	duration	Numeric	Measure of the bond with sensitivity of price, or other debt to change in interest rates.
3	credit_history	Categorical	Records of how a person maintained their credit history in the past.
4	purpose	Categorical	Purpose of the loan
5	credit_amount	Numeric	Amount the customer promises to repay
6	savings_status	Categorical	Status of the savings account
7	employment	Categorical	The customer's employment
8	installment_commitment	Numeric	Includes all the terms and conditions as per amount
9	personal_status	Categorical	Personal Status of the customer
10	other_parties	Categorical	All other parties included in the loan agreement
11	residence_since	Numeric	Dates of since when a person is living at a particular residence.
12	property_magnitude	Categorical	Type and importance of the loan
13	age	Numeric	Age of the customer
14	other_payment_plans	Categorical	Other payment plans included in the bank
15	housing	Categorical	Housing status of the customer
16	existing_credits	Numeric	Available information/ history of the customer
17	job	Categorical	A customer's basic job information
18	num_dependents	Numeric	Number of dependents included in the loan
19	own_telephone	Categorical	If customers have a contact number
20	foreign_worker	Categorical	If the customer is a foreign worker.
21	class	Categorical	Two different classes of loan-good/bad

visualizing the same will be done at this stage. Next step of using machine learning is a bit ambitious for us, with which we will try to find clusters within the dataset(s) under consideration to gain necessary wisdom to solve the problem under consideration.

V. Framework

Through the research done so far, we identified that the below attributes have a significant impact on determining the credit worthiness of a person.

- Education
- Marital Status
- Employment
- Income
- Property Type
- Rent and/or Utility payments
- Purchase history

In the next phase we are looking to find patterns among them and the weightage of those attributes on determining creditworthiness.

In typical situations, lenders use the credit score as the main factor to determine if people were eligible for a loan. Since 45 million people do not have a credit score, we will use the AI, ML to create a new framework called "qualification score". The qualification score is a calculation of multiple factors

33	total_rec_late_f ee	Numeric	Late fees received to date
34	last_pymnt_d	Date	Last month payment received
35	last_pymnt_am nt	Numeric	Last total payment amount received
36	next_pymnt_d	Date	Next schedule payment date
37	policy_code	Categorical	Policy code if publicly available or not.
38	application_typ e	Categorical	Individual application or Joint application
39	pub_rec_bankr uptcies	Numeric	Number of public record bankruptcies

such as social, educational, and financial factors to name a few. Based on this information, then the lender can use the "qualification score" instead of the credit score to determine creditworthiness for those who do not have one.

We will use AI to adopt a new system to find a proper way to rank each element based on the datasets we have. As a result, a will inherit and then calculate the "qualification score". [33], [39]

VI. Findings

VII. Feature walk through

All applications should be given an equal opportunity and accommodations to help gain access to the same lender model. With the help of the most powerful tool available in the world, artificial intelligence and bank institutions are collaborating to provide new and alternative approaches to help increase credit scores. With the help of alternative data source, Artificial Intelligence and Machine Learning will make the decision-making process for the lender much faster and provide an insight on whom will repay their loans. [40]–[42]

References

[1] D. Valatsas, "Coronavirus recovery: why non-us central banks must protect countries from rates cycle." https://www.scmp.com/comment/opinion/article/3127424/coronavirus-recovery-why-non-us-central-banks-must-protect, 29 March 2021.

[2] A. Iyyengar, "40% of financial services use ai for credit risk management. want to know why?." https://blog.aspiresys.com/banking-and-finance/40-financial-services-use-ai-for-credit-risk-management/, 11 August 2020.

[3] "45 million americans have no credit score.." https://www.cnbc.com/2015/05/05/credit-invisible-26- million-have-no-credit-score.html, 16 March 2021.

[4] Nina, "Credit cards are for transactions; not to borrow money." https://www.queercents.com/credit-cards-are-for-transactions-not-to-borrow-money/, 14 Agust 2009.

