Project proposal:

Modernizing credit risk modelling

Using machine learning to challenge industry norms

February 3, 2020
CS 890ES Winter 2020
Adam Kehler (200251114)

University of Regina

Outline

1. About the presenter
2. Background
3. Problem statement
4. Solution overview
5. Data & tools
6. Timelines
7. Expected outcomes

University of Regina

About the presenter

- Adam Kehler
- PhD student – statistics (computational)
- Director, Portfolio Modelling & Data Science @ Farm Credit Canada (FCC)
- 3 favorite things right now:
  - New puppy
  - New wife
  - New acreage

Background – standards & regulations

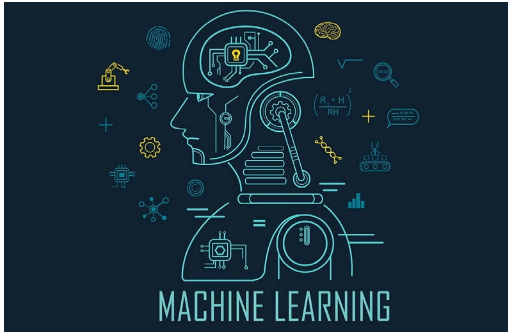
IFRS®

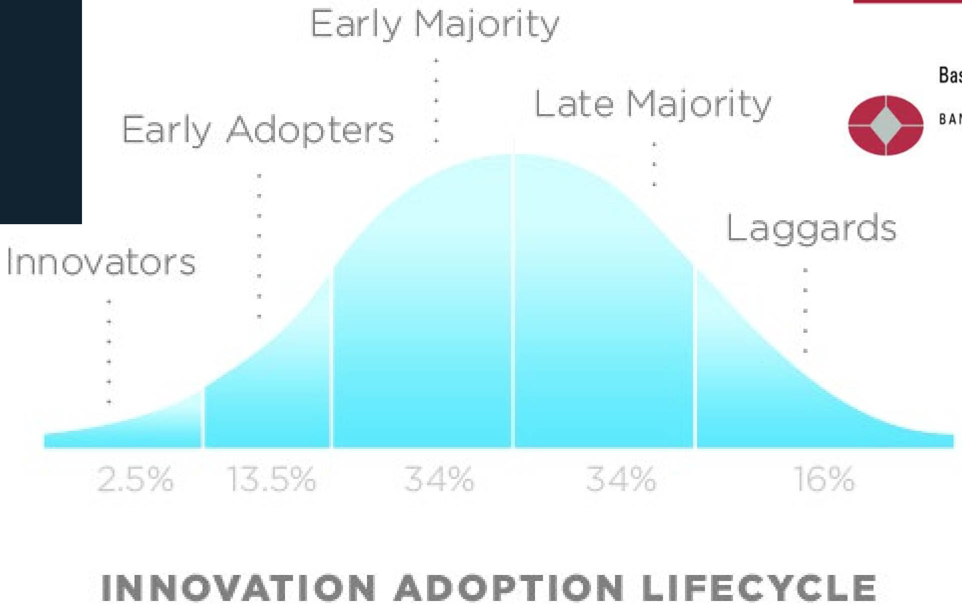
Basel Committee on Banking Supervision

BANK FOR INTERNATIONAL SETTLEMENTS

OSFI
BSIF

Background – slow adoption of ML





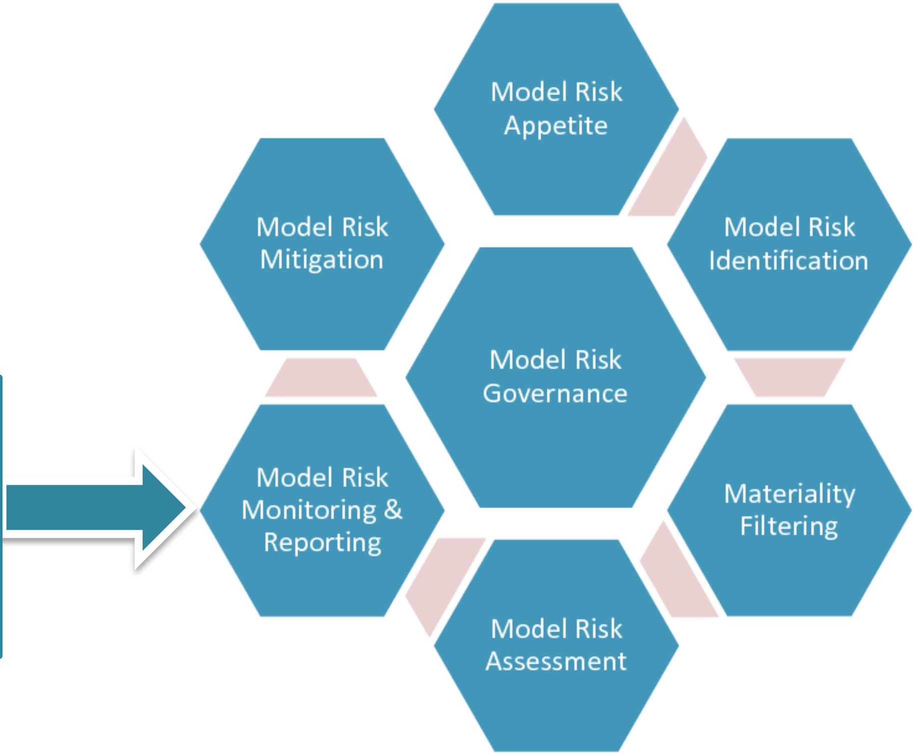
IFRS®

Basel Committee on Banking Supervision
BANK FOR INTERNATIONAL SETTLEMENTS

OSFI
BSIF

Background – Model risk management

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

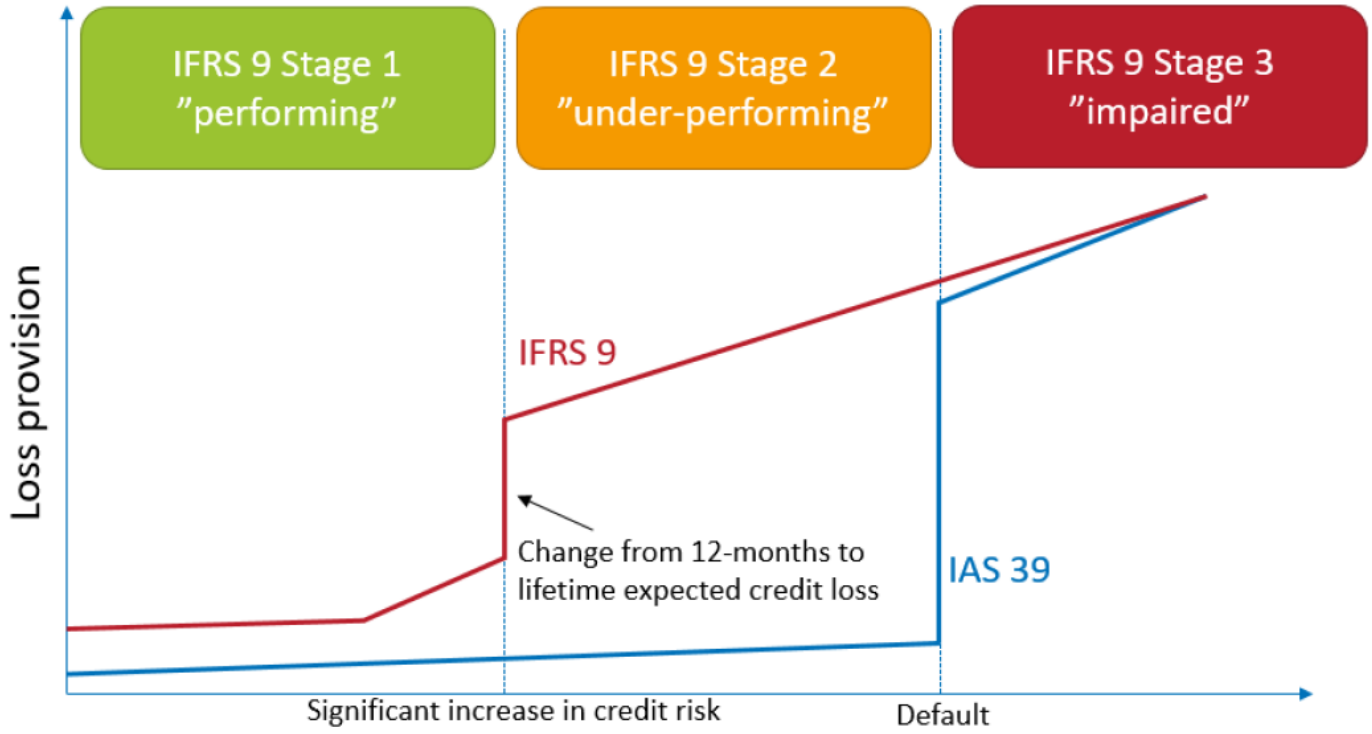


Problem Statement

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**.

