



Predicting the **Big Move**

*Predicting Rural Migration in India
using socio-demographic information*

*Akshad Viswanathan, Anjali Kumari, Imran Ali
Richa Gupta , Varun Sayal*

And the mover is.....

- The objective:

Key goal of this study is to identify the demographic and geographic traits of a typical Indian migrant.

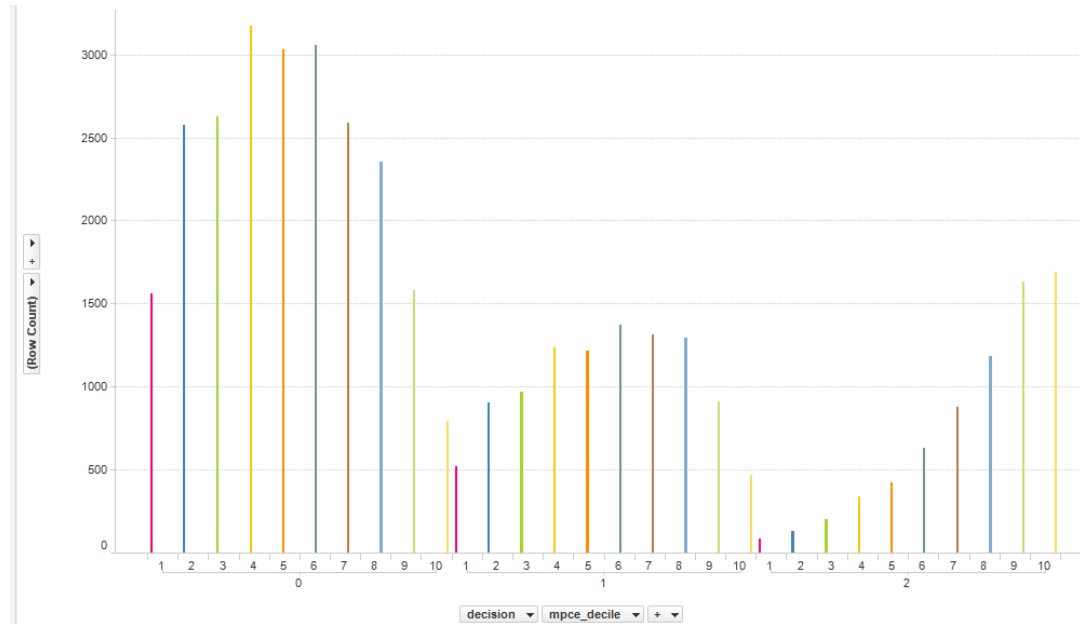
Using these we predict how likely is it a person from rural India will migrate and the direction of migration

- Who would benefit from this study

- The Government for schemes like NREGA
- Employment bureaus
- Urban Planning Departments
- Economists



Delving into the data.....



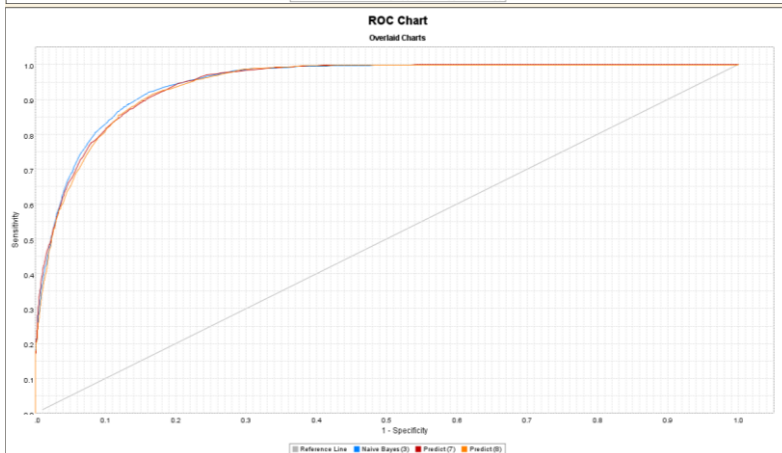
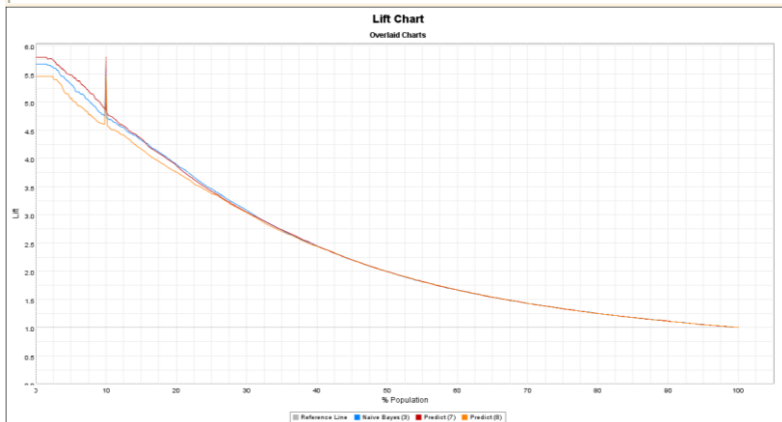
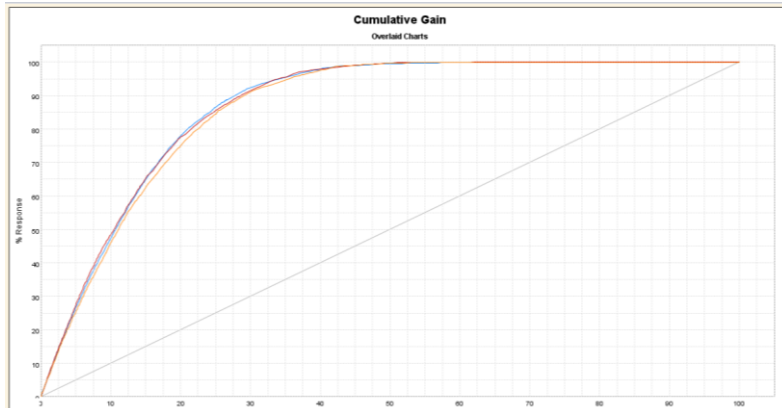
Predictors

