**Literature Review**

**Determinants**  
According to the credit risk modeling experts of SAS, elements of a good credit risk model cover speed, precision, confidence (Akhadov, Rogers & Filipenkov, 2018). Speed of credit decisions matter as competition means a lag of several minutes means a world of difference. Precision determines the extent of accurate credit decisions that maximize revenue and minimize defaults. A good measure of confidence is the balance between risk aversion and business development in a model to allow more creditworthy customers as well as ensuring validity of models through transparency and rigor in credit scoring.

**Traditional vs Modern Models**Traditional models often assume parameterization based on applicants having identical patterns or behaviors whereby this is not always the case, especially considering arbitrary classifications when taking into account what may be considered ‘good’ or ‘bad’ standing (Banasik, Crook & Thomas, 2003). Modern models are moving away from processes that allow for human biases in the manual processes and are incorporating as much alternative information and putting new algorithms and segmented models to the test. A study on dynamic credit risk modeling by (Moradi & Mokhatab Rafiei, 2019) to include generalized additive models (GAM) for classification and classification under supervised training and adaptive network-based fuzzy inference system (ANFIS) to adapt to input data and minimize error based on gradient descent training principle and fuzzy logic.

|  |  |
| --- | --- |
| Traditional Models | Modern Models |
| Tools consist of Logistic Regression, Naïve Bayes, etc. | Tools consist of Neural Networks, Recursive Partitioning Models, Radial-Basis Functions, etc. |
| Heavily based on historical data/samples, Boolean logic | Involves use of out-of-sample forecasts, makes use of consumer behavior data, uses fuzzy inference systems, etc. |
| Based on static loss assumptions | Use of segmented markets |
| Focused on systematic risk components | Integrated platform use |
| Open to human bias due to manual processes | Expansive use of automated processes |
| Manual processes bring heavy cost, slow decisions | Adaptive with anomaly detection and predictive analytics uncovering new forms of risk at any given time |

*Table 1: Characteristics of Current vs Modern Evolving Tools*

**Default Recovery Rates in Credit Risk**Traditional credit risk modeling assumes collateral values and recovery rates as a constant parameter or as a stochastic variable independent from probability of default. The authors Altman, Resti & Sironi (2004) recognize that three main variables affect the credit risk of a financial asset, which are probability of default (PD), loss given default (LGD), and exposure at default (EAD). Probability of default, found by one minus recovery rate, is typically given significant attention while much less so has been given to loss given default and its relationship with probability of default.

**Sample Selection Bias**In selecting a model as well, sample selection bias is still evident in traditional models. This ties in with how much less of a focus is given to those approved for loans rather than those who failed to receive one (Ditrich, 2015). While a focus on reject inference has been attempted by several studies, techniques such as augmentation and extrapolation remain scrutinized by academics despite some results improving accuracy and discriminative power of models.

**Variable Importance (VI)**Variable importance allows for a quantitative ranking of how each predictor contributes to the model but at times heuristics have to be applied to include specific variables despite poor results. Feature engineering can be conducted to improve predictor existence.

**Multicollinearity**Use of a computational score variance inflation factor (VIF) allows for a measure of inflation of variance of regression coefficients due to multicollinearity in the model (Yoo et al., 2014). This serves as an independence test across variables. By removal of factors with VIF values deemed high, at approximate cut-off point of ten, it allows for simplification of model through the reduction of redundant data.

**Random Forests**Random Forests is one of the techniques we picked to try as a bagging classifier to allow better aggregation (Grennepois, Alvirescu & Bombail, 2018). It enables us to limit overfitting, presents high accuracy, provide easy choice of variables and is stable. Its drawbacks of parameter choice and lower interpretability were less factors to consider.

**Measures & Optimal Cut-Off**Determinants found to be of use in optimal decision threshold of models are sensitivity and specificity, of which the maximum sum of determines the best point to minimize overall error of modelling credit risk. These two factors are determined by the possible outcomes consisting of true positives, true negatives, false positives, and false negatives. Kolmogorov-Smirnov tests (Rezac & Rezac, 2011) provide a visual for credit scorecard models. References

Akhadov, A., Rogers, D., & Filipenkov, N. (2018). 6 Keys to Credit Risk Modeling for the Digital Age. Retrieved 14 October 2019, from https://www.sas.com/content/dam/SAS/en\_us/doc/whitepaper1/credit-risk-modeling-digital-age-109772.pdf

Altman, E., Resti, A., & Sironi, A. (2004). Default Recovery Rates in Credit Risk Modelling: A Review of the Literature and Empirical Evidence. *Economic Notes*, *33*(2), 183-208. doi: 10.1111/j.0391-5026.2004.00129.x

Banasik, J., Crook, J., & Thomas, L. (2003). Sample selection bias in credit scoring models. *Journal Of The Operational Research Society*, *54*(8), 822-832. doi: 10.1057/palgrave.jors.2601578

Ditrich, J. (2015). SELECTION BIAS REDUCTION IN CREDIT SCORING MODELS. In *The 9th International Days of Statistics and Economics*. Prague, Czech Republic: University of Economics, Prague. Retrieved from <https://pdfs.semanticscholar.org/0e5d/1b35645625e166764565ed5018057f73eefb.pdf>

Grennepois, N., Alvirescu, A., & Bombail, M. (2018). Point of View: Using Random Forest for credit risk models. Retrieved 14 October 2019, from https://www2.deloitte.com/content/dam/Deloitte/nl/Documents/financial-services/deloitte-nl-fsi-using-random-forest-for-credit-risk-models.pdf

Moradi, S., & Mokhatab Rafiei, F. (2019). A dynamic credit risk assessment model with data mining techniques: evidence from Iranian banks. *Financial Innovation*, *5*(1). doi: 10.1186/s40854-019-0121-9

Rezac, M., & Rezac, F. (2011). How to Measure the Quality of Credit Scoring Models.

Yoo, W., Mayberry, R., Bae, S., Singh, K., Peter He, Q., & Lillard, J. W., Jr (2014). A Study of Effects of MultiCollinearity in the Multivariable Analysis. *International journal of applied science and technology*, *4*(5), 9–19.