Modernizing credit risk modelling:

Using machine learning to challenge industry norms:

Loss Given Default (LGD) Models

CS 890ES Winter 2020

Adam Kehler (200251114)

University of Regina

Outline

1. About the presenter
2. Problem Statement
3. Introduction & Background
4. Solution overview
5. Data, Modelling, & Analysis
6. Outcomes – conclusions & impacts
7. Next steps

University of Regina

1. About the presenter

- Adam Kehler
- PhD student – statistics (computational)
- Director, Portfolio Modelling & Data Science @ Farm Credit Canada (FCC)

2. Problem Statement

Use **modern data science** and **machine learning** techniques to benchmark, **challenge**, refine, and enhance the more **traditional statistical modelling** methods employed currently at my organization for processes related to certain **accounting standards** and **banking regulations**.

University of Regina

3. Introduction & Background

Constrained by standards & regulations:

IFRS®

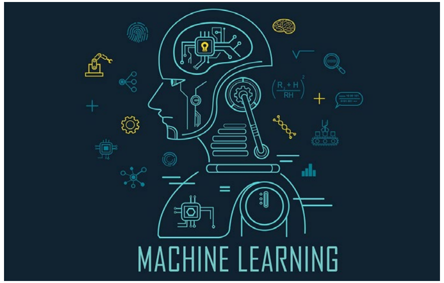
Basel Committee on Banking Supervision

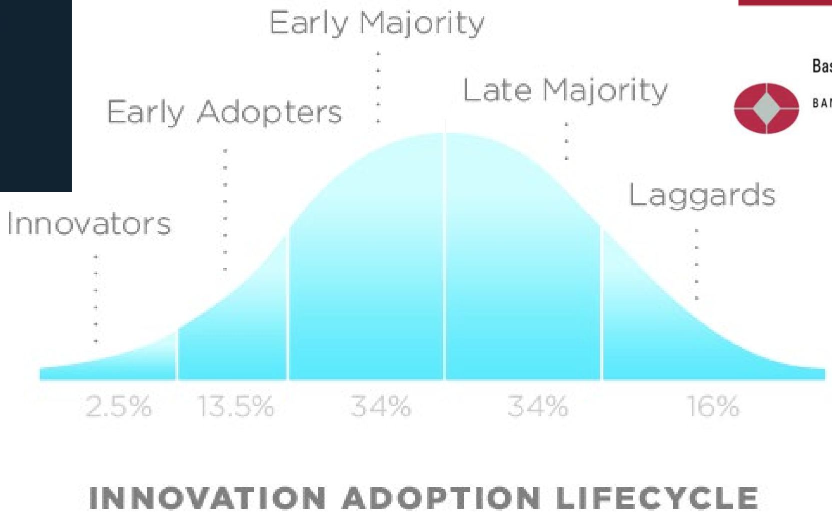
BANK FOR INTERNATIONAL SETTLEMENTS

OSFI
BSIF

3. Introduction & Background

Resulting in slower adoption:





Basel Committee on Banking Supervision
BANK FOR INTERNATIONAL SETTLEMENTS
IFRS®
OSFI
BSIF

University of Regina

3. Introduction & Background

Initial opportunity:

- Model Performance
- Model benchmarking
- “Challenger” models
- Alternative models

4. Solution Overview

Scope:

4. Solution Overview

Scope:

Allowance for Credit Losses (ACL) = Expect lifetime credit loss (ECL)

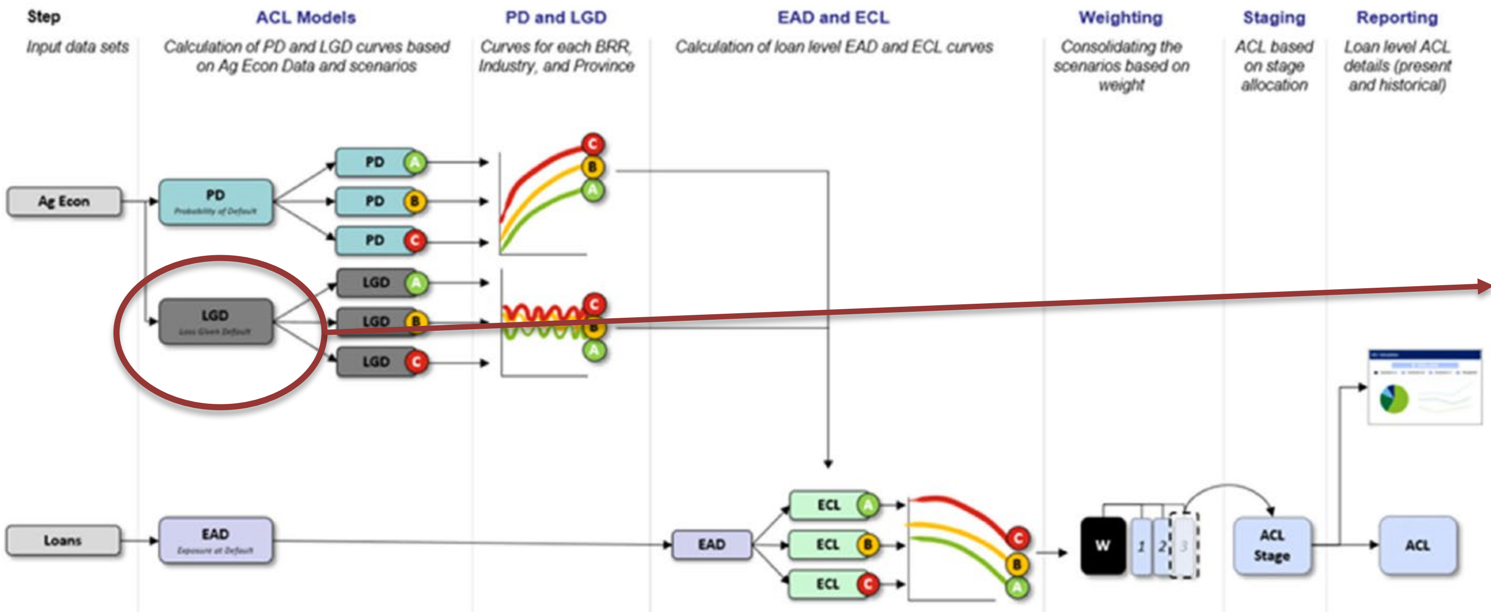
$$ACL = \sum \left[\frac{PD_i \times LGD_i \times EAD_i}{(1+r)^n} \right]$$