Solution overview – scope

Solution overview – Allowance



Solution overview – Allowance

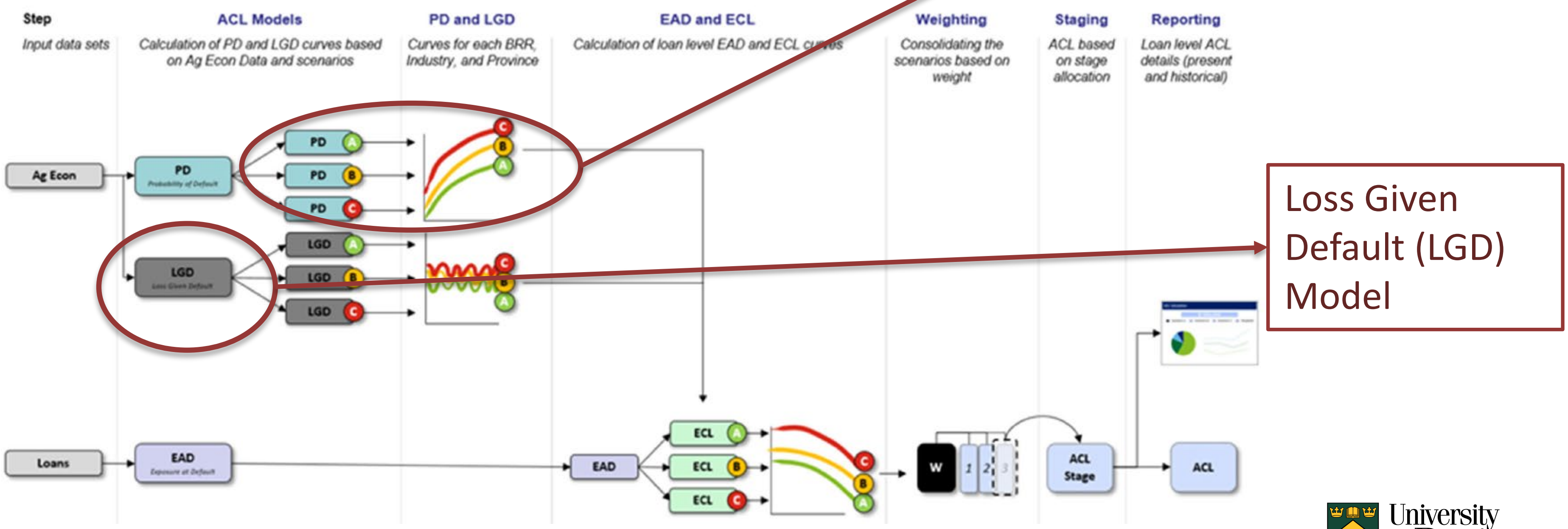
Allowance for Credit Losses (ACL) = Expect lifetime credit loss (ECL)

