Modeling Maternal Mortality Rates (MMR)

Data Science Part Time Course

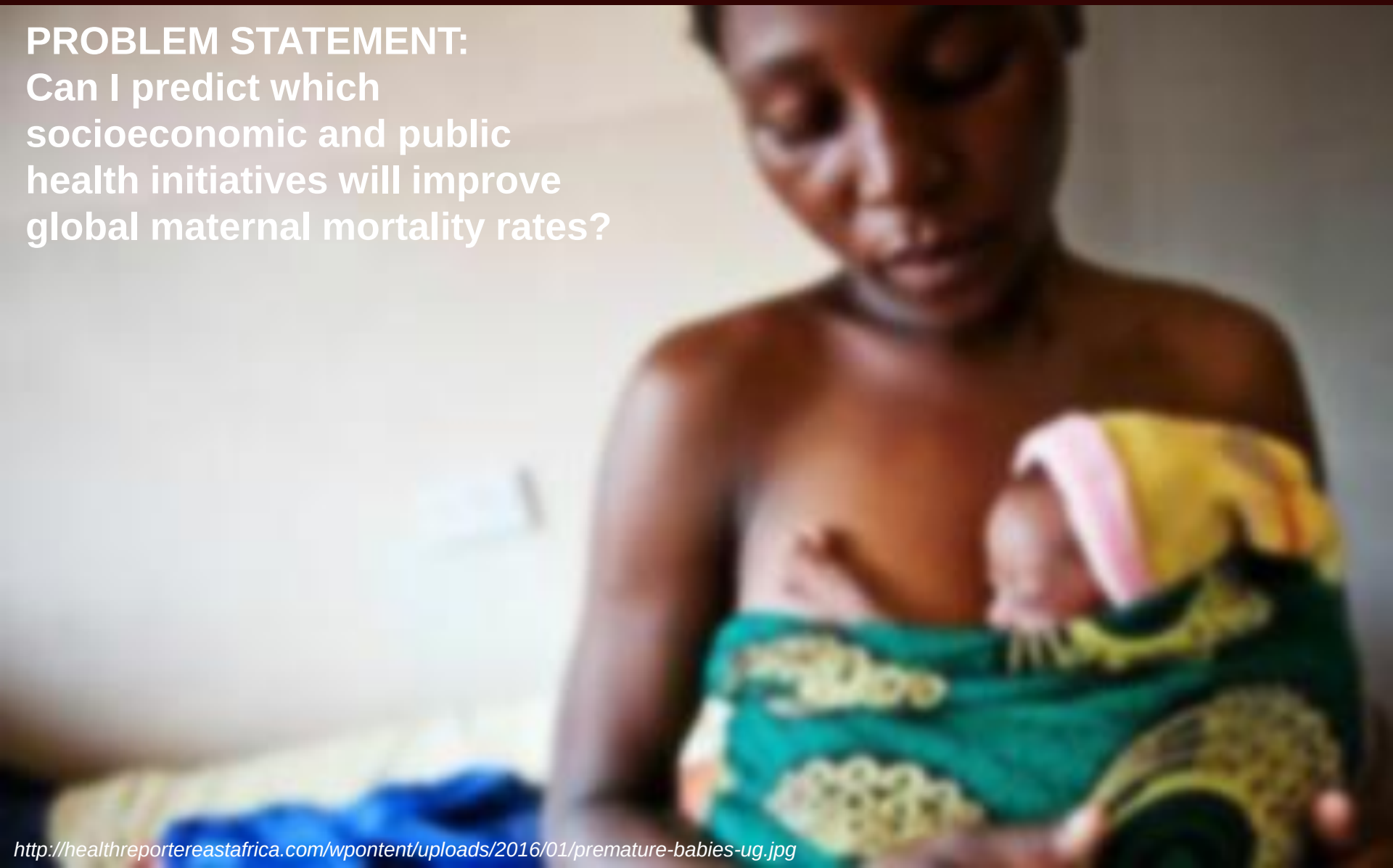
Rebecca Minich

11.29.16

https://assets.weforum.org/article/image/large_b05pCeiWDULOlBGVxJ52gFYvy6EPYx8R5WyOhPz1lNQ.jpg

What factors contribute to global MMR?

