CENTRE FOR
SOCIAL DATA ANALYTICS

Predictive Analytics in Child Welfare:

Implementation of a predictive model to support child maltreatment hotline screening decisions: methods, challenges and lessons learned

Diana Benavides Prado,
Data Scientist

June 2nd, 2017

AUT UNIVERSITY
NEW ZEALAND

Context

- Child maltreatment is an international public health problem
- **~40 million children worldwide** are subject to some kind of **maltreatment**, per year
- There are over **3 million calls** per year made to **child welfare agencies** concerning abuse or neglect in the US alone.

Children in the Public Benefit System at Risk of Maltreatment

Identification Via Predictive Modeling

Rhema Vaithianathan, PhD, Tim Maloney, PhD, Emily Putnam-Hornstein, PhD, Nan Jiang, PhD

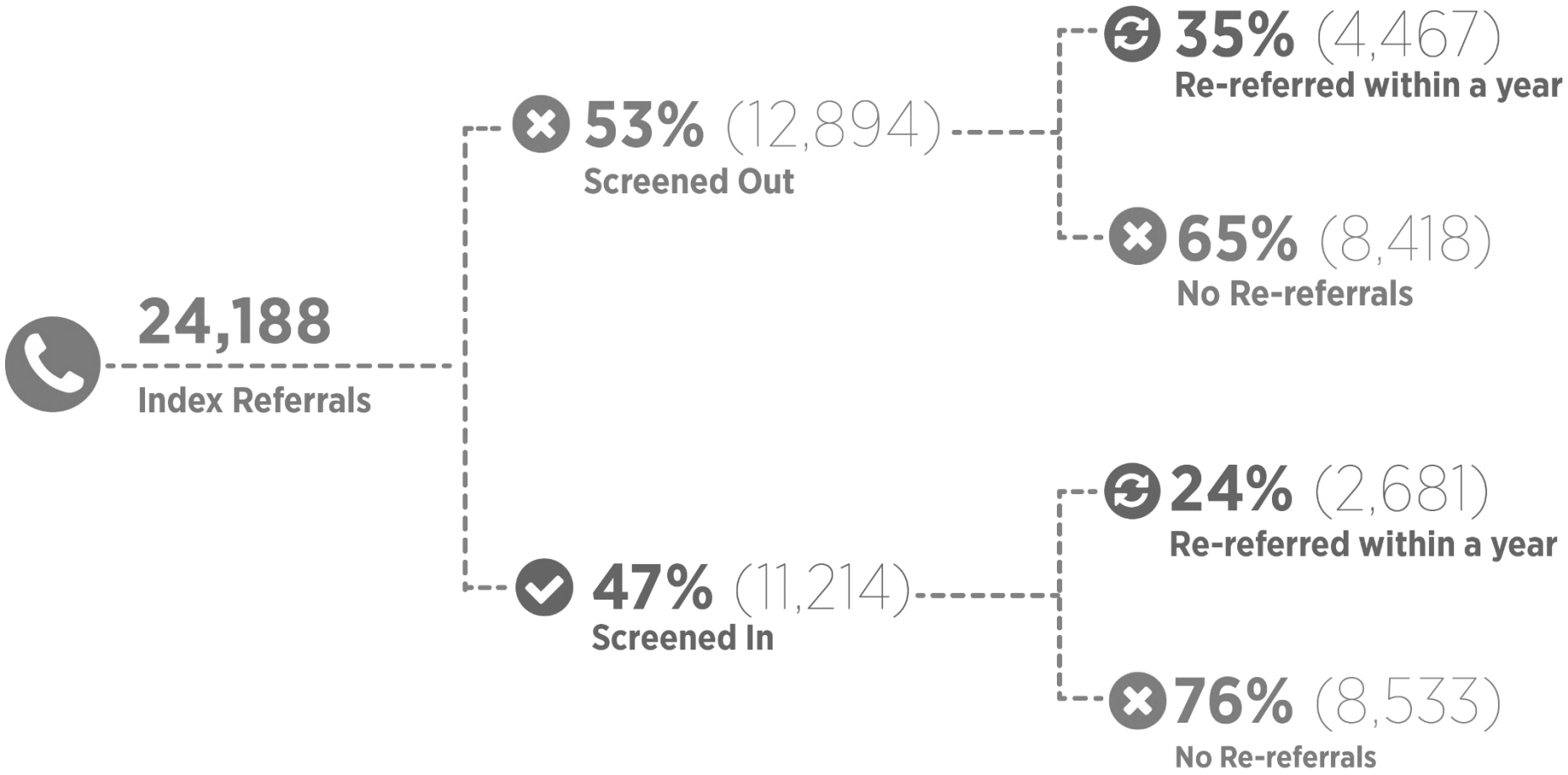
Abstract: A growing body of research links child abuse and neglect to a range of negative short- and long-term health outcomes. Determining a child's risk of maltreatment at or shortly after birth provides an opportunity for the delivery of targeted prevention services. This study presents findings from a predictive risk model (PRM) developed to estimate the likelihood of substantiated maltreatment among children enrolled in New Zealand's public benefit system. The objective was to explore the potential use of administrative data for targeting prevention and early intervention services to children and families.

A data set of integrated public benefit and child protection records for children born in New Zealand between January 1, 2003, and June 1, 2006, was used to develop a risk algorithm using stepwise probit modeling. Data were analyzed in 2012. The final model included 132 variables and produced an area under the receiver operating characteristic curve of 76%. Among children in the top decile of risk, 47.8% had been substantiated for maltreatment by age 5 years. Of all children substantiated for maltreatment by age 5 years, 83% had been enrolled in the public benefit system before age 2 years. This analysis demonstrates that PRMs can be used to generate risk scores for substantiated maltreatment. Although a PRM cannot replace more-comprehensive clinical assessments of abuse and neglect risk, this approach provides a simple and cost-effective method of targeting early prevention services.