[5] M. Heidari and S. Rafatirad, "Using transfer learning approach to implement convolutional neural network model to recommend airline tickets by using online reviews," in *2020 15th International Workshop on Semantic and Social Media Adaptation and Personalization (SMA*, pp. 1–6, 2020.

[6] J. Wang, H. Song, and X. Zhou, "A collaborative filtering recommendation algorithm based on biclustering," in *2015 IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems (CYBER)*, pp. 803–807, 2015.

[7] M. Heidari and S. Rafatirad, "Bidirectional transformer based on online text-based information to implement convolutional neural network model for secure business investment," in *IEEE 2020 International Symposium on Technology and Society (ISTAS20), ISTAS20 2020*, 2020.

[8] S. Chen, S. Owusu, and L. Zhou, "Social network based recommendation systems: A short survey," in *2013 International Conference on Social Computing*, pp. 882–885, 2013.

[9] S. Chen, S. Owusu, and L. Zhou, "Social network based recommendation systems: A short survey," in *2013 International Conference on Social Computing*, pp. 882–885, 2013.

[10] S. Lin, C. Liu, and Z.-K. Zhang, "Multi-tasking link prediction on coupled networks via the factor graph model," in *IECON 2017 - 43rd Annual Conference of the IEEE Industrial Electronics Society*, pp. 5570–5574, 2017.

[11] M. Heidari and S. Rafatirad, "Semantic convolutional neural network model for safe business investment by using bert," in *IEEE 2020 Seventh International Conference on Social Networks Analysis, Management and Security, SNAMS 2020*, 2020.

[12] Y. Chu, F. Huang, H. Wang, G. Li, and X. Song, "Short-term recommendation with recurrent neural networks," in *2017 IEEE International Conference on Mechatronics and Automation (ICMA)*, pp. 927–932, 2017.

[13] M. Heidari, S. Zad, and S. Rafatirad, "Ensemble of supervised and unsupervised learning models to predict a profitable business decision," in *IEEE 2021 International IOT, Electronics and Mechatronics Conference, IEMTRONICS 2021*, 2021.

[14] C. Yang, X. Chen, T. Song, B. Jiang, and Q. Liu, "A hybrid recommendation algorithm based on heuristic similarity and trust measure," in *2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/ 12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE)*, pp. 1413–1418, 2018.

[15] S. Ji and J. Liu, "Interpersonal ties and the social link recommendation problem," in *2019 6th International Conference on Systems and Informatics (ICSAI)*, pp. 456–462, 2019.

[16] M. Heidari, S. Zad, B. Berlin, and S. Rafatirad, "Ontology creation model based on attention mechanism for a specific business domain," in *IEEE 2021 International IOT, Electronics and Mechatronics Conference, IEMTRONICS 2021*, 2021.

[17] M. Backman, "1 in 5 americans have no credit score. here's why that's a problem." https://www.fool.com/the-ascent/credit-cards/articles/1-5-americans-have-no-credit-score-heres-why-problem/, 5 February 2021.

[18] A. Randazzo and C. Young, "What caused the meltdown: A financial crisis faq." https://reason.org/faq/what-caused-the-meltdown-a-fin/, 25 January 2010.

[19] M. Tatham, "How to qualify for new credit with no credit score." https://www.experian.com/blogs/ask-experian/how-to-qualify-for-new-credit-with-no-credit-score/, 18 December 2018.

[20] M. Heidari, J. H. J. Jones, and O. Uzuner, "Deep contextualized word embedding for text-based online user profiling to detect social bots on twitter," in *IEEE 2020 International Conference on Data Mining Workshops (ICDMW), ICDMW 2020*, 2020.

[21] E. Hermes, "How to assess the creditworthiness of a customer," 14 February 2021.

[22] M. Heidari and J. H. Jones, "Using bert to extract topic-independent sentiment features for social media bot detection," in *2020 11th IEEE Annual Ubiquitous Computing, Electronics Mobile Communication Conference (UEMCON)*, pp. 0542–0547, 2020.