University of Regina

4. Solution Overview

$$ACL = \sum \left[\frac{PD_i \times LGD_i \times EAD_i}{(1+r)^n} \right]$$

Allowance for Credit Loss Calculation | Model



Loss Given Default (LGD) Model

4. Solution Overview

Current Approach

- Two-stage micro-structure
- Likelihood X Severity
- Assumes independence

$$E(LGD) = E(PWOD \times ELWO) = E(PWOD) \times E(ELWO)$$

- Scorecard approaches based on logistic regression

Proposed Approaches

1. Two-step models
    a) RF class with RF regress
    b) RF class with k-NN regress

2. Two-step models
    a) RF regression
    b) k-NN regression
    c) Multi-layer neural network
    d) Simulation-based

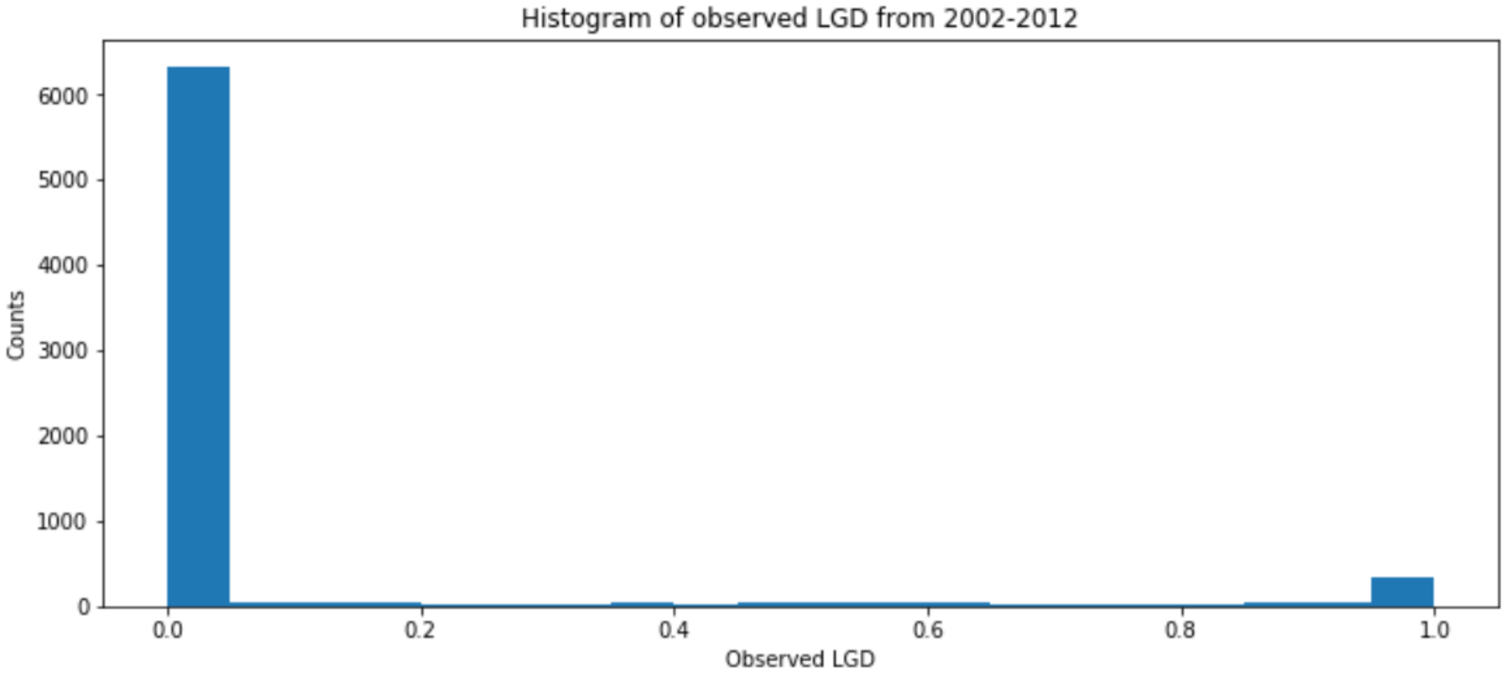
5. Data, Modelling, & Analysis

LGD Data:

- Observed loss data from 2002 to 2012
- 7,264 instances with 172 features (~1.25 M)
- Narrowed to 21 useful & quality features (~150 K)
- From SAP, Oracle DBs, Teradata DBs, and Excel (combined using SQL in SAS)
- Due to system access limitations, CSV into Python JN