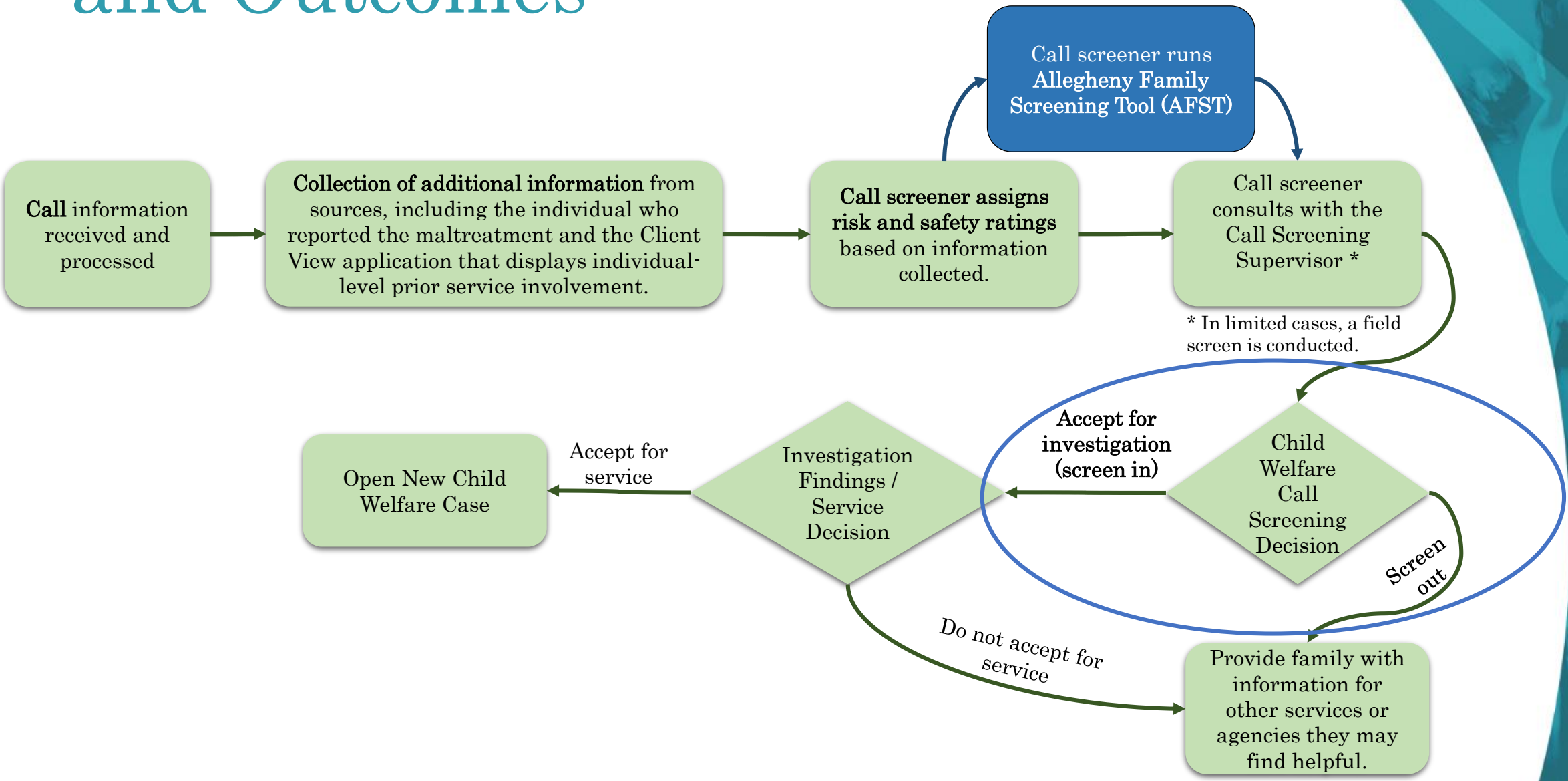
(Am J Prev Med 2013;45(3):354–359)

Context: Allegheny County (Pittsburgh, PA, US)

