CUSTOMER CHURN PREDICTION

Supervisor : Dr. Dinesh Gopalani

Honey Duhar (2012ucp1531)
Hemant Kumar Meena (2012ucp1160)
Pavan Mahawar (2012ucp1211)

Problem Statement

- Consumer brands often offer discounts to attract new shoppers to buy their products or to retain the existing customer. The most valuable customers are those who return after this initial incentive purchase.
- So our task is to provide a prediction model to predict which shoppers are most likely to repeat purchase. In other words **Customer Churn Prediction.**

Dataset

- To create the prediction, a minimum of a year of shopping history prior to each customer's incentive, as well as the purchase histories of many other shoppers is provided.
- The transaction history contains all items purchased, not just items related to the offer.
- This data captures the process of offering incentives to a large number of customers and forecasting those who will become loyal to the product.

Files :

We are provided with four relational files:

- **transactions.csv** - contains transaction history for all customers for a period of at least 1 year prior to their offered incentive.
- **trainHistory.csv -** contains the incentive offered to each customer and information about the behavioral response to the offer.
- **testHistory.csv** - contains the incentive offered to each customer but does not include their response (we need predict the repeater column for each id in this file.
- **offers.csv** - contains information about the offers.

Fields

Data Files	Properties
Past Transactions	Customer ID, store, product department, product company, product category, product brand, date of purchase, product size, product size, product measure, purchase quantity, purchaseAmount
Training History	Customer ID, store, offer ID, geographical region, number of repeat trips, repeater, offer date
Testing History	Customer ID, store, offer ID, geographical region, number of repeat trips, repeater, offer date
Offers	Offer ID, offer category, offerquantity, offer company, offer value, offer brand

Algorithms Used

- In our classification problem we are using –
  - Naive Bayes Classification
  - Decision Tree Classification
  - Bagging Tree
  - Random Forest Classification

Naive Bayes Classification

- Naive Bayes classifier is a simple probabilistic classifier based on applying Bayes' theorem with strong (naive) independence assumptions between the features.

- Bayes theorem -

$$p(C_k|X) = p(C_k)\frac{p(X|C_k)}{p(X)}$$

- p(x) is same for all features, so

$$p(C_k|X) \propto p(C_k)p(X|C_k)$$

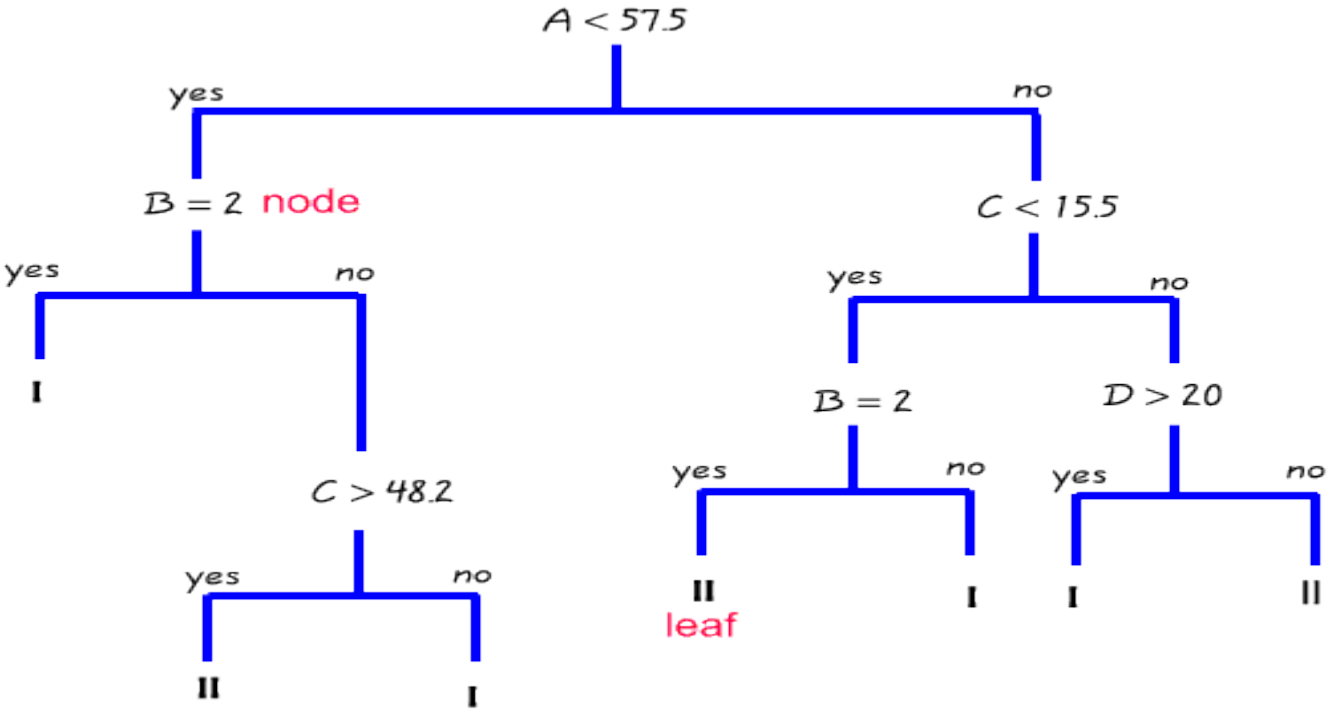
- Now we assumed that each sample attribute is independent to each other which is a strong assumption in naïve Bayes, so the final classification hypothesis becomes.

$$y = \underset{k\varepsilon\{1,\ldots.K\}}{argmax}\ p(C_k)\ \prod_{i=1}^{n} p(x_i|C_k)$$

