1) Adaptive boosted model

Parameters = max\_depth = 5, minsplit = 5

Overall Gini coefficient = 0.05391558

Feature gain =

feature\_29 feature\_16

36.65518952 17.78671623

feature\_65 feature\_30

12.11424712 9.98239296

feature\_53 feature\_70

9.67581923 4.02308294

feature\_15 feature\_48

1.60296958 1.55198446

feature\_39 feature\_7

1.19350425 0.73973006

feature\_40 feature\_41

0.61853764 0.50390800

feature\_3 num\_of\_times\_overdue

0.46569247 0.33790282

mean\_diff\_lastpayment\_opened\_dt mean\_diff\_open\_enquiry\_dt

0.27622382 0.20145656

median\_high\_credit avg\_diff\_payment

0.18597915 0.18591612

dt\_opened Ratio\_currbalance\_creditlimit

0.17202936 0.17186467

entry\_time payment\_history\_mean\_length

0.15226528 0.14120835

mean\_paymenthistory\_avg\_dpd\_0\_20\_bucket.x median\_enq\_amtt

0.13891449 0.13761541

feature\_14 feature\_36

0.10825586 0.09384692

feature\_37 feature\_25

0.08789725 0.08743211

feature\_2 feature\_27

0.07761589 0.07747999

mean\_rateofinterest feature\_33

0.07372245 0.05253556

feature\_31 feature\_12

0.04956820 0.04693846

perc\_secured feature\_26

0.04509080 0.04290776

feature\_35 feature\_13

0.04118183 0.03511001

feature\_52 feature\_72

0.03354799 0.03171842

2) Adaptive boosted model (all numeric predictors scaled and centred)

Parameters = max\_depth = 10, minsplit = 5

Overall Gini coefficient = 0.04830945

Feature gain =

feature\_29 feature\_16

37.24396854 18.40461484

feature\_65 feature\_30

12.41108957 9.66604074

feature\_53 feature\_70

9.07014119 3.96518175

feature\_15 feature\_39

1.65072898 1.25086328

feature\_48 feature\_41

1.20806484 0.65564552

feature\_40 feature\_7

0.56129455 0.45850728

feature\_3 num\_of\_times\_overdue

0.44002892 0.33561463

median\_high\_credit avg\_diff\_payment

0.23859080 0.23774664

feature\_2 Ratio\_currbalance\_creditlimit

0.21628235 0.18053120

median\_enq\_amtt mean\_diff\_open\_enquiry\_dt

0.17451892 0.16375374

feature\_37 mean\_paymenthistory\_avg\_dpd\_0\_20\_bucket.x

0.16041834 0.14536883

payment\_history\_mean\_length feature\_31

0.12780031 0.12502781

feature\_14 feature\_26

0.11594730 0.10849441

dt\_opened mean\_diff\_lastpayment\_opened\_dt

0.10521900 0.08643456

feature\_35 total\_enquiries

0.08142256 0.07775804

feature\_52 feature\_12

0.07453875 0.04699121

feature\_1 feature\_51

0.04242497 0.03945120

feature\_72 feature\_27

0.03486990 0.03343656

feature\_42 feature\_36

0.03125574 0.02993220

3) Adaptive boosted model (all numeric predictors scaled and centred)

Parameters = max\_depth = 10, minsplit = 10

Overall Gini coefficient = 0.04542818

Feature gain =

feature\_29 feature\_16

37.88811804 17.85327670

feature\_65 feature\_30

11.77370557 10.03017136

feature\_53 feature\_70

8.61139650 3.81105046

feature\_15 feature\_48

1.86174514 1.60542091

feature\_39 feature\_7

1.31087294 0.81838297

feature\_41 feature\_40

0.48451784 0.39291447

avg\_diff\_payment feature\_3

0.30710605 0.30608325

median\_high\_credit Ratio\_currbalance\_creditlimit

0.26141586 0.25788405

dt\_opened feature\_2

0.24326957 0.18092182

mean\_paymenthistory\_avg\_dpd\_0\_20\_bucket.x mean\_diff\_lastpayment\_opened\_dt

0.17147347 0.16477076

num\_of\_times\_overdue feature\_37

0.16220825 0.15734340

mean\_diff\_open\_enquiry\_dt feature\_26

0.14668763 0.14397321

payment\_history\_mean\_length median\_enq\_amtt

0.14097708 0.12396677

entry\_time feature\_51

0.11176334 0.11033077

feature\_31 feature\_12

0.09283270 0.08973470

mean\_rateofinterest perc\_secured

0.08345962 0.06339743

feature\_13 feature\_27

0.05897746 0.05304960

feature\_25 feature\_72

0.05099910 0.04217915

feature\_35 feature\_1

0.03362204 0.00000000

4) Adaptive boosted model (all numeric predictors scaled and centred)