*Region,
Age,
Householdtype,
Social group,
Landpossession,
Education,
Agricultural Labour,
Student,
Casual Labour etc*

- **Data Source:**
We used data from the national survey on migration with over 500,000 respondents from all over India. It had over 150 fields which could be potential predictors
- **Area of focus**
 - We created a categorical variable “decision” to affectively capture migration and its direction
 - We saw “data finger prints” using histograms and ran classification tree to analyze which predictors can bring value to the model
 - We finally pruned the data to identify 19 predictors

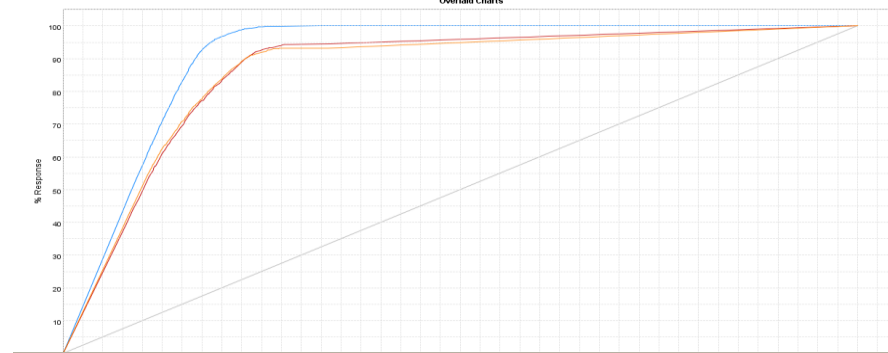
Seeing the light - Models

Naïve Bayes

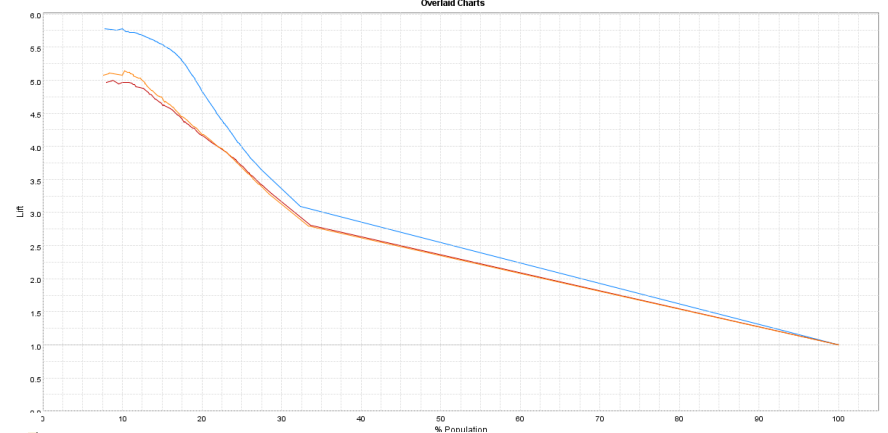


CART

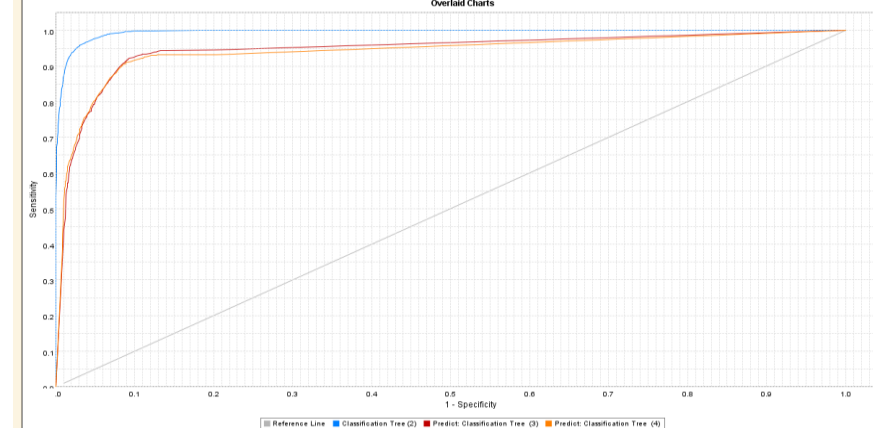
Cumulative Gain



Lift Chart



ROC Chart

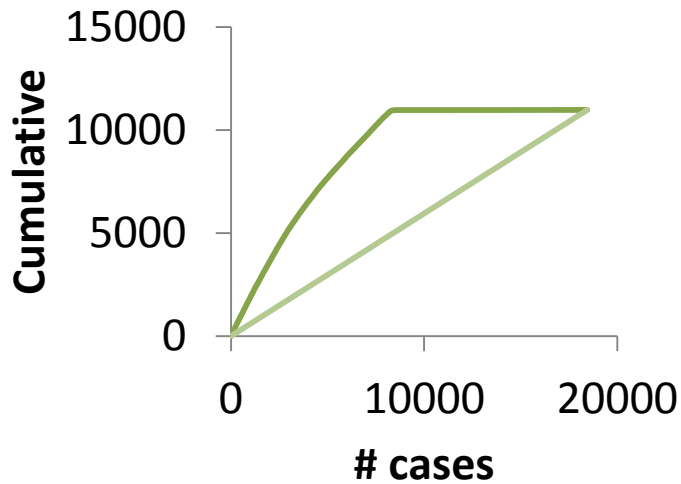


Seeing the light - Models

KNN

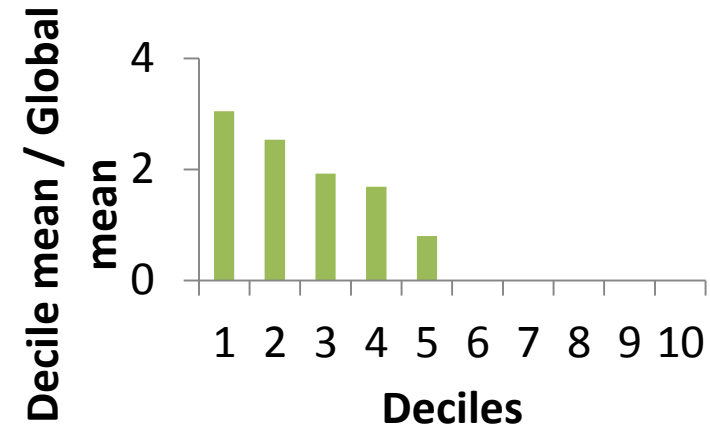
Training Data scoring - Summary Report (for k=16)		
Total sum of squared errors	RMS Error	Average Error
0	0	0
Validation Data scoring - Summary Report (for k=16)		
Total sum of squared errors	RMS Error	Average Error
1985.233204	0.328114412	0.00702096
Test Data scoring - Summary Report (for k=16)		
Total sum of squared errors	RMS Error	Average Error
1319.135507	0.327578782	0.014186566

Gains chart



— Cumulative decision when sorted using predicted values

Decile-wise lift chart



Measuring Performance.....

Metric	Naïve Bayes	K-NN	CART
%age Agreement	86.30%	87.74%	89.55%
Def Ratio 0	93.30%	95.91%	95.40%
Def Ratio1	76.50%	79.74%	83.80%
Def Ratio 2	78.30%	73.15%	78.40%

0: No Migration **1:** Migration from Rural to Rural **2:** Migration from Rural to Urban

Conclusions

- ✓ Both classification tree and K-NN came out to be more effective than Naïve Bayes
- ✓ CART fared better on Agreement and Definitude ratios than K-NN
- ✓ Result: Given the 19 predictors which can be easily known for an individual our models can strongly predict the migration and it's direction

Questions??