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

Solution overview – Allowance

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

Allowance for Credit Loss Calculation | Model

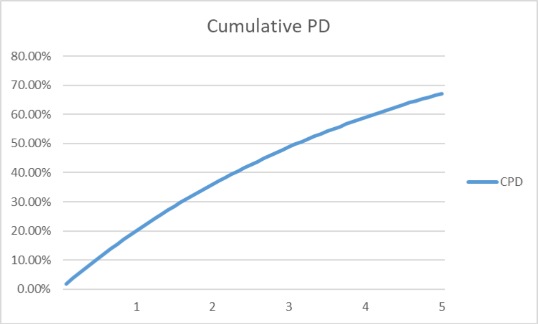


Solution overview – Lifetime PD

Current Approach

- "Constant Intensity Model"
- Based on an underlying Poisson Process

$$F(t) = 1 - e^{-\lambda t}$$
$$\lambda = -ln\big(1 - F(1)\big)$$



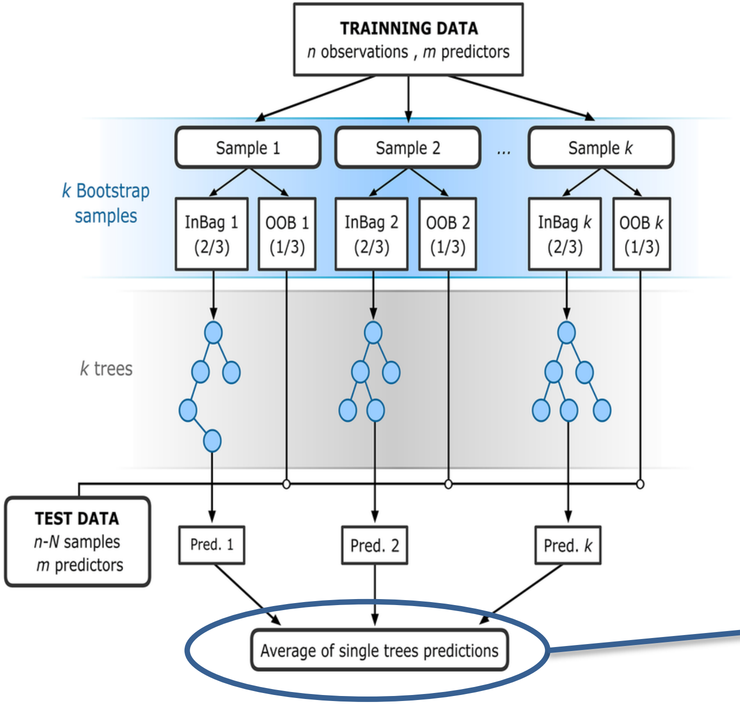
Proposed Approach

- "Random Survival Forest" – mix between Random Forest and Survival Analysis
- Use Random Forest algorithm with hazard rate to inform loss function being optimized