Parameters = max\_depth = 10, minsplit = 50

Overall Gini coefficient = 0.04639377

Feature Gain =

feature\_29 feature\_16

36.33970778 18.24126127

feature\_65 feature\_30

11.74474862 11.15436171

feature\_53 feature\_70

8.65099886 4.14848349

feature\_15 feature\_39

2.02797025 1.38981743

feature\_48 feature\_40

1.27249784 0.68124877

feature\_7 feature\_41

0.57119607 0.52491975

feature\_3 num\_of\_times\_overdue

0.33087981 0.29234395

Ratio\_currbalance\_creditlimit dt\_opened

0.27153590 0.26999480

payment\_history\_mean\_length feature\_2

0.23462877 0.16150415

mean\_diff\_lastpayment\_opened\_dt mean\_paymenthistory\_avg\_dpd\_0\_20\_bucket.x

0.15869301 0.15646295

median\_high\_credit mean\_diff\_open\_enquiry\_dt

0.14634015 0.13557520

feature\_37 feature\_27

0.12628194 0.12538388

avg\_diff\_payment feature\_14

0.10890132 0.09590330

median\_enq\_amtt feature\_52

0.07907359 0.07675223

feature\_12 feature\_31

0.07568020 0.07395872

mean\_rateofinterest total\_enquiries

0.06225866 0.04485894

median\_credit\_limit feature\_26

0.04168285 0.04043433

feature\_72 feature\_36

0.03327435 0.03253154

feature\_51 median\_cash\_limit

0.02774304 0.02697901

perc\_secured entry\_time

0.02313158 0.00000000

5) Random forest model (all numeric features scaled and centered)

Parameters = ntrees = 100, mtry = 10

Overall Gini coefficient = 0.04365031

Feature gain =

MeanDecreaseGini

dt\_opened 66.1341526

entry\_time 63.7981195

feature\_1 9.6244141

feature\_2 60.0000716

feature\_3 62.6765110

feature\_4 8.6805562

feature\_7 64.8151628

feature\_8 12.1860180

feature\_9 11.5608867

feature\_10 0.0000000

feature\_11 3.8874599

feature\_12 13.7674656

feature\_13 7.9879932

feature\_14 8.6793387

feature\_18 0.2569186

feature\_19 3.7651070

feature\_23 4.7989304

feature\_25 5.4935832

feature\_26 20.1154246

feature\_27 22.9966213

feature\_31 20.7641822

feature\_32 12.4157493

feature\_33 2.6976758

feature\_34 2.1579145

feature\_35 39.2009775

feature\_36 22.0366755

feature\_37 23.4436745

feature\_39 63.8799297

feature\_40 63.6280895

feature\_41 59.6247200

feature\_42 7.5456264

feature\_46 9.2453271

feature\_50 3.3652680

feature\_51 21.6448872

feature\_52 26.5722472

feature\_54 0.0000000

feature\_55 7.7161223

feature\_56 22.2668996

feature\_57 2.9411974

feature\_58 1.0291880

feature\_59 4.7508815

feature\_60 0.1705982

feature\_61 0.3426873

feature\_62 0.0000000

feature\_64 54.6234318

feature\_67 3.0417030

feature\_68 1.7471008

feature\_69 39.4238643

feature\_71 28.1174187

feature\_72 5.1039380

feature\_73 5.2732328

feature\_76 2.3425715

feature\_78 4.3448162

feature\_79 0.0000000

mean\_paymenthistory\_avg\_dpd\_0\_20\_bucket.x 32.3100159

last\_365day\_enquiries 8.8726341

last\_90day\_enquiries 4.2699023

mean\_diff\_lastpayment\_opened\_dt 79.5490871

mean\_diff\_open\_enquiry\_dt 70.3055337

num\_of\_times\_overdue 42.3804576

payment\_history\_mean\_length 68.7739912

Ratio\_currbalance\_creditlimit 66.4607831

mean\_paymenthistory\_avg\_dpd\_0\_20\_bucket.y 31.9536581

total\_enquiries 55.6345506

total\_secured 35.6325001

perc\_secured 45.4803172

utilisation\_trend 66.3527144

avg\_diff\_payment 67.8602193

median\_high\_credit 65.1945537

median\_credit\_limit 20.1960034

median\_cash\_limit 15.9239991

mean\_rateofinterest 47.2073686

median\_actualpaymentamount 25.9246859

median\_enq\_amtt 51.9726656

6) Random forest model (all numeric features scaled and centered)

