Credit Scores using machine learning techniques

Nadine Charmel May (nadine@aims.ac.za)
African Institute for Mathematical Sciences (AIMS)
Supervised by: Yabebal Fantaye, AIMS South Africa
Co-supervised by: Thuso Simon, Zoona Mobile Transactions

12 October 2018
Submitted in partial fulfillment of a structured masters degree at AIMS South Africa

Abstract

The new age of computers has brought forth the extensive capabilities of machine learning. In this essay we use some of this capabilities to obtain an improved credit scoring model. There has already been an extensive work on credit scoring for traditional banks, but credit scoring for Fintech data is an emerging area. Fintechs lack of consumer descriptive information makes application scoring particularly difficult. In this essay we investigate a Machine Learning approach to credit scoring on Fintech data. In particular, we use the Random forest algorithm for a binary classification of credits. We use a confusion matrix and an area under the receiver operator characteristic curve as our metric to measure the performance of our model.

Declaration

I, the undersigned, hereby declare that the work contained in this research project is my original work, and that any work done by others or by myself previously has been acknowledged and referenced accordingly.

Nadine Charmel May, 12 October 2018

Contents

Acknowledgements

I would like to thank my supervisors Yabebal Fantaye and Thuso Simon, this project has been so much fun. I had a wonderful time working with both of you. I am grateful for the guidance, support and kind words of encouragement in times of need. You guys are the best mentors anyone could ask for. I am appreciative for the time my tutor Josephine Tetteh has taken to go through this thesis with me in making corrections.

To Zoona Mobile Transactions for the internship presented to me and giving me the opportunity to work with this data I am very thankful and grateful. I am also thankful to AIMS South Africa directors, Jeff Sanders and Barry Green for dealing with my special requests due to the NDA of this work.

My family who has offered me abundant support throughout my years of studies especially my father for being my helping hand and thesis writing guide. The Friendships developed in my honors years has become stronger here at AIMS, you guys have become my family, the value you add to my life is inconceivable.

References

Adolphson, A., Sperber, S., and Tretkoff, M., editors. *p-adic Methods in Number Theory and Algebraic Geometry*. Number 133 in Contemporary Mathematics. American Mathematical Society, Providence, RI, 1992.

Aitkin, M. A., Francis, B., Hinde, J., and Darnell, R. *Statistical modelling in R*. Oxford University Press Oxford, 2009.

Baesens, B., Gestel, T. V., Viaene, S., Stepanova, M., Suykens, J., and Vanthienen, J. Benchmarking state-of-the-art classification algorithms for credit scoring. *Journal of the Operational Research Society*, 54(6):627–635, 2003. doi: 10.1057/palgrave.jors.2601545. URL https://doi.org/10.1057/palgrave.jors.2601545.

Beardon, A. From problem solving to research, 2006. Unpublished manuscript.

Bhatia, S., Sharma, P., Burman, R., Hazari, S., and Hande, R. Credit scoring using machine learning techniques. *International Journal of Computer Applications*, 161(11):1–4, Mar 2017. ISSN 0975-8887. doi: 10.5120/ijca2017912893. URL http://www.ijcaonline.org/archives/volume161/number11/27189-2017912893.

Bidgoli, H. Encyclopedia of information systems, four-volume set, 07 2002.

CIA,world fact book. Zambia. Central Intelligence Agency, https://www.cia.gov/library/publications/resources/the-world-factbook/geos/za.html, Accessed:2018-09-20.

Dabla-Norris, E., Ji, Y., Townsend, R. M., and Unsal, D. F. Distinguishing constraints on financial inclusion and their impact on gdp, tfp, and inequality. Working Paper 20821, National Bureau of Economic Research, January 2015. URL http://www.nber.org/papers/w20821.

Finlay, S. Multiple classifier architectures and their application to credit risk assessment. 210:368–378, 04 2011.

Keown, A. J. *Personal Finance Turning Money into Wealth*. Pearson, 2013.

Lamport, L. *LaTeX: A Document Preparation System*. Addison-Wesley, 1986.

MacKay, D. Statistical testing of high precision digitisers. Technical Report 3971, Royal Signals and Radar Establishment, Malvern, Worcester. WR14 3PS, 1986a.

MacKay, D. A free energy minimization framework for inference problems in modulo 2 arithmetic. In Preneel, B., editor, *Fast Software Encryption (Proceedings of 1994 K.U. Leuven Workshop on Cryptographic Algorithms)*, number 1008 in Lecture Notes in Computer Science Series, pages 179–195. Springer, 1995b.

Manyika, J., Lund, S., and Marc Singer andOlivia White, C. B. Digital finance for all: Powering inclusive growth in emerging economies. Technical report, McKinsey Global Institute (MGI), 2016.

Mashanovich, N. Credit scoring: The development process from end to end, Sep 2017. URL https://www.worldprogramming.com/blog/credit_scoring_development_pt1.

Radcliffe, B. How education and training affect the economy, 2018. URL http://www.educationviews.org/education-training-affect-economy/.

Rahmatian, S. Transaction processing sytems.

SADC.

Sammut, C. and Webb, G. I. *Encylcypedia of Machine Learning and Data mining*. Springer, Boston, MA, 2011.

Shannon, C. A mathematical theory of communication. *Bell Sys. Tech. J.*, 27:379–423, 623–656, 1948.

Shannon, C. The best detection of pulses. In Sloane, N. J. A. and Wyner, A. D., editors, *Collected Papers of Claude Shannon*, pages 148–150. IEEE Press, New York, 1993.

stein, P., Susan Dorasil, A. K., and Gamster, M. small and medium enterprise finance: New findings, trends and g-20 /global partnership for financial inclusion progress. Technical report, Interntionl finance corporation, 2013.

Tia Linda Zuze, J. H. Age,gender and financial inclusion in zambia. Technical report, Financial Sector Deepening Zambia, 2018.