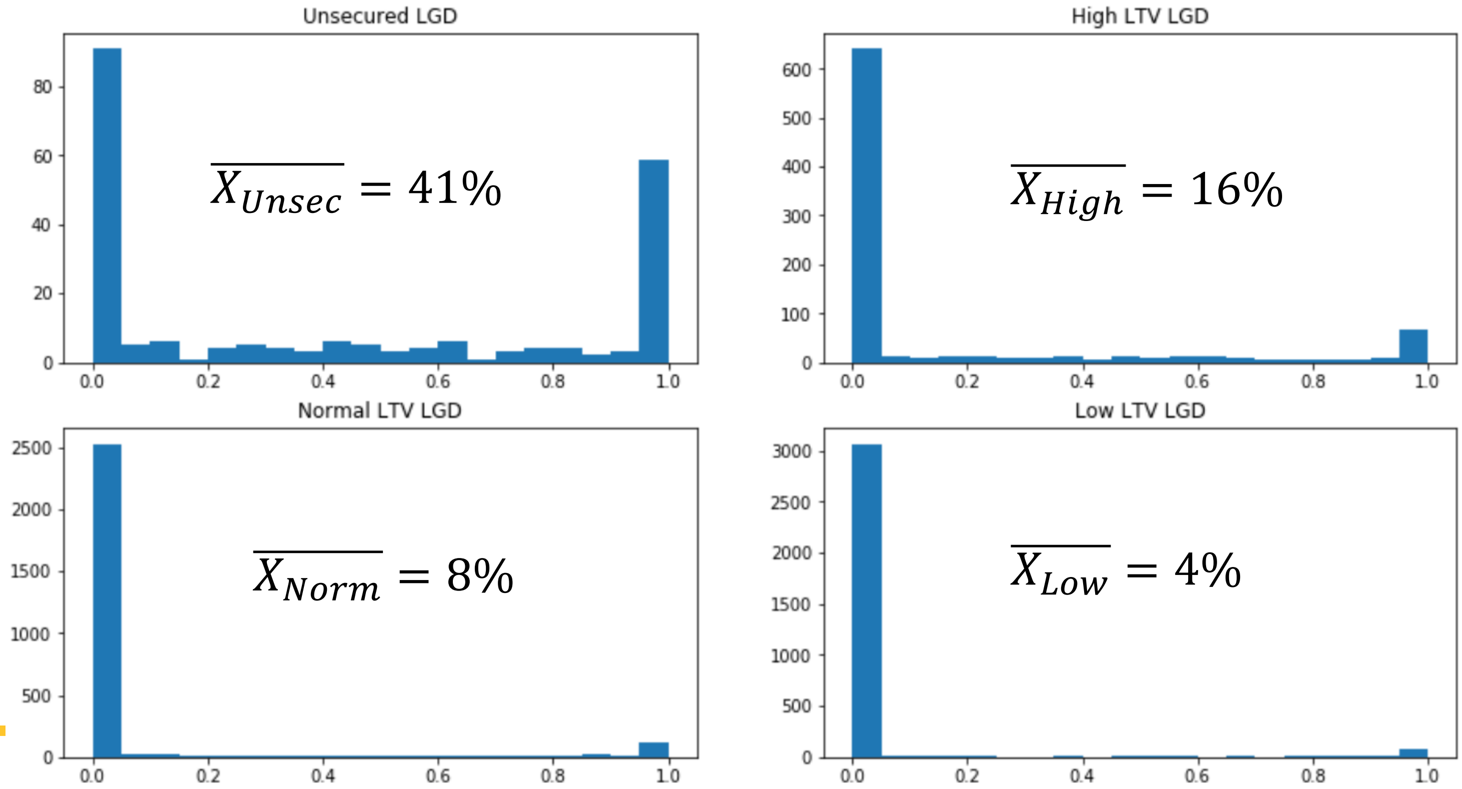
5. Data, Modelling, & Analysis

LGD Data:



5. Data, Modelling, & Analysis

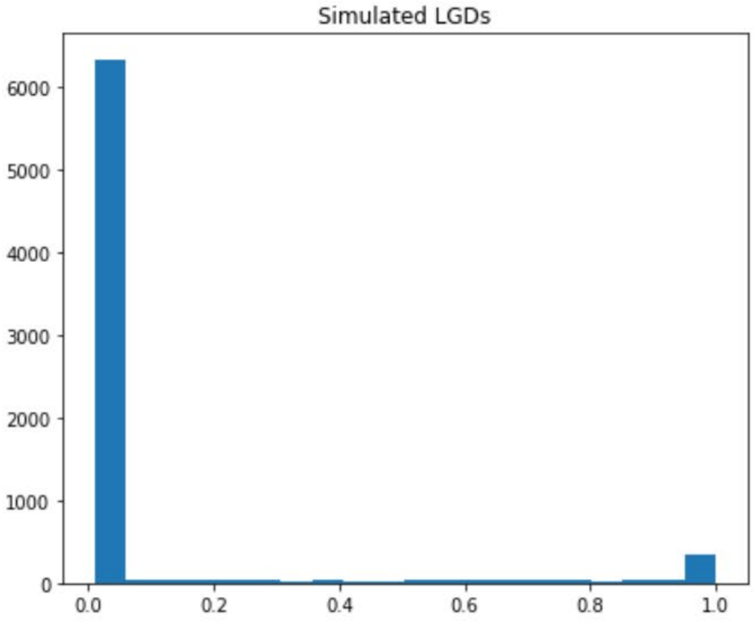
LGD Data:

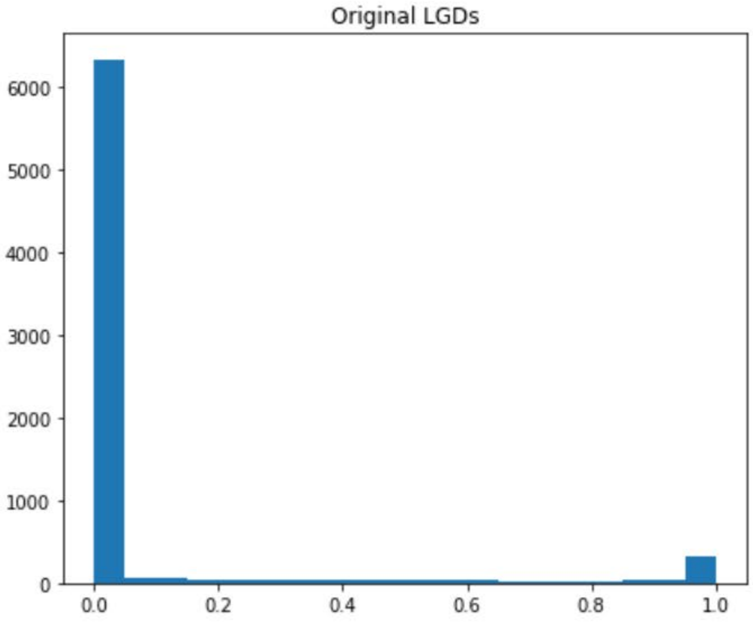


University of Regina

5. Data, Modelling, & Analysis

Random LGD Generator:

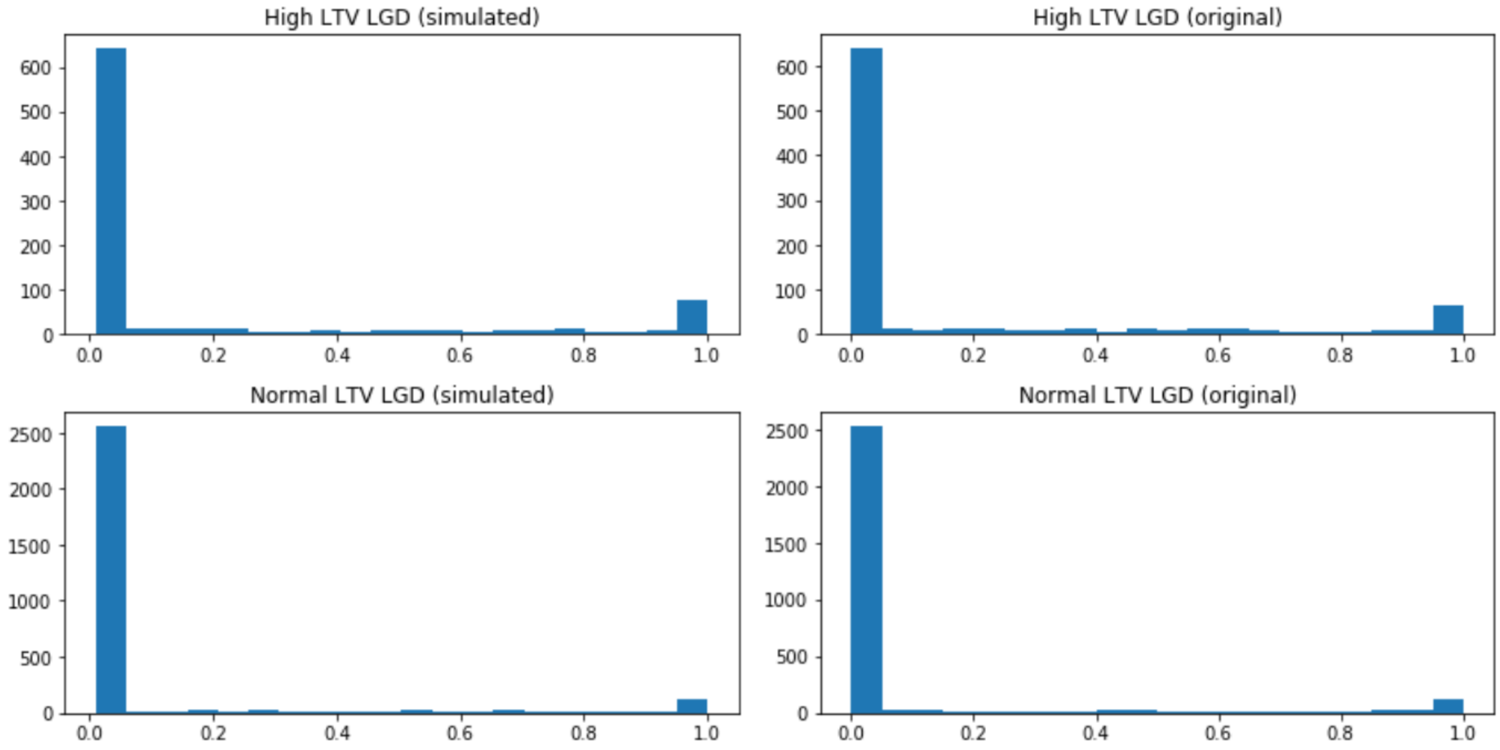




University of Regina

5. Data, Modelling, & Analysis

Random LGD Generator:



University of Regina

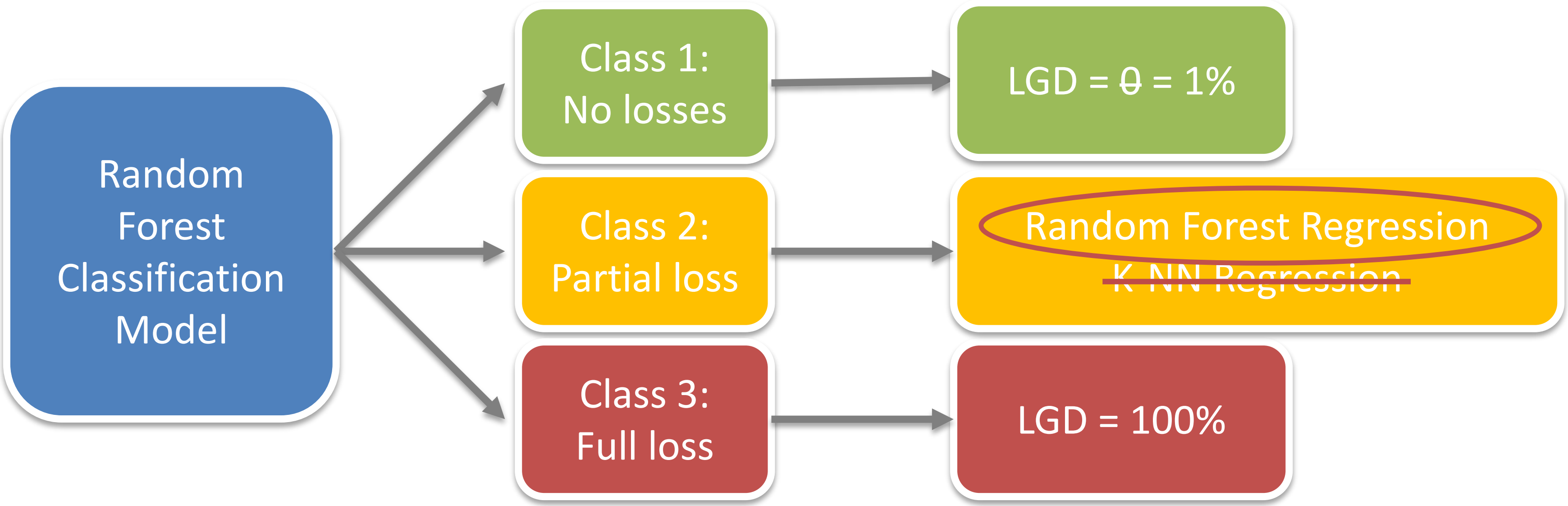
5. Data, Modelling, & Analysis

Predictive model data prep:

- One-hot encoding
- Missing values (set to 0)
- Bin LGDs into 3 groups (0 – no loss, 1 – partial loss, 2 – full loss)
- Split Train-Test (70-30)
- Scale features

5. Data, Modelling, & Analysis

Two-step models:



Classification then severity

5. Data, Modelling, & Analysis

Two-step models:

Random Forest Classification Model

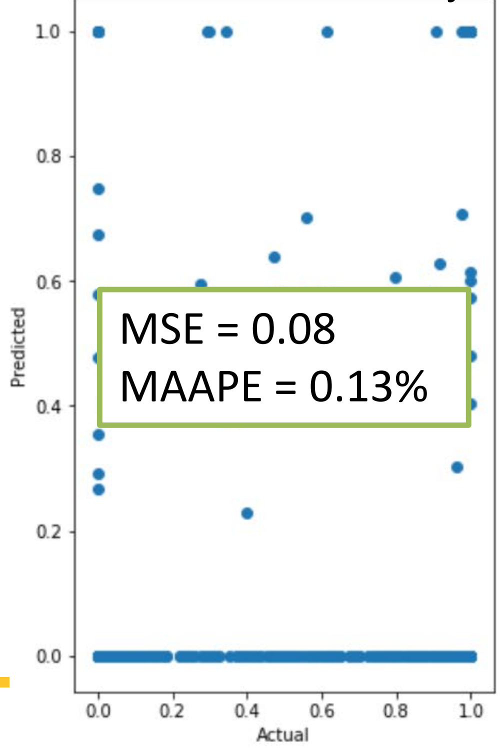
Probability	Severity
Class 1: No losses	LGD = ~~0~~ = 1%
Class 2: Partial loss	RF Regression 1% < LGD < 100%
Class 3: Full loss	LGD = 100%

Probability weighted

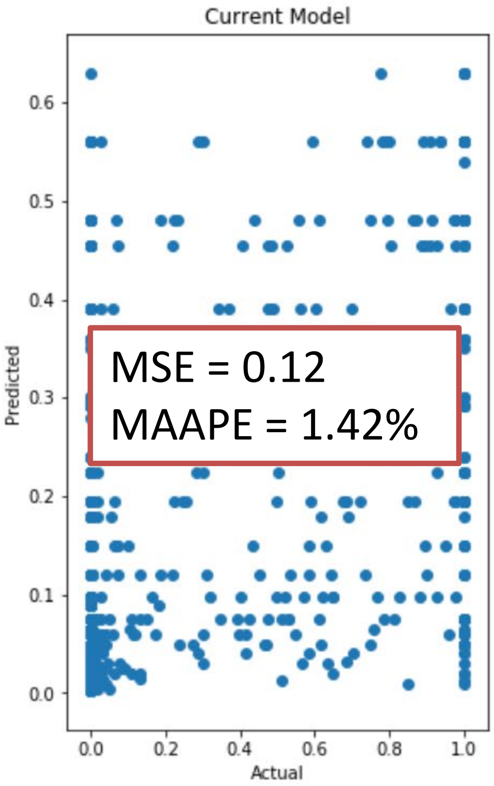
5. Data, Modelling, & Analysis

Two-step models:

Class then severity

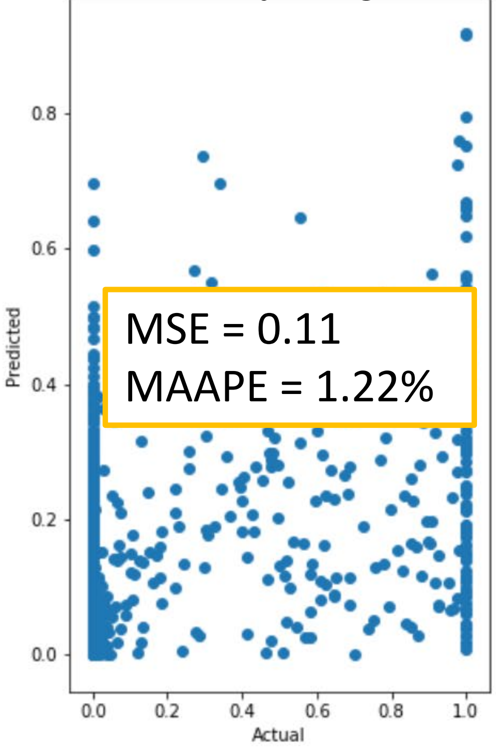


MSE = 0.08
MAAPE = 0.13%



MSE = 0.12
MAAPE = 1.42%

Probability weighted

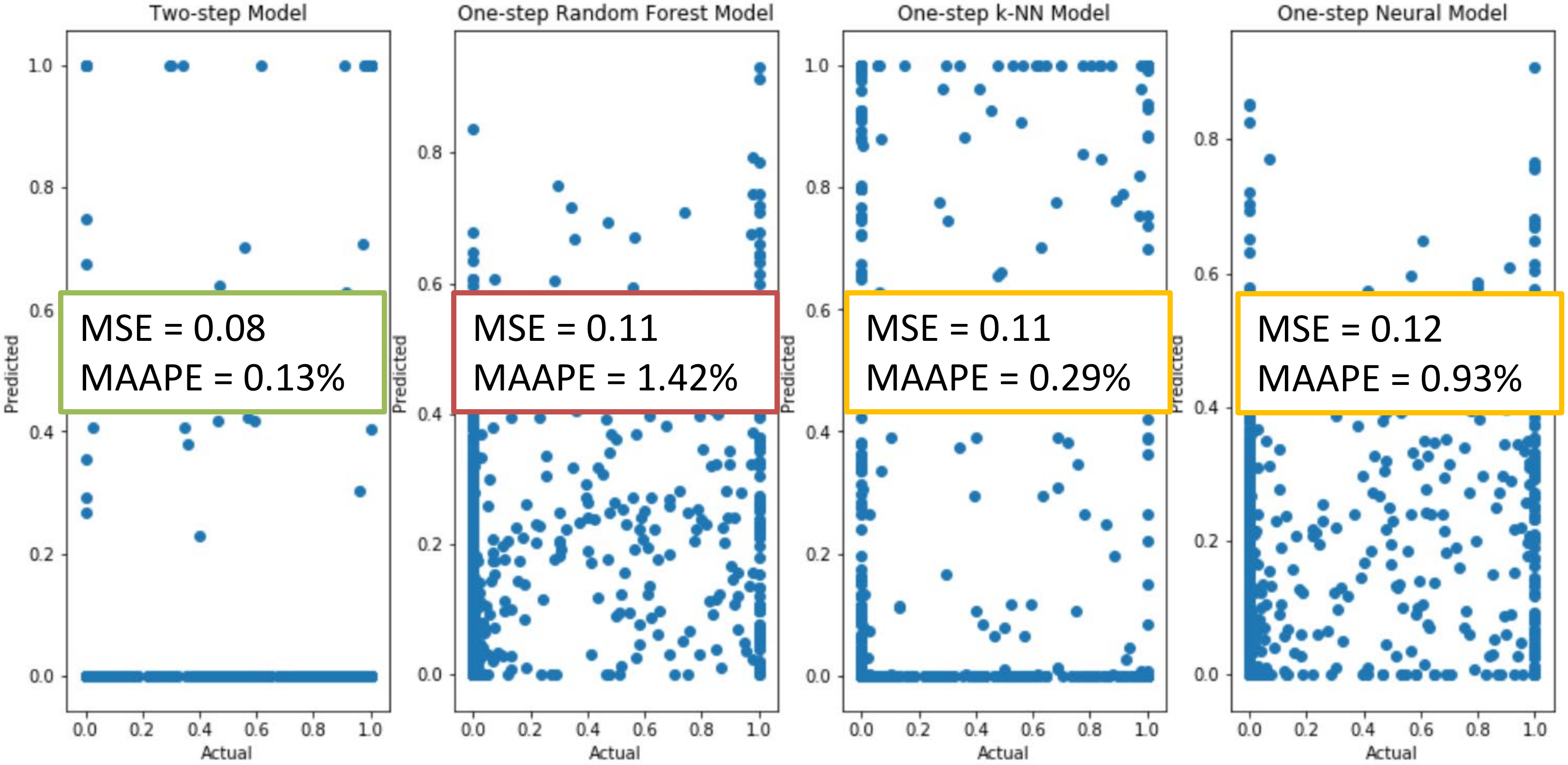


MSE = 0.11
MAAPE = 1.22%

University of Regina