Parameters = ntrees = 100, mtry = 20

Overall Gini coefficient = 0.04365031

Feature gain =

MeanDecreaseGini

dt\_opened 71.4436013

entry\_time 61.1473508

feature\_1 8.3213406

feature\_2 61.8354391

feature\_3 64.1184588

feature\_4 5.8461985

feature\_7 68.4773203

feature\_8 11.8771729

feature\_9 12.1918433

feature\_10 0.0000000

feature\_11 3.2498259

feature\_12 12.5380504

feature\_13 6.1601419

feature\_14 6.2341262

feature\_18 0.1313543

feature\_19 2.4244239

feature\_23 3.9670231

feature\_25 4.3400893

feature\_26 18.9647171

feature\_27 21.7613639

feature\_31 22.4930895

feature\_32 10.8953898

feature\_33 1.8767682

feature\_34 1.0584635

feature\_35 34.5967572

feature\_36 20.7176570

feature\_37 23.7519616

feature\_39 65.9971962

feature\_40 68.0754000

feature\_41 62.2886516

feature\_42 8.3181202

feature\_46 9.1504039

feature\_50 2.3547712

feature\_51 24.3480989

feature\_52 27.2249905

feature\_54 0.0000000

feature\_55 8.8890976

feature\_56 18.1251810

feature\_57 1.8198140

feature\_58 1.2166533

feature\_59 3.7340602

feature\_60 0.1181912

feature\_61 0.4982895

feature\_62 0.0000000

feature\_64 54.1730755

feature\_67 1.8347196

feature\_68 1.4278543

feature\_69 36.0451712

feature\_71 24.0503564

feature\_72 3.0466915

feature\_73 4.4258609

feature\_76 2.5900059

feature\_78 5.1931652

feature\_79 0.0000000

mean\_paymenthistory\_avg\_dpd\_0\_20\_bucket.x 34.6483573

last\_365day\_enquiries 9.7885258

last\_90day\_enquiries 5.3410561

mean\_diff\_lastpayment\_opened\_dt 84.5424158

mean\_diff\_open\_enquiry\_dt 74.1364841

num\_of\_times\_overdue 38.7232662

payment\_history\_mean\_length 71.9723283

Ratio\_currbalance\_creditlimit 60.6211365

mean\_paymenthistory\_avg\_dpd\_0\_20\_bucket.y 34.3165000

total\_enquiries 58.2850244

total\_secured 33.4291176

perc\_secured 44.1834782

utilisation\_trend 61.8119166

avg\_diff\_payment 72.5959681

median\_high\_credit 67.4308650

median\_credit\_limit 22.4567346

median\_cash\_limit 18.7327308

mean\_rateofinterest 53.3995785

median\_actualpaymentamount 30.2609739

median\_enq\_amtt 51.8713953

7) Random forest model (all numeric features scaled and centered)

Parameters = ntrees = 1000, mtry = 20

Overall Gini coefficient = 0.04365031

Feature gain =

MeanDecreaseGini

dt\_opened 70.839153153

entry\_time 60.738957949

feature\_1 8.297303935

feature\_2 61.354852271

feature\_3 65.830456985

feature\_4 5.768796824

feature\_7 67.613750436

feature\_8 11.761542693

feature\_9 11.536295418

feature\_10 0.000000000

feature\_11 3.539519974

feature\_12 12.507970535

feature\_13 6.117821193

feature\_14 6.630901856

feature\_18 0.424844611

feature\_19 2.431336955

feature\_23 4.103648092

feature\_25 4.362571631

feature\_26 18.600352339

feature\_27 22.758008566

feature\_31 23.586050259

feature\_32 9.976314857

feature\_33 1.771464412

feature\_34 1.443681259

feature\_35 36.815138147

feature\_36 20.461731207

feature\_37 23.002031144

feature\_39 65.567888766

feature\_40 68.384126237

feature\_41 62.303557479

feature\_42 8.360303249

feature\_46 9.403800601

feature\_50 2.450978047

feature\_51 21.991551690

feature\_52 26.919044888

feature\_54 0.000000000

feature\_55 8.165176731

feature\_56 19.747280617

feature\_57 1.849148679

feature\_58 1.344808557

feature\_59 3.439786665

feature\_60 0.151584564

feature\_61 0.264854762

feature\_62 0.000000000

feature\_64 55.031404536

feature\_67 1.740291679

feature\_68 1.205464984

feature\_69 36.645003531

feature\_71 23.639488311

feature\_72 3.138291825

feature\_73 4.023312280

feature\_76 2.312955967

feature\_78 4.673923908

feature\_79 0.004626263

mean\_paymenthistory\_avg\_dpd\_0\_20\_bucket.x 33.347937974

last\_365day\_enquiries 10.617581804

last\_90day\_enquiries 4.953062154

mean\_diff\_lastpayment\_opened\_dt 85.206622137

mean\_diff\_open\_enquiry\_dt 73.231577282

num\_of\_times\_overdue 39.388495347

payment\_history\_mean\_length 70.764678825

Ratio\_currbalance\_creditlimit 62.015131046

mean\_paymenthistory\_avg\_dpd\_0\_20\_bucket.y 33.467667522

total\_enquiries 56.391416033

total\_secured 34.715501316

perc\_secured 44.783493720

utilisation\_trend 62.038970404

avg\_diff\_payment 71.482421236

median\_high\_credit 67.264437551

median\_credit\_limit 21.817064298

median\_cash\_limit 18.424193700

mean\_rateofinterest 52.304876797

median\_actualpaymentamount 31.064551411

median\_enq\_amtt 52.890925349

8) Other models I tried running but could not finish (due to memory and/or time issues)

- Support Vector machines with radial kernel

- Flexible discriminant analysis

- Mixture disciminant analysis

- Extreme gradient boosting (Xgboost)

**So far, the best model appears to be a random forest classifier with 1000 trees and a predictor space of 20 at each split.**