Problem: Screening Decisions and Outcomes

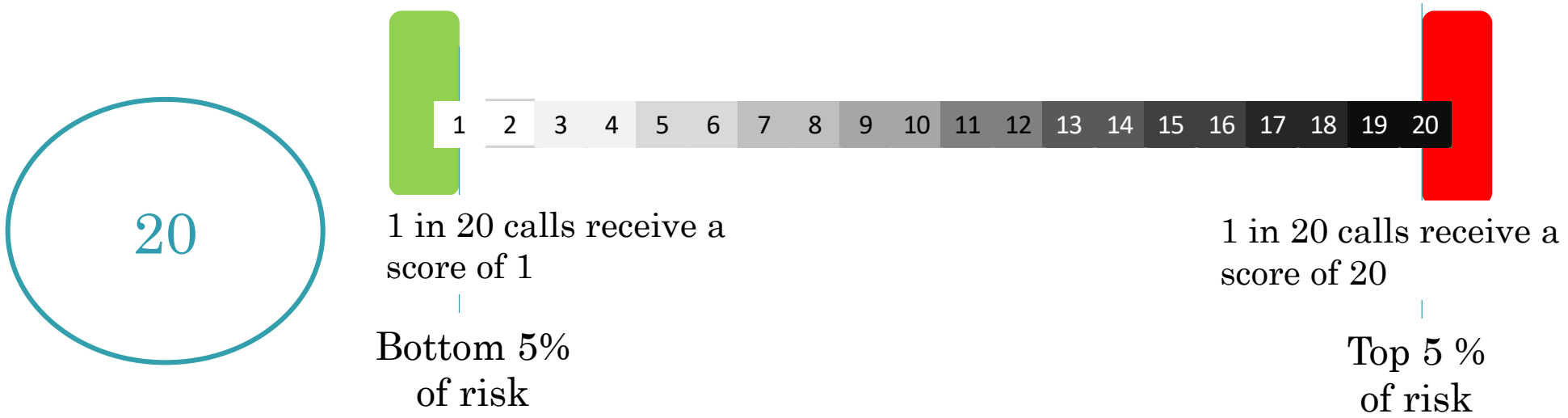


Problem: Screening Decisions and Outcomes

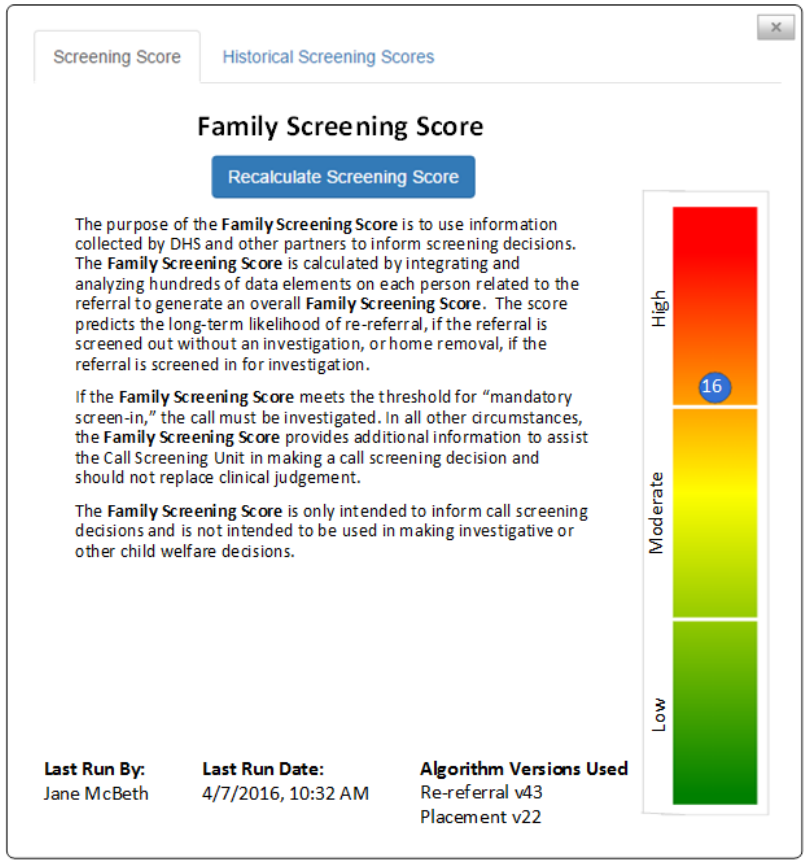


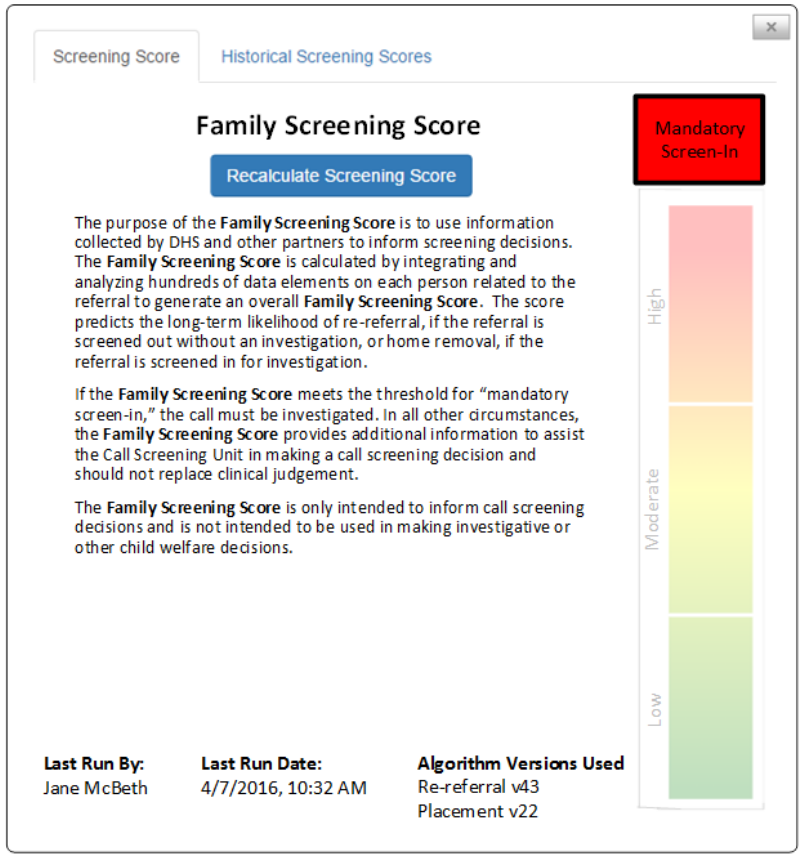
Problem: Screening Decisions and Outcomes

- What's the right question that a predictive model should help to answer?
  - What's the risk that a child will be **re-referred conditional on being screened-out**.
  - What's the risk that a child will be **placed in foster care conditional on being screened-in**.



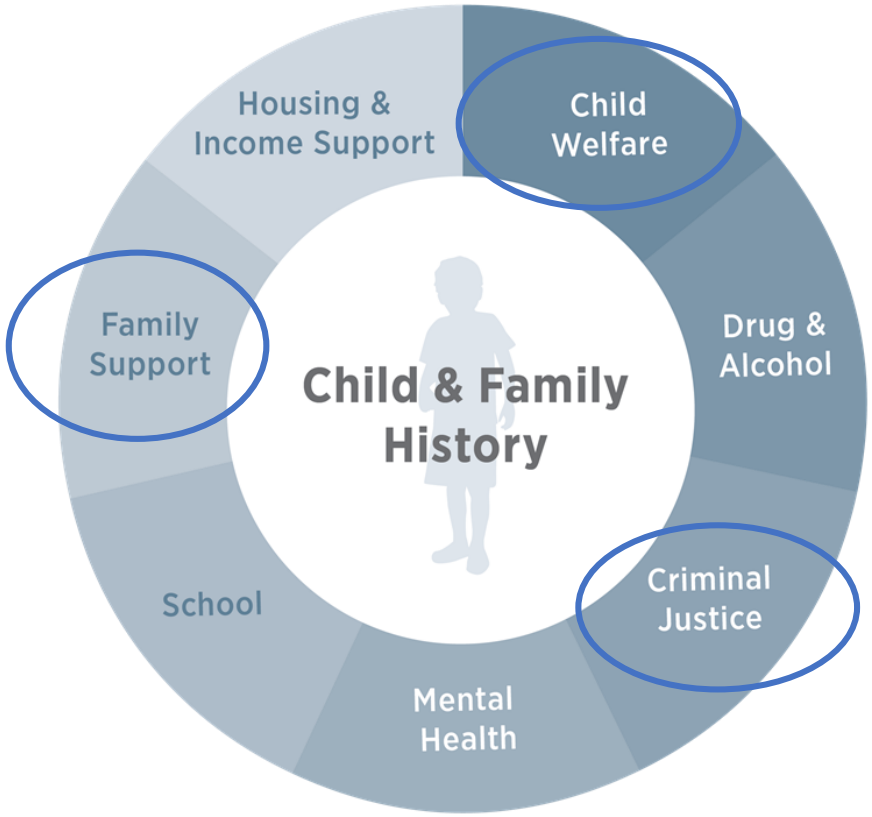
Problem: Screening Decisions and Outcomes





A subset of high scores that are particularly correlated with possible critical incidents are flagged for **mandatory screen-in**, with an override option for unique circumstances

Data



Referral ID	Individual ID	Role	Individual Info…	Referral Info…

- Several administrative data sources
- More than 800 variables/predictors
- De-identified data for research/analysis purposes
- What's the ideal "shape" of the data for analysis?

Modelling v1.0

Extract:

The “right” patterns

Identify correctly the high risk population

Initial criteria:

“Good” balance between true positives and false negatives

Modelling v1.0

- Logistic Regression

$$\hat{p} = \frac{\exp(b_0 + b_1X_1 + b_2X_2 + \ldots + b_pX_p)}{1 + \exp(b_0 + b_1X_1 + b_2X_2 + \ldots + b_pX_p)}$$

59 predictors for the re-referral model

71 predictors for the placement model

Modelling v2.0

- Random Forests

(Breiman, 2001)

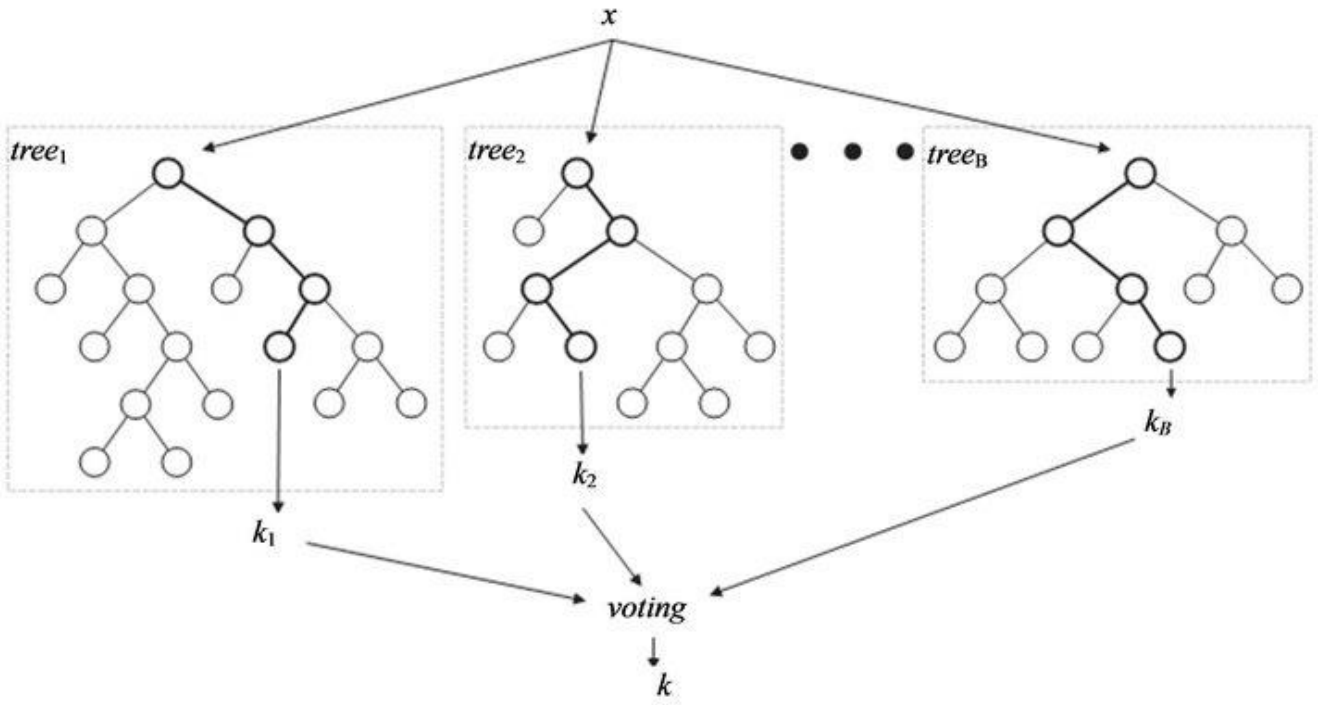


Figure:
http://file.scirp.org/Html/6-9101686_31887.htm

Modelling v2.0

- XGBoost

(Chen, Guestrin, 2016)