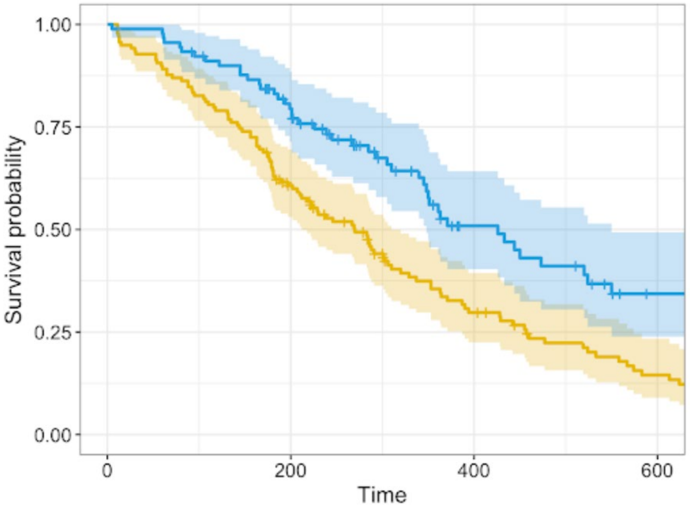
University of Regina

Solution overview – Lifetime PD

Random Forest Algorithm



Survival Analysis Theory



$$\Lambda(t|X_i) = \lambda_0(t)e^{\boldsymbol{X_i\beta}}$$

Solution overview – LGD Model

Current Approach

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

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

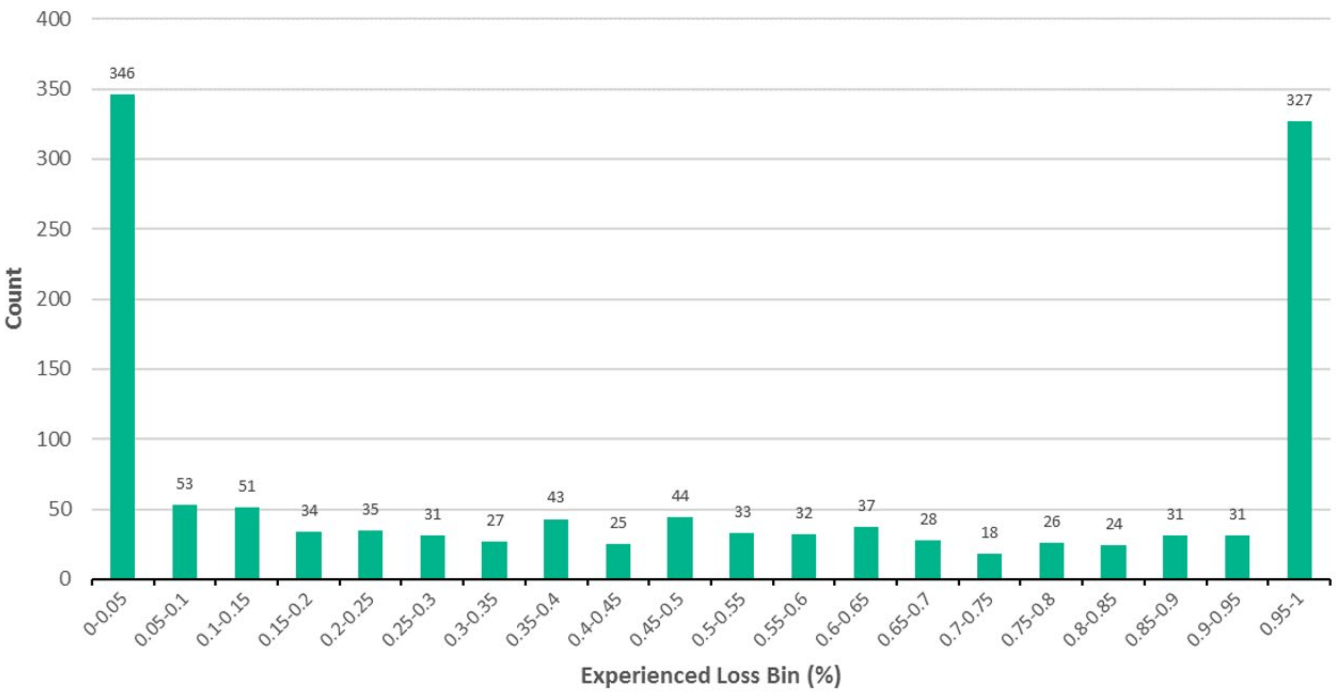
- Scorecard approaches based on logistic regression

Proposed Approach

a) Two-stage model
   - CART
   - SVM

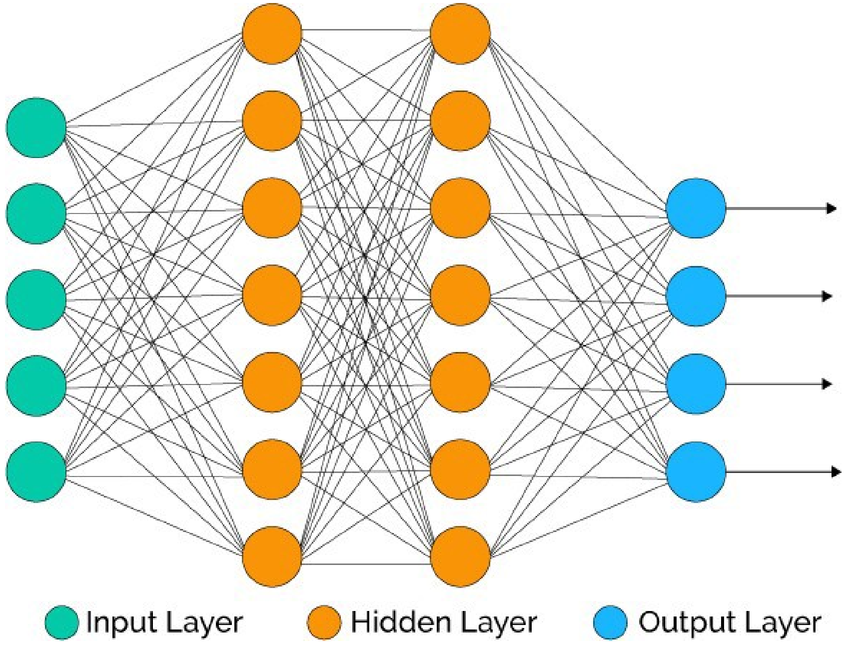
b) One-stage model
   - Random Forest (Forward-looking)
   - Deep neural networks