[23] A. Gatzioura, J. Vinagre, A. M. Jorge, and M. Sànchez-Marrè, "A hybrid recommender system for improving automatic playlist continuation," *IEEE Transactions on Knowledge and Data Engineering*, vol. 33, no. 5, pp. 1819–1830, 2021.

[24] Z. Liao, Y. Song, Y. Huang, L.-w. He, and Q. He, "Task trail: An effective segmentation of user search behavior," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 12, pp. 3090–3102, 2014.

[25] M. Heidari, J. H. J. Jones, and O. Uzuner, "An empirical study of machine learning algorithms for social media bot detection," in *IEEE 2021 International IOT, Electronics and Mechatronics Conference, IEMTRONICS 2021*, 2021.

[26] C.-Y. Chi, Y.-S. Wu, W.-r. Chu, D. C. Wu, J. Y.-j. Hsu, and R. T.-H. Tsai, "The power of words: Enhancing music mood estimation with textual input of lyrics," in *2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*, pp. 1–6, 2009.

[27] A. Gatzioura, J. Vinagre, A. M. Jorge, and M. Sànchez-Marrè, "A hybrid recommender system for improving automatic playlist continuation," *IEEE Transactions on Knowledge and Data Engineering*, vol. 33, no. 5, pp. 1819–1830, 2021.

[28] S. Zad, M. Heidari, J. H. J. Jones, and O. Uzuner, "Emotion detection of textual data: An interdisciplinary survey," in *IEEE 2021 World AI IoT Congress,AIIoT2021*, 2021.

[29] H. Yang, C. He, H. Zhu, and W. Song, "Prediction of slant path rain attenuation based on artificial neural network," in *2000 IEEE International Symposium on Circuits and Systems (ISCAS)*, vol. 1, pp. 152–155 vol.1, 2000.

[30] S. Zad, M. Heidari, J. H. J. Jones, and O. Uzuner, "A survey on concept-level sentiment analysis techniques of textual data," in *IEEE 2021 World AI IoT Congress,AIIoT2021*, 2021.

[31] M. Heidari, J. H. J. Jones, and O. Uzuner, "Offensive behaviour detection on social media platforms by using natural language processing models," 2021.

[32] "New research shows that your looks, creditworthiness may go hand in hand." https://phys.org/news/2009-03-creditworthiness.html, 12 March 2009.

[33] A. Klein, "Reducing bias in ai-based financial services." https://www.brookings.edu/research/reducing-bias-in-ai-based-financial-services/, 10 July 2020.

[34] bigml.com, “Loan risk data,” 14 February 2021.
[35] S. Gupta, “Lending club loan dataset 2007_2011.” https://www.kaggle.com/imsparsh/lending-club-loan-dataset-2007-2011?select=loan.csv, 29 May 2020.
[36] bigml.com, “Loanstats’ dataset,” 2021.
[37] A. Smith, “7 fundamental steps to complete a data analytics project.” https://blog.dataiku.com/2019/07/04/fundamental-steps-data-project-success, 4 July 2019.
[38] P. Meena, “Step-by-step guide to exploratory data analysis in python.” https://towardsdatascience.com/an-introduction-to-exploratory-data-analysis-in-python-9a76f04628b8, 14 August 2020.
[39] P. Crosman, “Can ai be programmed to make fair lending decisions?.” https://www.americanbanker.com/news/can-ai-be-programmed-to-make-fair-lending-decisions, 27 September 2016.
[40] A. Dobrin D.S.W., “It’s not fair! but what is fairness?.” https://www.psychologytoday.com/us/blog/am-i-right/201205/its-not-fair-what-is-fairness, 11 May 2012.
[41] R. Chawla, “How ai supports financial institutions for deciding creditworthiness.” https://www.entrepreneur.com/article/310262, 11 March 2018.
[42] The New York State Senate, “Unlawful discriminatory practices in relation to credit.” https://www.nysenate.gov/legislation/laws/EXC/296-A, 2021.