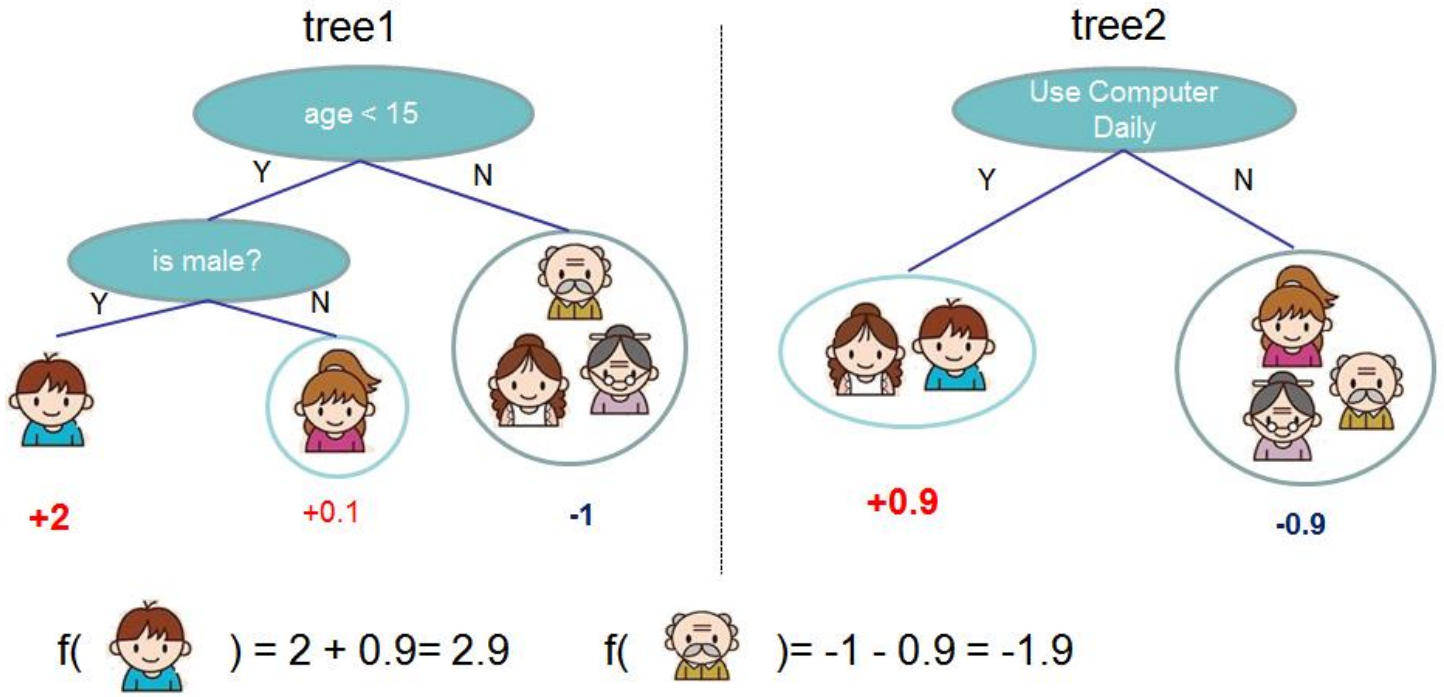
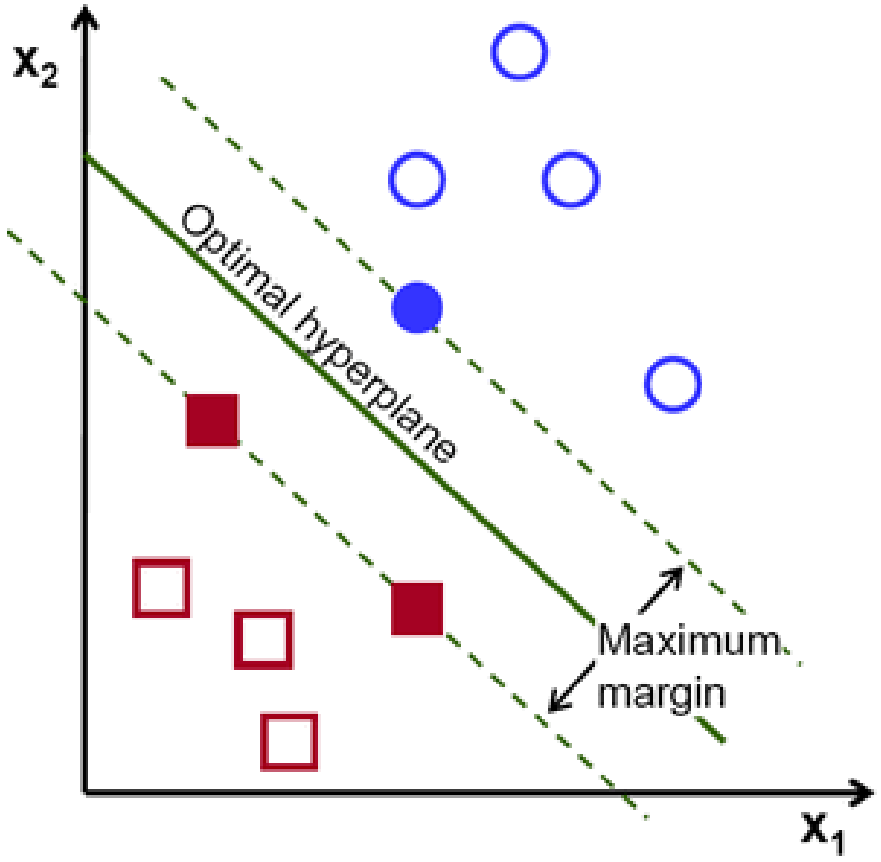


Figure: http://xgboost.readthedocs.io/en/latest/model.html

Modelling v2.0

- SVM

(Vapnik, Chervonenkis, 1963)