PROBLEM STATEMENT:
Can I predict which socioeconomic and public health initiatives will improve global maternal mortality rates?



http://healthreportereastafrica.com/wpontent/uploads/2016/01/premature-babies-ug.jpg

Could other factors contribute to MMR?

Datasets

Maternal Mortality Rates (MMR):

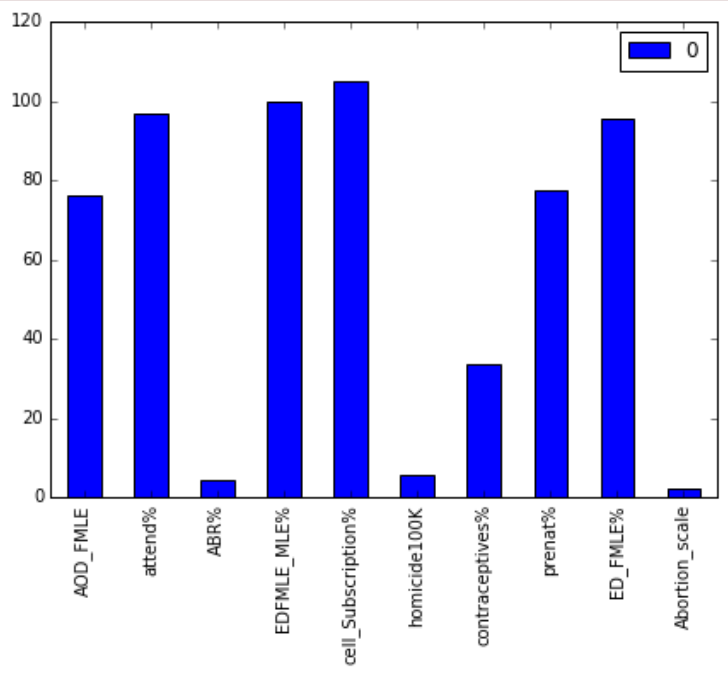
A. Birth attendant
B. Antenatal (prenatal) care
C. Adolescent Birth Rates
D. Contraceptives modern methods
E. Abortion policy scale
F. Education: Primary School Enrollment
G. Technology: Cellphone Subscribers
H. War: Homicides
I. Health: Life Expectancy
J. Wealth: GNI Per Capita Data by Country(PPP)

World Health Organization

CENTER FOR REPRODUCTIVE RIGHTS

Dataframe and Feature Engineering

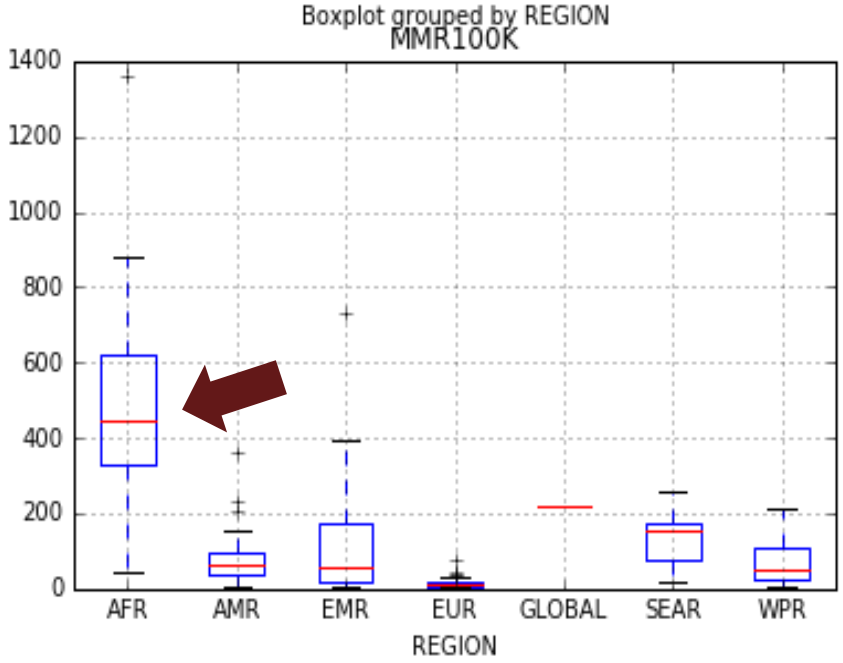
- 190 countries
- 12 different datasets, 33 total columns
- Missing data filled with median values:



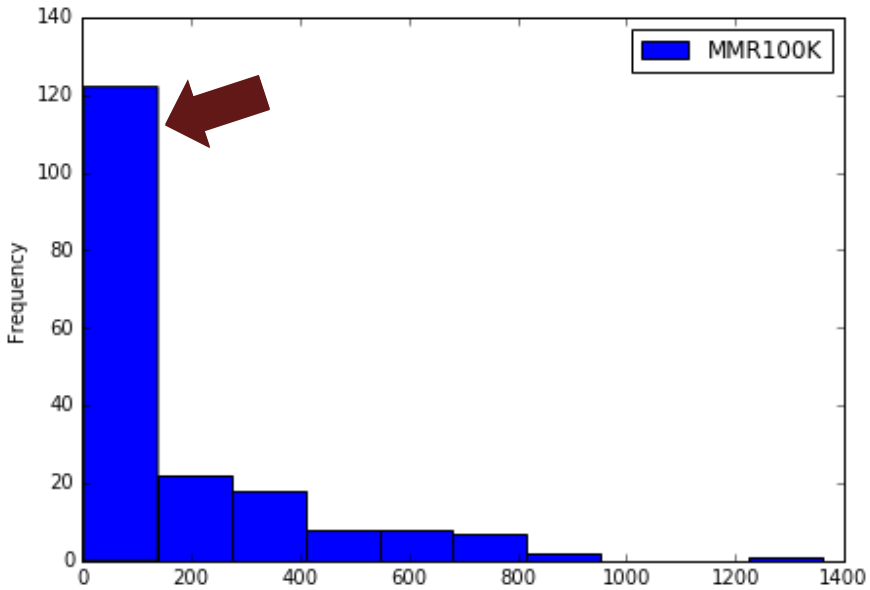
AOD_FMLE	76.06
attend%	96.60
GNI	10080.00
ABR%	4.54
EDFMLE_MLE%	99.82
cell_Subscription%	105.00
homicide100K	5.40
contraceptives%	33.73
prenat%	77.60
ED_FMLE%	95.37
Abortion_scale	2.00