Solution overview – LGD Model

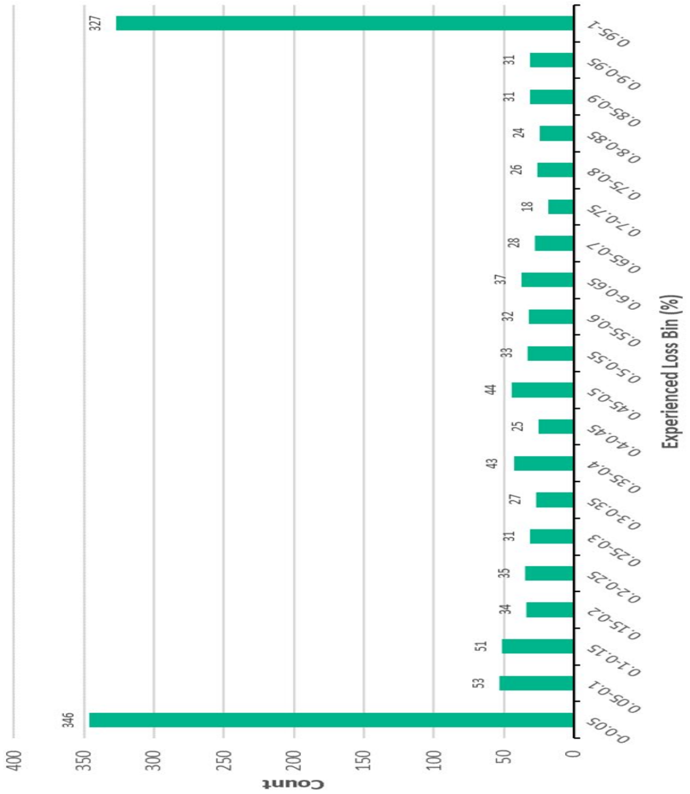


University of Regina

Solution overview – LGD Model

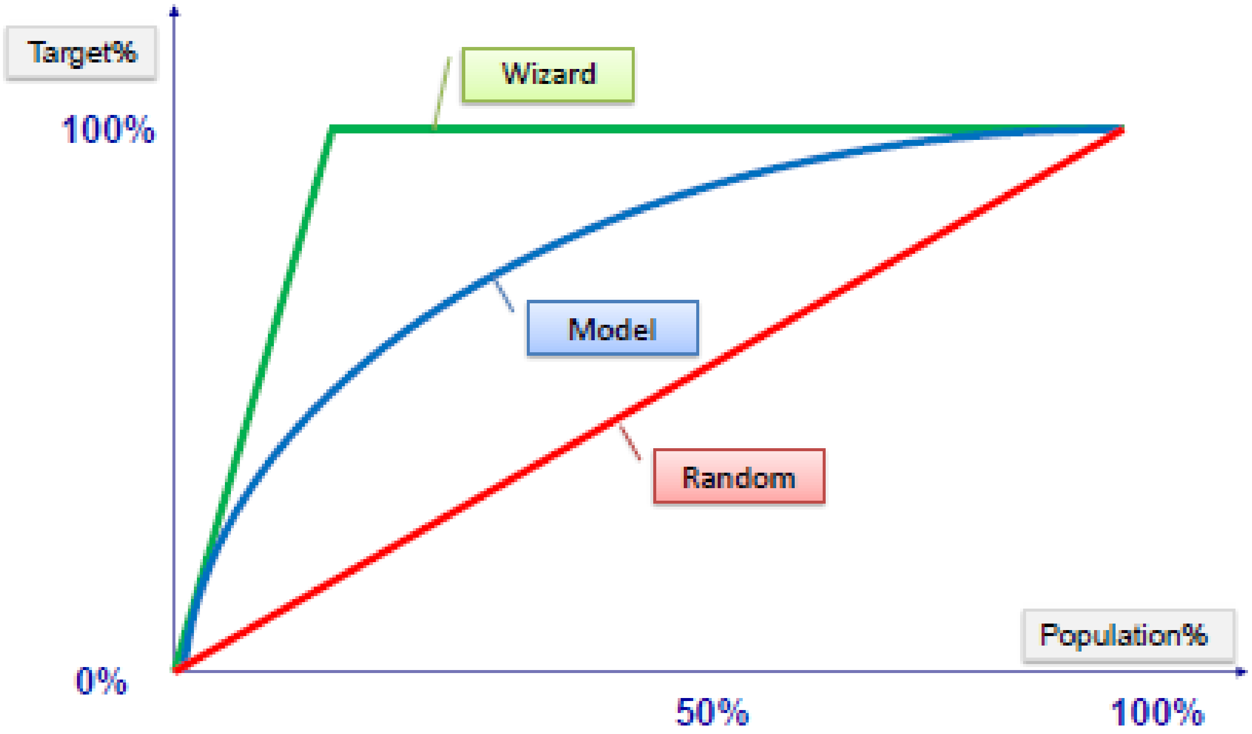


Deep Neural Network



University of Regina

Solution overview – Performance



Data & tools

jupyter

python

NumPy

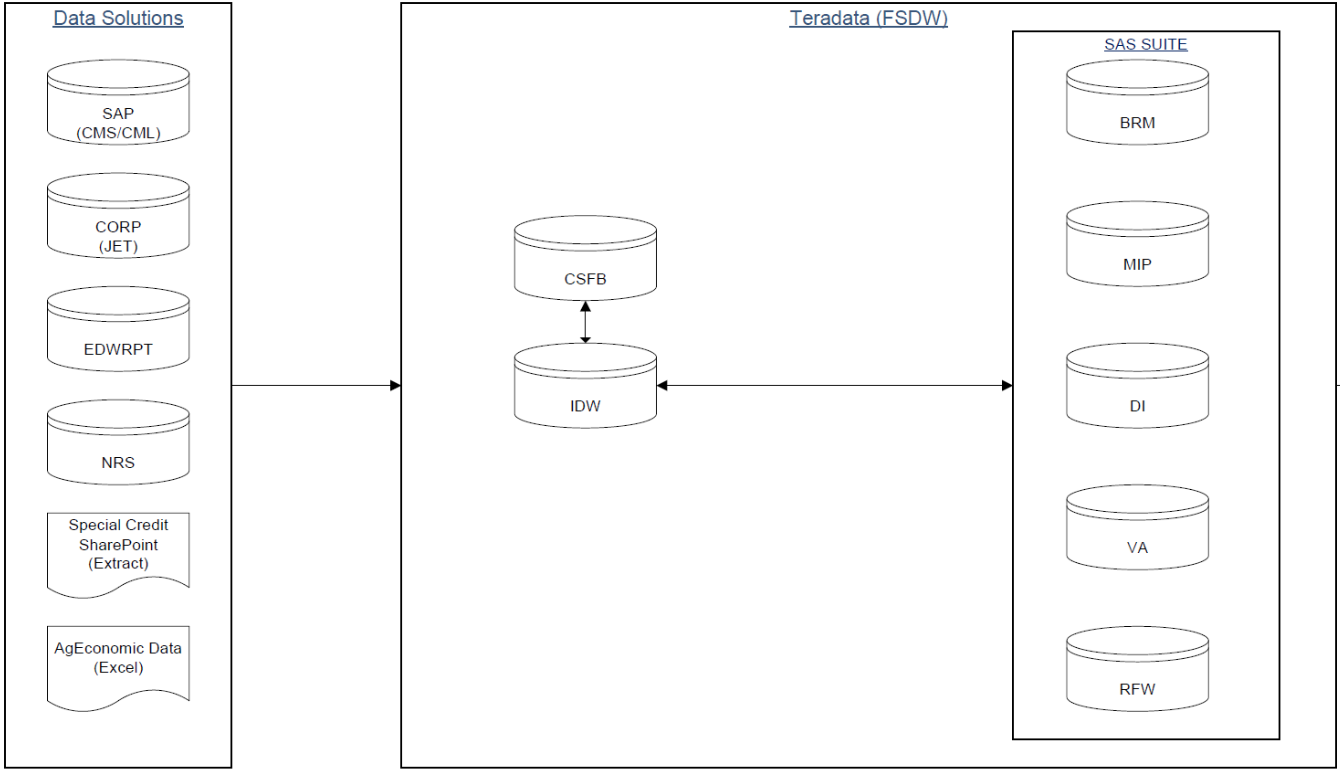
sas

R Studio

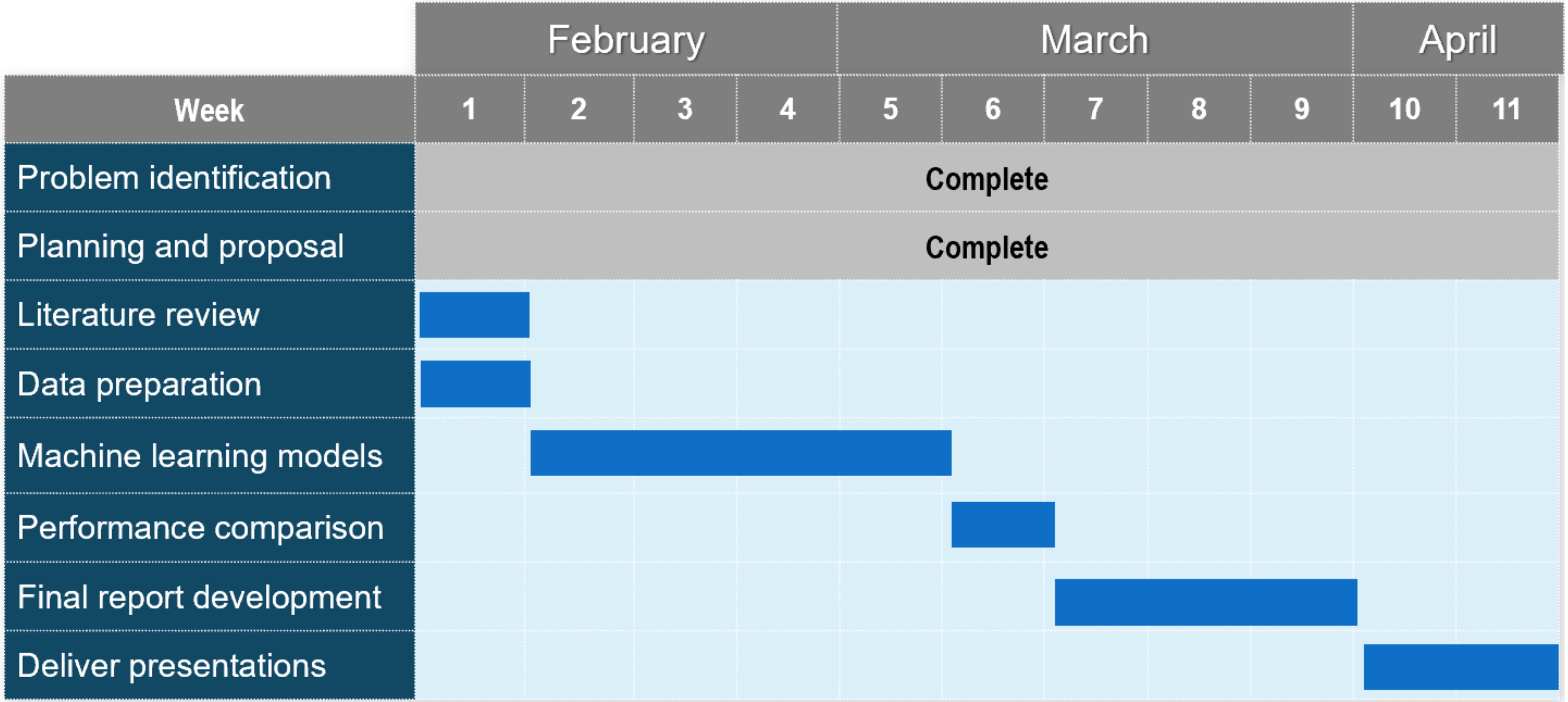
GitHub

Data & tools - data

Data & tools - data



Timelines


February
March
April
Week
1
2
3
4
5
6
7
8
9
10
11
Problem identification
Complete
Planning and proposal
Complete
Literature review
Data preparation
Machine learning models
Performance comparison
Final report development
Deliver presentations

Expected outcomes

1. Machine learning based model for Lifetime PD Term Structure

2. Machine learning based model for LGD

3. Know how current and proposed models compare in performance
   a) If better, potential to use proposed in production
   b) If not better (but still “good”), use as “challenger” model

4. Useable code that can be leveraged by my organization

5. Senior management has improved confidence in models and progress on ML adoption

University of Regina