There's NFL…

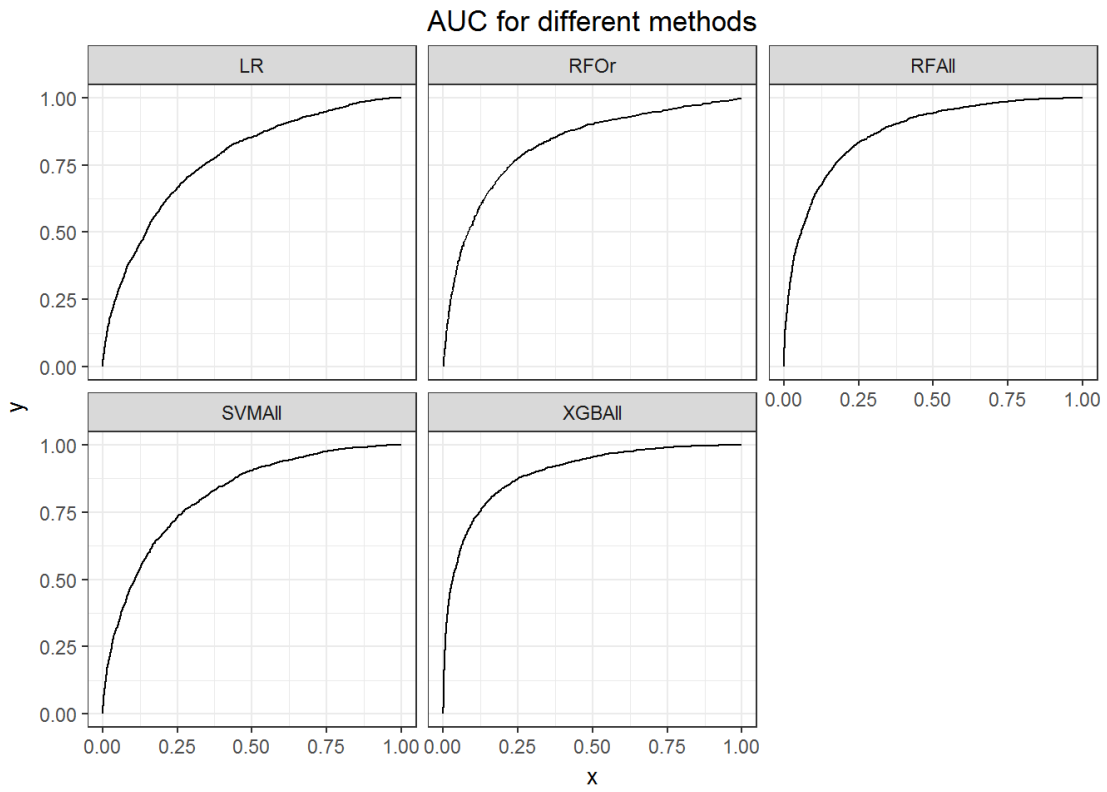
Figure: http://docs.opencv.org/2.4/doc/tutorials/ml/introduction_to_svm/introduction_to_svm.html

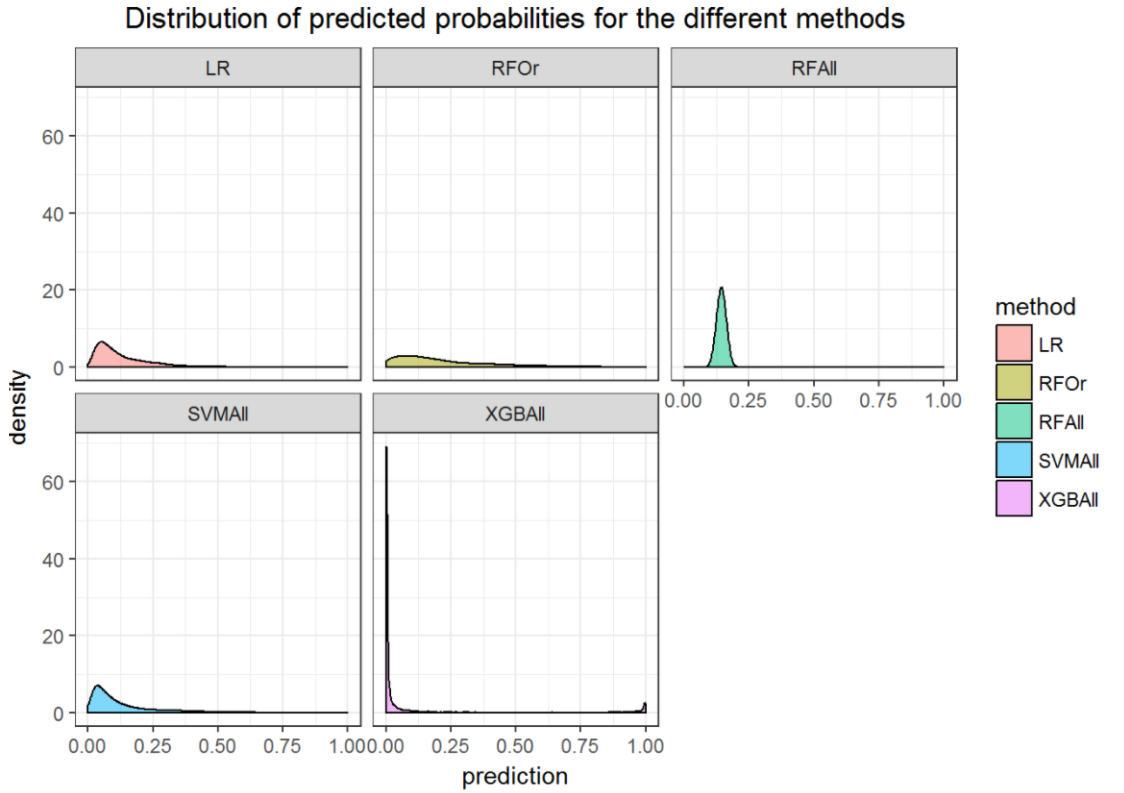
Modelling v2.0

- Logistic Regression
  - More familiar to a wider audience
  - Not so accurate
- Random Forests
  - Human-readable, to an extent (rules)
  - Usually accurate for high dimensional data
- XGBoost
  - Super-extreme and scalable
  - Often complex, though human-readable (to an extent)
- SVM
  - Less explanatory
  - Usually accurate for high dimensional numeric data

Evaluating what's been learned…

- What are the right metrics to compare – what do we want to achieve?