5. Data, Modelling, & Analysis

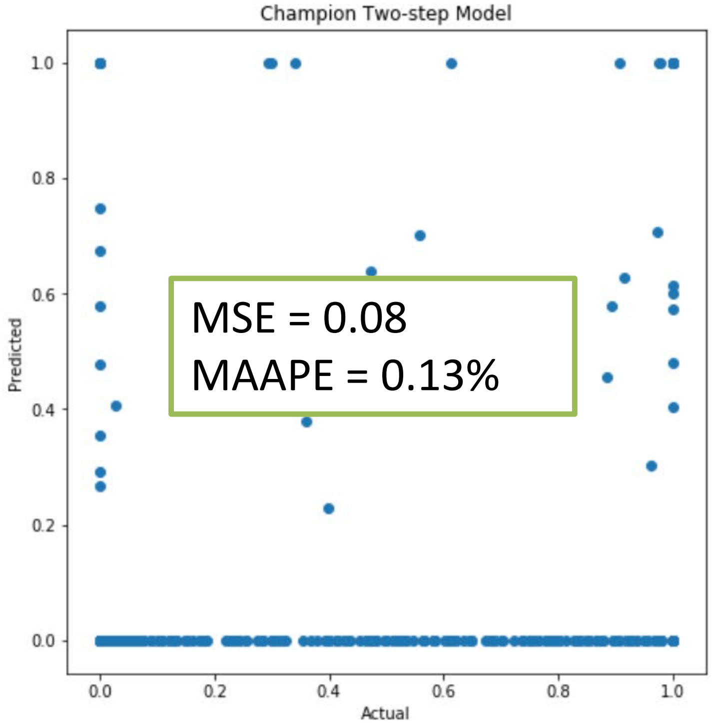
One-step models:



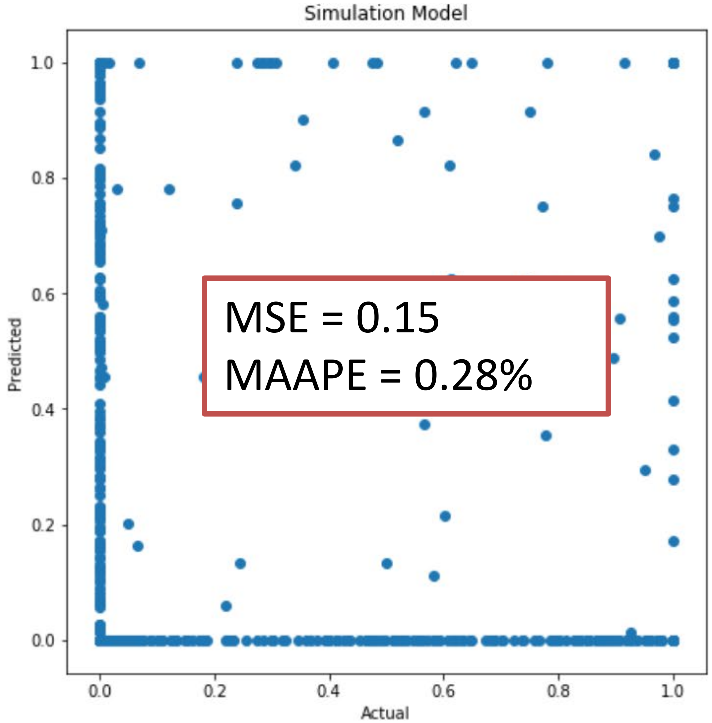
University of Regina

5. Data, Modelling, & Analysis

Simulation-based model:



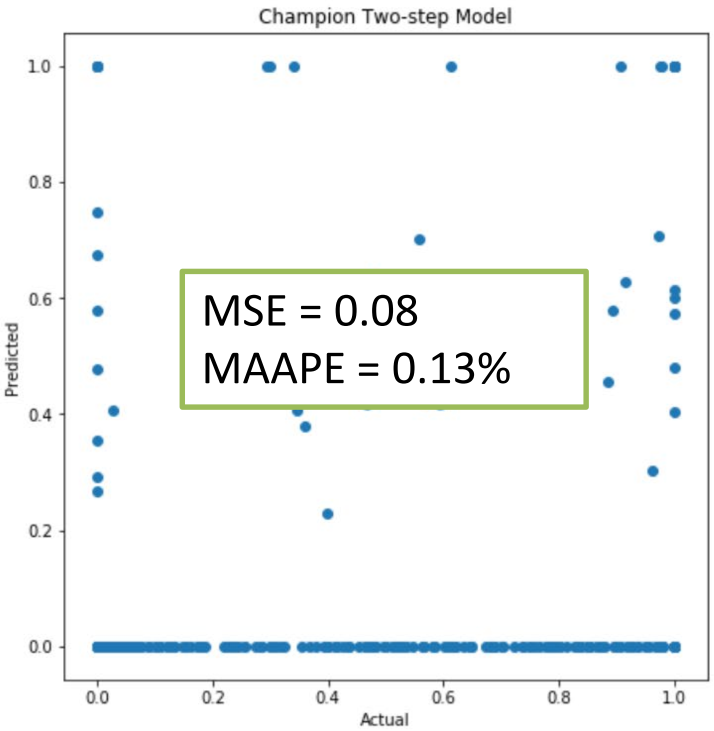
MSE = 0.08
MAAPE = 0.13%

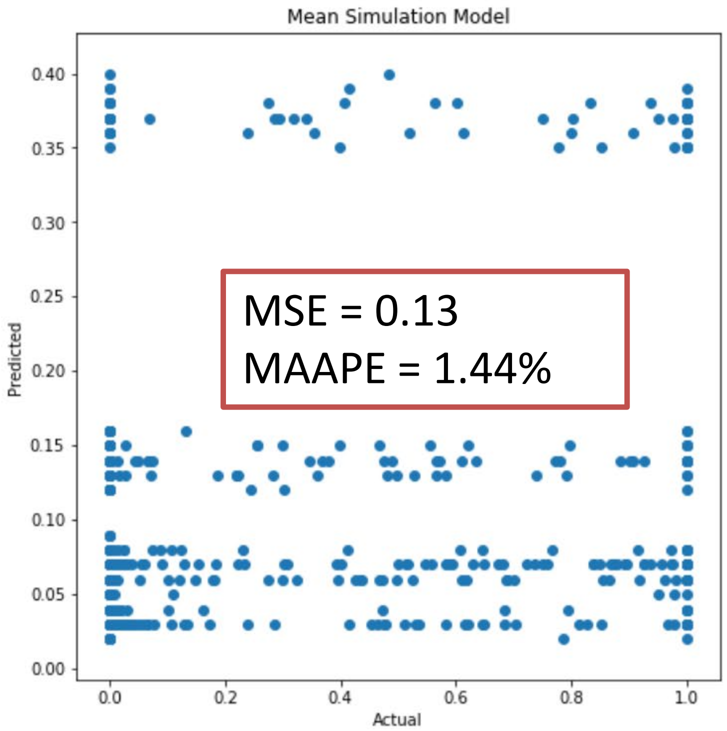


MSE = 0.15
MAAPE = 0.28%

5. Data, Modelling, & Analysis

Simulation-based model:

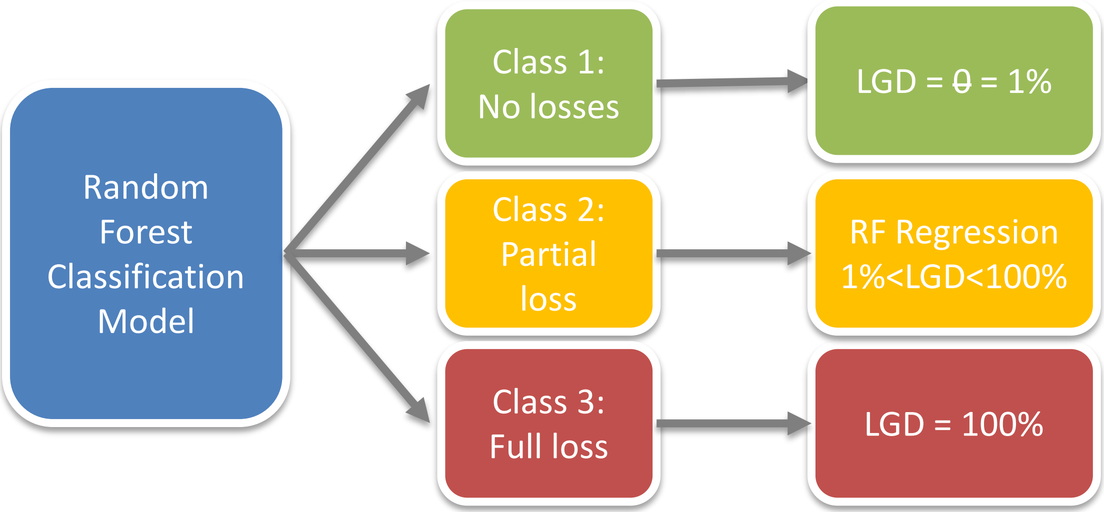


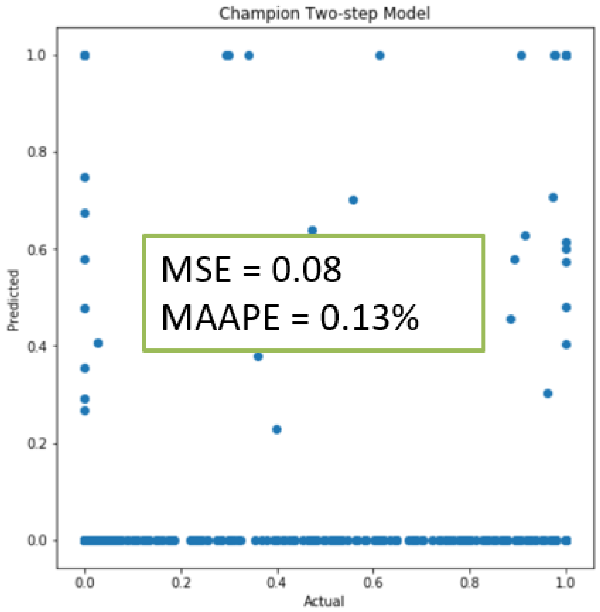


University of Regina

6. Outcomes – conclusions

Champion Model: RF Classification, then RF Severity





6. Outcomes – impacts

- Enables more accurate Financial Statements
- Enables more refined pricing
- Improved Model Performance Monitoring
- Facilitate change management and adoption of new techniques
- Inform next round of model development (this upcoming year)
- Simulation engine will be used in credit economic capital (ECAP) model

University of Regina

7. Next Steps

1. Partner with Model Development on LGD simulation (for credit ECAP)
2. Work with IT and Finance to get 2013 – 2019 data, then re-run analysis
3. Incorporate "Challenger" model into Model Performance Monitoring
4. Explore enhancements by bringing in macro-economic variables (forward-looking)
5. "Familiarize" stakeholders with new approaches (to help move Challenger to Production)
6. Work with IT and Finance to determine best practices for deployment