Response Variable Engineering

Regional MMR



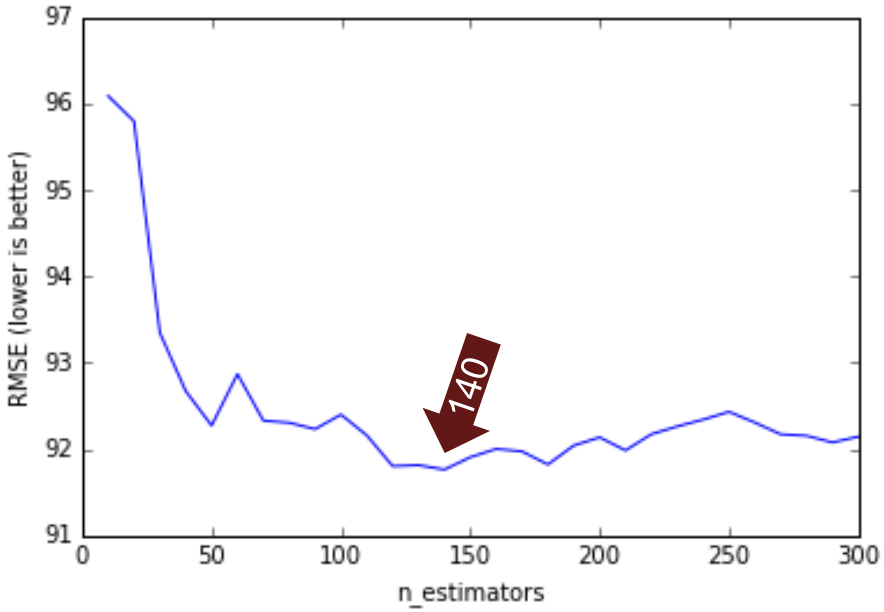
MMR Frequency



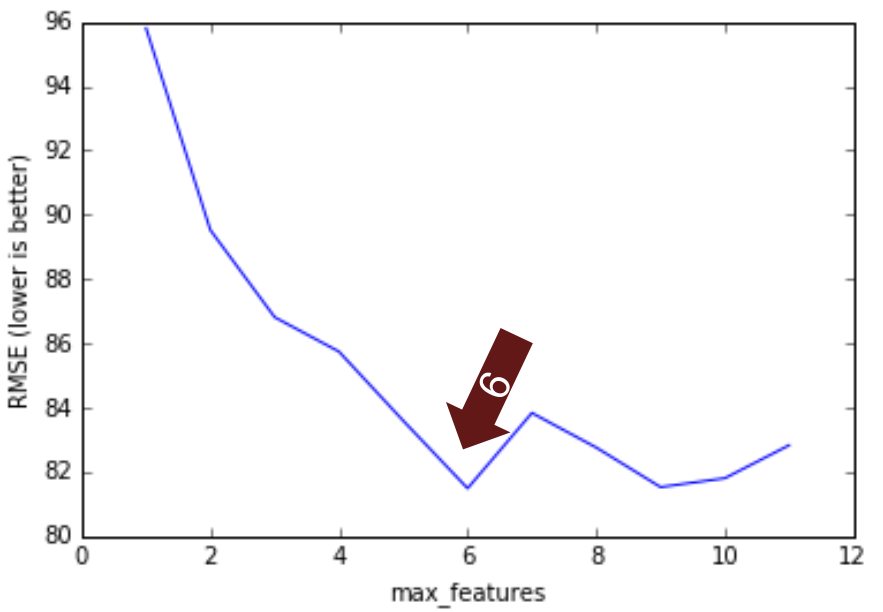
1. MMR100K - Maternal deaths per 100,000 births
2. MMRClassifier - Created for logistic regression model based on min, max and 25, 50, 75% of MMR.
3. MMRBinary - Binary variable for logistic regression model.

Optimizing Features: Random Forest Regressor

Estimator Tuning



Feature Tuning



	RMSE Null	RMSE Total	RMSE Optimized
MMR100K	**228.39**	91.88	82.65

Table 2: Optimized demographics include: 'ABR%', 'AOD_FMLE', 'GNI', 'EDFMLE_MLE%', 'cell_Subscription%', 'attend%'. The Random Forest Regressor model was. Null RMSE is with y_predict = mean.

Model Comparisons

	Null	Log Reg	Log Reg (Optimized)	Ran For	Ran For (Optimized)
MMRBinary	**0.526316**	0.873684211	0.857894737	0.8895	0.8947
MMRClassifier	**0.268421**	0.692464986	0.674336644	0.7558	0.7482
MMR100K (oob)	**0.052632**	Na	Na	0.815*	0.8348*

*Table 2: Optimized demographics include: 'ABR%', 'AOD_FMLE', 'GNI', 'EDFMLE_MLE%', 'cell_Subscription%', 'attend%'. The Random Forest Classifier model was used for MMRBinary and MMRClassifier. *Random Forest Regressor was used as the model for MMR100K. Null accuracy scores are the percent chance of selecting the most frequent value (MMRBinary = 1, MMRClassifier = 2, MMR100K = 54).*

- All of the models improve null accuracy.
- The random forest classifier performs better than the logistic regression model.
- Reducing the number of demographics improves model accuracy.

**The random forest regressor model was used for MMR100K and is the only model run that produces MMR predictions.*

Conclusions

OPTIMIZED FEATURES included in the model:
* AOD_FMLE - 0.810598
* attend% - 0.077110
* GNI - 0.029861
* ABR% - 0.019747
* EDFMLE_MLE% - 0.012234
* cell_Subscription% - 0.011530

Features excluded from the model:
* ED_FMLE% - 0.011345
* prenat% - 0.010079
* homicide100K - 0.008285
* contraceptives% - 0.006246
* Abortion_scale - 0.002966

Demographic contributions to model:

- Successful hypothesis: Education
- Surprising outcomes: Abortion policy, cell phone subscriptions

Future Work

- Build a model that lacks well known causal demographics and determine the predictability.
- Add more demographics from the WHO website.