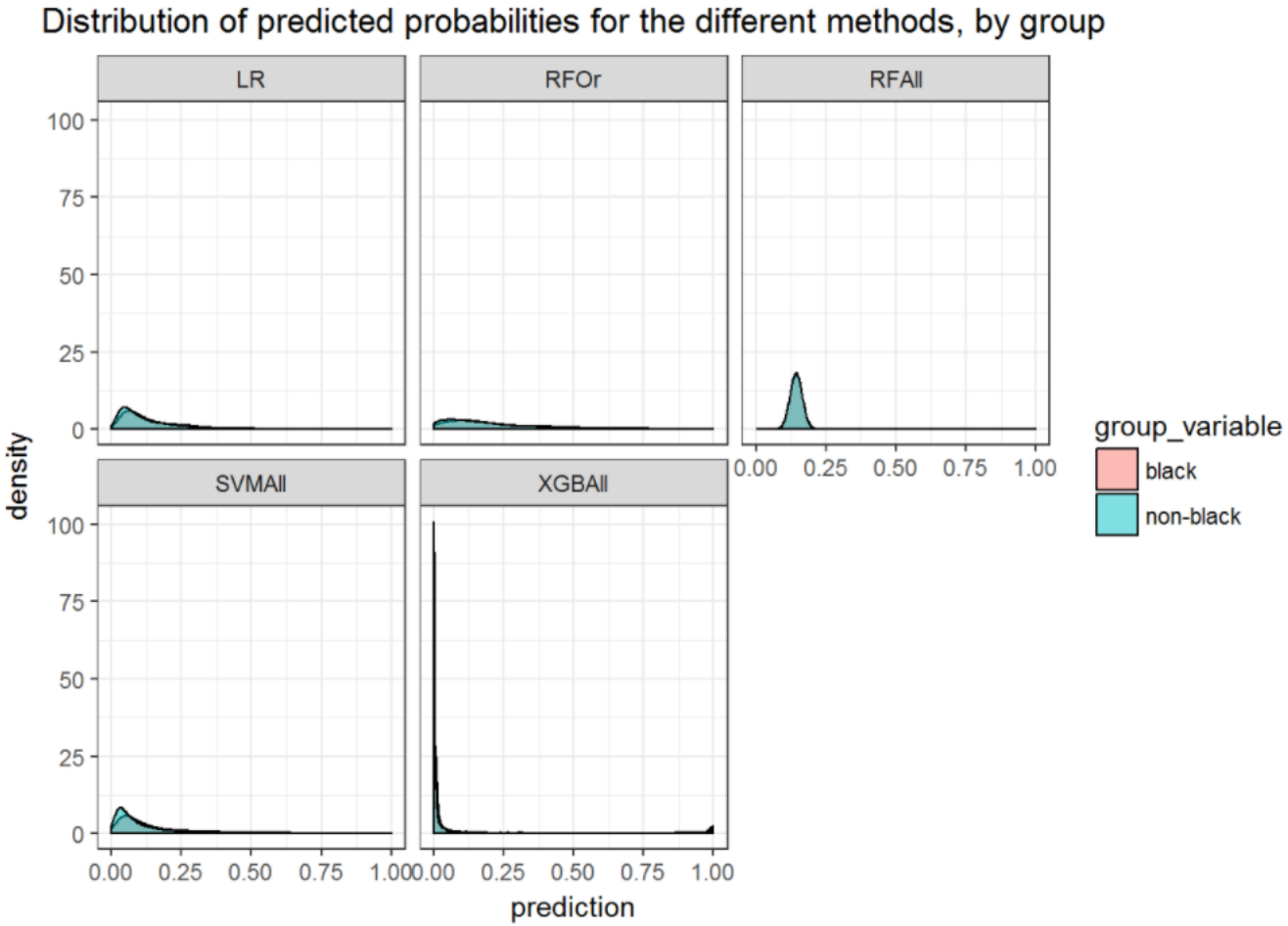




Evaluating what's been learned…

- What's the impact of what we are learning from data?

Disparities and Fairness…



Evaluation and Implementation v2.0

Extract:

The “right” patterns

Identify correctly the high risk population

Subject to:

A “good” explanation

Additional criteria:

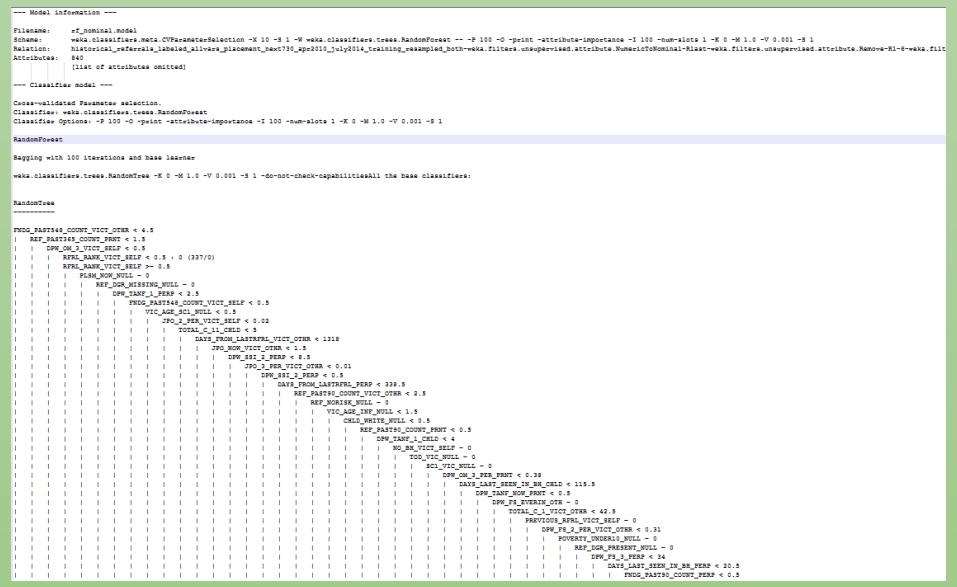
Interpretability/Transparency… Fairness

Implementation v2.0

- How can we explain a pattern, a prediction?

  The right to an explanation…

Simple rules (RF), lots of them…



```
--- Model information ---

Filename:     rf_nominal.model
Scheme:       weka.classifiers.meta.CVParameterSelection -X 10 -S 1 -W weka.classifiers.trees.RandomForest -- -P 100 -O -print -attribute-importance -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1
Relation:     historical_referrals_labeled_allvars_placement_new730_apr2010_july2014_training_resampled_both-weka.filters.unsupervised.attribute.NumericToNominal-Rlast-weka.filters.unsupervised.attribute.Remove-R1-6-weka.filt
Attributes:   640
              [list of attributes omitted]

--- Classifier model ---

Cross-validated Parameter selection.
Classifier: weka.classifiers.trees.RandomForest
Classifier Options: -P 100 -O -print -attribute-importance -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1

RandomForest

Bagging with 100 iterations and base learner

weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1 -do-not-check-capabilitiesAll the base classifiers:


RandomTree
==========

FNDG_PAST548_COUNT_VICT_OTHR < 4.5
|   REF_PAST548_COUNT_PRNT < 1.5
|   |   DPW_ON_3_VICT_SELF < 0.5
|   |   |   RFRL_RANK_VICT_SELF < 0.8 : 0 (337/0)
|   |   |   RFRL_RANK_VICT_SELF >= 0.8
|   |   |   |   PLSN_NON_NULL = 0
|   |   |   |   REF_DGR_MISSING_NULL = 0
|   |   |   |   |   DPW_TANF_1_PERP < 2.5
|   |   |   |   |   |   FNDG_PAST548_COUNT_VICT_SELF < 0.5
|   |   |   |   |   |   |   VIC_AGE_SCI_NULL < 0.5
|   |   |   |   |   |   |   |   JPO_2_PER_VICT_SELF < 0.02
|   |   |   |   |   |   |   |   |   TOTAL_C_11_CHLD < 5
|   |   |   |   |   |   |   |   |   |   DAYS_FROM_LASTRFRL_VICT_OTHR < 1318
|   |   |   |   |   |   |   |   |   |   |   JPO_NON_VICT_OTHR < 1.5
|   |   |   |   |   |   |   |   |   |   |   |   DPW_SEI_3_PERP < 0.5
|   |   |   |   |   |   |   |   |   |   |   |   |   JPO_2_PER_VICT_OTHR < 0.01
|   |   |   |   |   |   |   |   |   |   |   |   |   |   DPW_SEI_3_PERP < 0.5
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   DAYS_FROM_LASTRFRL_PERP < 338.5
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   REF_PAST90_COUNT_VICT_OTHR < 2.5
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   REF_NONRSK_NULL = 0
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   VIC_AGE_INF_NULL < 1.5
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   CHLD_WHITE_NULL < 0.5
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   REF_PAST90_COUNT_PRNT < 0.5
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   DPW_TANF_1_CHLD < 4
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   RC_BH_VICT_SELF = 0
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   TOD_VIC_NULL = 0
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   SCI_VIC_NULL = 0
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   DPW_ON_3_PER_PRNT < 0.38
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   DAYS_LAST_SEEN_IN_BH_CHLD < 113.5
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   DPW_TANF_NON_PRNT < 0.5
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   DPW_FS_EVERIN_OTR = 0
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   TOTAL_C_1_VICT_OTHR < 42.5
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   PREVIOUS_RFRL_VICT_SELF = 0
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   DPW_FS_2_PER_VICT_OTHR < 0.31
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   POVERTY_UNDER10_NULL = 0
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   REF_DGR_PRESENT_NULL = 0
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   DPW_FS_3_PERP < 34
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   DAYS_LAST_SEEN_IN_BH_PERP < 20.5
|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   FNDG_PAST90_COUNT_PERP < 0.5
```

Complex formulas (SVM)

```
svm_type c_svc
kernel_type rbf
gamma 0.0011919
nr_class 2
total_sv 18949
rho 0.0621671
label 1 -1
probA -1.3873
probB 1.31897
nr_sv 3805 15144
SV
0.8 1:-0.333333 7:-1 8:-1 9:-1 10:-1 11:-1 12:1 13:-1 14:-1 15:0.68 16:-1 17:-1 18:-1 19:-1 20:1 22:-1 24:-1 25:1 26:-1 27:-1 28:-1 29:-1 30:-1 31:1 32:1 33:-1 34:-1 35:-1 36:1 37:-1 38:1 39:-1 40:
0.8 1:-0.438596 7:-1 8:-1 9:-1 10:-1 11:1 12:-1 13:-1 14:-1 15:0.68 16:-1 17:-1 18:-1 19:-1 20:1 22:-1 24:-1 25:1 26:-1 27:-1 28:-1 29:-1 30:-1 31:1 32:1 33:-1 34:-1 35:-1 36:1 37:-1 38:1 39:-1 40:
0.8 1:-0.754386 7:-1 8:-1 9:1 10:-1 11:-1 12:-1 13:-1 14:1 15:-0.164 16:-1 17:-1 18:-1 19:1 20:-1 22:-1 24:-1 25:1 26:-1 27:-1 28:-1 29:-1 30:-1 31:1 32:1 33:-1 34:1 35:1 36:-1 37:-1 38:-1 39:-1 40
0.8 1:-0.649123 7:-1 8:-1 9:-1 10:1 11:-1 12:-1 13:-1 14:-1 15:-0.164 16:-1 17:-1 18:-1 19:1 20:-1 22:-1 24:-1 25:1 26:-1 27:-1 28:-1 29:-1 30:-1 31:1 32:1 33:-1 34:1 35:1 36:-1 37:-1 38:-1 39:-1 4
0.8 1:-0.508772 7:-1 8:-1 9:-1 10:-1 11:1 12:-1 13:-1 14:1 15:-0.164 16:-1 17:-1 18:-1 19:1 20:-1 22:-1 24:-1 25:1 26:-1 27:-1 28:-1 29:-1 30:-1 31:1 32:1 33:-1 34:1 35:1 36:-1 37:-1 38:-1 39:-1 40
0.8 1:-0.438596 7:-1 8:-1 9:-1 10:-1 11:1 12:-1 13:-1 14:1 15:-0.164 16:-1 17:-1 18:-1 19:1 20:-1 22:-1 24:-1 25:1 26:-1 27:-1 28:-1 29:-1 30:-1 31:1 32:1 33:-1 34:1 35:1 36:-1 37:-1 38:-1 39:-1 40
0.8 1:-0.859649 7:1 8:-1 9:-1 10:-1 11:-1 12:-1 13:-1 14:1 15:0.68 16:-1 17:-1 18:-1 19:-1 20:1 22:-1 24:-1 25:1 26:-1 27:-1 28:-1 29:-1 30:-1 31:1 32:-1 33:-1 34:1 35:-1 36:1 37:-1 38:-1 39:1 40:-
0.8 1:-0.649123 7:-1 8:-1 9:-1 10:1 11:-1 12:-1 13:-1 14:-1 15:0.68 16:-1 17:-1 18:-1 19:-1 20:1 22:-1 24:-1 25:1 26:-1 27:-1 28:-1 29:-1 30:-1 31:1 32:-1 33:-1 34:1 35:-1 36:1 37:-1 38:-1 39:1 40:
0.8 1:-0.333333 7:-1 8:-1 9:-1 10:-1 11:-1 12:1 13:-1 14:-1 15:-0.788 16:-1 17:1 18:-1 19:-1 20:-1 22:-1 24:1 25:-1 26:-1 27:-1 28:-1 29:-1 30:-1 31:1 32:1 33:-1 34:1 35:-1 36:1 37:-1 38:-1 39:1 40
0.8 1:-0.614035 7:-1 8:-1 9:-1 10:1 11:-1 12:-1 13:-1 14:-1 15:-0.024 16:-1 17:-1 18:-1 19:1 20:-1 22:-1 24:-1 25:1 26:-1 27:-1 28:-1 29:1 30:-1 31:1 32:-1 33:-1 34:-1 35:1 36:-1 37:-1 38:-1 39:-1
0.8 1:-0.649123 7:-1 8:-1 9:-1 10:1 11:-1 12:-1 13:-1 14:-1 15:-0.024 16:-1 17:-1 18:-1 19:1 20:-1 22:-1 24:-1 25:1 26:-1 27:-1 28:-1 29:1 30:-1 31:1 32:-1 33:-1 34:-1 35:1 36:-1 37:-1 38:-1 39:-1
0.8 1:-0.368421 7:-1 8:-1 9:-1 10:-1 11:-1 12:1 13:-1 14:1 15:-0.42 16:-1 17:-1 18:1 19:-1 20:-1 22:-1 24:-1 25:1 26:-1 27:-1 28:-1 29:-1 30:-1 31:1 32:-1 33:-1 34:1 35:-1 36:1 37:-1 38:-1 39:1 40:
0.8 1:-0.403509 7:-1 8:-1 9:-1 10:-1 11:-1 12:1 13:-1 14:1 15:-0.488 16:-1 17:-1 18:1 19:-1 20:-1 22:-1 24:-1 25:1 26:-1 27:-1 28:-1 29:-1 30:-1 31:1 32:1 33:-1 34:1 35:1 36:-1 37:-1 38:-1 39:1 40:
0.8 1:-0.333333 7:-1 8:-1 9:-1 10:-1 11:-1 12:1 13:-1 14:1 15:-0.056 16:-1 17:-1 18:-1 19:1 20:-1 22:-1 24:-1 25:1 26:-1 27:-1 28:-1 29:-1 30:-1 31:1 32:1 33:-1 34:-1 35:-1 36:1 37:-1 38:-1 39:1 40
0.8 1:-0.684211 7:-1 8:-1 9:1 10:-1 11:-1 12:-1 13:-1 14:-1 15:-0.144 16:-1 17:-1 18:-1 19:1 20:-1 22:-1 24:-1 25:1 26:-1 27:-1 28:-1 29:-1 30:-1 31:1 32:1 33:-1 34:1 35:1 36:-1 37:-1 38:-1 39:1 40
0.8 1:-0.824561 7:-1 8:1 9:-1 10:-1 11:-1 12:-1 13:-1 14:-1 15:-0.048 16:-1 17:-1 18:-1 19:1 20:-1 22:-1 24:-1 25:1 26:-1 27:-1 28:-1 29:-1 30:-1 31:1 32:-1 33:-1 34:-1 35:1 36:-1 37:-1 38:-1 39:-1
0.8 1:-0.719298 7:-1 8:-1 9:1 10:-1 11:-1 12:-1 13:-1 14:1 15:-0.552 16:-1 17:-1 18:1 19:-1 20:-1 22:-1 24:1 25:-1 26:-1 27:-1 28:-1 29:-1 30:-1 31:1 32:1 33:-1 34:1 35:1 36:-1 37:-1 38:-1 39:1 40:
0.8 1:-0.754386 7:-1 8:-1 9:1 10:-1 11:-1 12:-1 13:-1 14:-1 15:-0.552 16:-1 17:-1 18:1 19:-1 20:-1 22:-1 24:1 25:-1 26:-1 27:-1 28:-1 29:-1 30:-1 31:1 32:1 33:-1 34:1 35:1 36:-1 37:-1 38:-1 39:1 40
0.8 1:-0.859649 7:1 8:-1 9:-1 10:-1 11:-1 12:-1 13:-1 14:1 15:-0.2 16:-1 17:-1 18:1 19:-1 20:-1 22:-1 24:-1 25:1 26:-1 27:-1 28:-1 29:-1 30:-1 31:1 32:-1 33:-1 34:-1 35:-1 36:-1 37:-1 38:-1 39:-1 40
0.8 1:-0.684211 7:-1 8:-1 9:1 10:-1 11:-1 12:-1 13:-1 14:1 15:-0.608 16:-1 17:1 18:-1 19:-1 20:-1 22:-1 24:1 25:-1 26:-1 27:-1 28:-1 29:-1 30:-1 31:1 32:-1 33:-1 34:-1 35:-1 36:-1 37:1 38:-1 39:-1
0.8 1:-0.789474 7:-1 8:1 9:-1 10:-1 11:-1 12:-1 13:-1 14:1 15:-0.608 16:-1 17:1 18:-1 19:-1 20:-1 22:-1 24:1 25:-1 26:-1 27:-1 28:-1 29:-1 30:-1 31:1 32:-1 33:-1 34:-1 35:-1 36:-1 37:1 38:-1 39:-1
0.8 1:-0.614035 7:-1 8:-1 9:-1 10:1 11:-1 12:-1 13:-1 14:-1 15:-0.488 16:-1 17:-1 18:1 19:-1 20:-1 22:-1 24:1 25:-1 26:-1 27:-1 28:-1 29:-1 30:-1 31:1 32:1 33:-1 34:1 35:-1 36:1 37:-1 38:1 39:-1 40:
0.8 1:-0.298246 7:-1 8:-1 9:-1 10:-1 11:-1 12:1 13:-1 14:-1 15:-0.488 16:-1 17:-1 18:1 19:-1 20:-1 22:-1 24:1 25:-1 26:-1 27:-1 28:-1 29:-1 30:-1 31:1 32:1 33:-1 34:1 35:-1 36:1 37:-1 38:1 39:-1 40
0.8 1:-0.403509 7:-1 8:-1 9:-1 10:-1 11:-1 12:1 13:-1 14:-1 15:-0.216 16:-1 17:-1 18:1 19:-1 20:-1 22:-1 24:-1 25:1 26:-1 27:-1 28:-1 29:-1 30:-1 31:1 32:1 33:-1 34:1 35:-1 36:1 37:-1 38:1 39:-1 40
0.8 1:-0.438596 7:-1 8:-1 9:-1 10:-1 11:1 12:-1 13:-1 14:-1 15:-0.2 16:-1 17:-1 18:1 19:-1 20:-1 22:-1 24:-1 25:1 26:-1 27:-1 28:-1 29:-1 30:-1 31:1 32:1 33:-1 34:1 35:-1 36:1 37:-1 38:1 39:1 40:-
0.8 1:-0.789474 7:-1 8:1 9:-1 10:-1 11:-1 12:-1 13:-1 14:-1 15:-0.632 16:-1 17:1 18:-1 19:-1 20:-1 22:-1 24:1 25:-1 26:-1 27:-1 28:-1 29:-1 30:-1 31:-1 32:-1 33:-1 34:-1 35:-1 36:-1 37:1 38:-1 39:-
0.8 1:-0.438596 7:-1 8:-1 9:-1 10:-1 11:1 12:-1 13:-1 14:1 15:-0.02 16:-1 17:-1 18:-1 19:1 20:-1 22:-1 24:-1 25:1 26:-1 27:-1 28:-1 29:-1 30:-1 31:-1 32:-1 33:-1 34:1 35:-1 36:1 37:-1 38:1 39:-1 40
0.8 1:-0.859649 7:1 8:-1 9:-1 10:-1 11:-1 12:-1 13:-1 14:1 15:0.68 16:-1 17:-1 18:-1 19:-1 20:1 22:-1 24:-1 25:1 26:-1 27:-1 28:-1 29:-1 30:-1 31:1 32:-1 33:-1 34:-1 35:1 36:-1 37:-1 38:-1 39:1 40:
0.8 1:-0.824561 7:-1 8:1 9:-1 10:-1 11:-1 12:-1 13:-1 14:1 15:-0.024 16:-1 17:-1 18:-1 19:1 20:-1 22:-1 24:1 25:1 26:-1 27:-1 28:-1 29:-1 30:-1 31:1 32:1 33:-1 34:-1 35:-1 36:1 37:-1 38:-1 39:-1 4
0.8 1:-0.438596 7:-1 8:-1 9:-1 10:-1 11:1 12:-1 13:-1 14:1 15:0.576 16:-1 17:-1 18:-1 19:-1 20:1 22:-1 24:-1 25:1 26:-1 27:-1 28:-1 29:-1 30:-1 31:1 32:1 33:-1 34:1 35:1 36:-1 37:-1 38:-1 39:-1 40:
0.8 1:-0.438596 7:-1 8:-1 9:-1 10:-1 11:1 12:-1 13:-1 14:1 15:-0.632 16:-1 17:1 18:-1 19:-1 20:-1 22:-1 24:1 25:-1 26:-1 27:-1 28:-1 29:-1 30:-1 31:1 32:1 33:-1 34:1 35:-1 36:1 37:-1 38:-1 39:1 40
0.8 1:-0.719298 7:-1 8:-1 9:1 10:-1 11:-1 12:-1 13:-1 14:1 15:-0.02 16:-1 17:-1 18:-1 19:1 20:-1 22:-1 24:-1 25:1 26:-1 27:-1 28:-1 29:-1 30:-1 31:1 32:-1 33:-1 34:1 35:1 36:-1 37:-1 38:-1 39:-1 4
```

The (near) future…

- Challenges at the State level (CA, US) - more data, possibly more concerns about disparities
- 'Hot' topics:
  - Ethics: disparities, fairness, stigmatization
  - Interpretability/Transparency, for a wide variety of audiences
  - Accountability
  - Engagement with, and high-level knowledge of, ML from the involved parties

Machine learning: the power and promise of computers that learn by example, Royal Society, April 2017.

Thanks.

Allegheny's Methodology Paper:

http://www.alleghenycountyanalytics.us/index.php/2017/04/17/developing-predictive-risk-models-support-child-maltreatment-hotline-screening-decisions/

Visit/follow CSDA:

https://csda.aut.ac.nz/

@AUTCSDA

@rvaithianathan