- For predicting $p(X_i|C_k)$ when X_i is continuous variable we fit a Gaussian curve to the predict the required probabilities.
- $$p(x = v|c) = \left(\frac{1}{\sqrt{2\pi\sigma_c^2}}\right) e^{-\frac{(v-\mu_c)^2}{2\sigma_c^2}}$$

Decision Tree

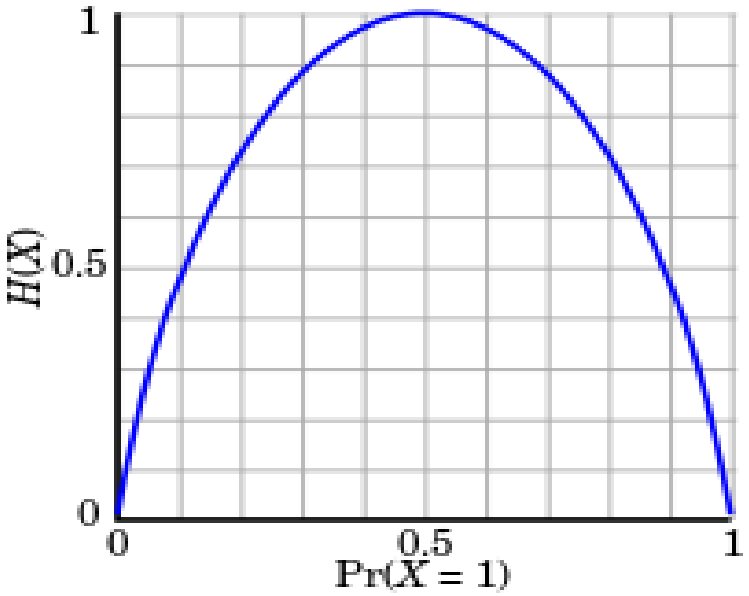
- Commonly used machine learning approach for classification and regression.
- It creates a hypothesis tree providing classification probabilities in the leaf nodes.



Splitting Criteria: Entropy

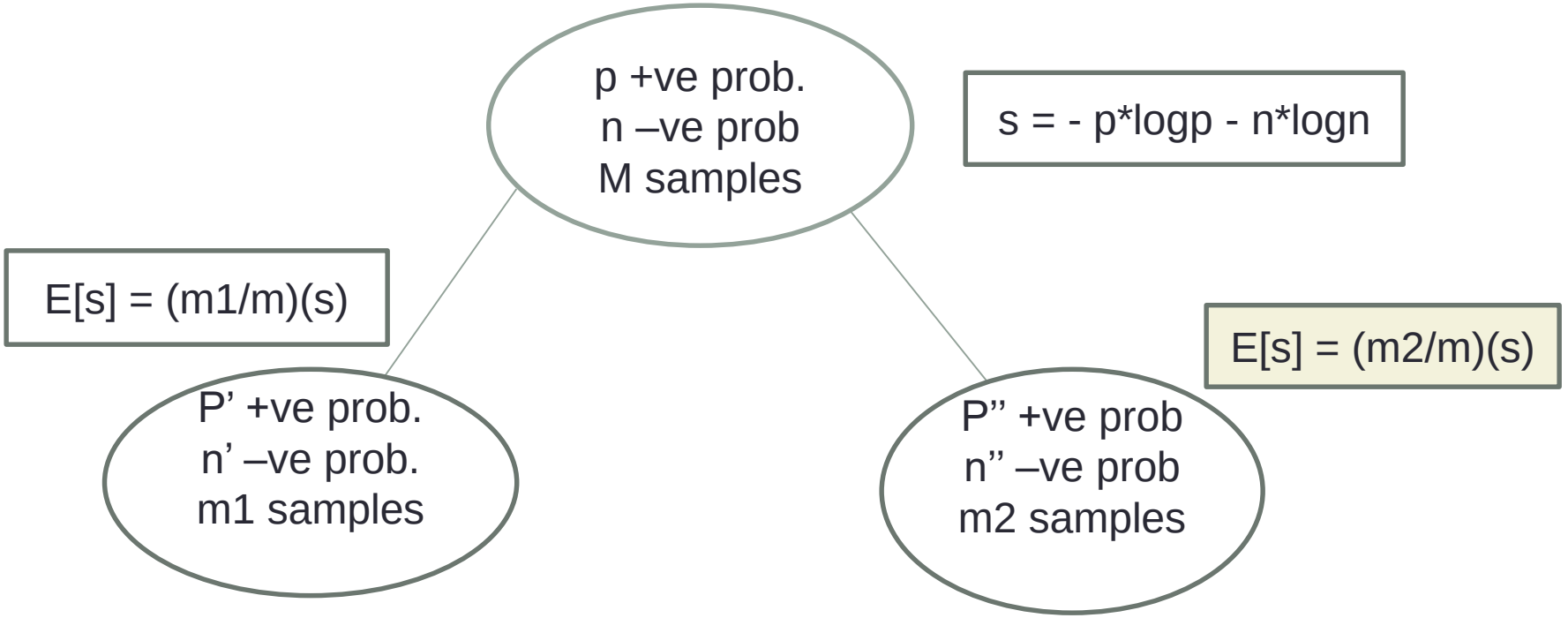
- Entropy is a measure of *unpredictability* of *information content*.

$$Entropy(s) = -\Sigma p log_2 p$$



- A split variable is selected which has maximum information gain.

$$Information\ gain = Entropy(parent) - Expected\ entropy(children)$$



- Split for continuous variables :
  - Sort data according to that variable.
  - Calculate information gain only when there is a change in class.
- Split for discrete variables:
  - Data is split for available discrete values.

Bagging Trees

- A number of decision trees are created for different bootstrap samples.
- Probabilities calculated by taking averages.

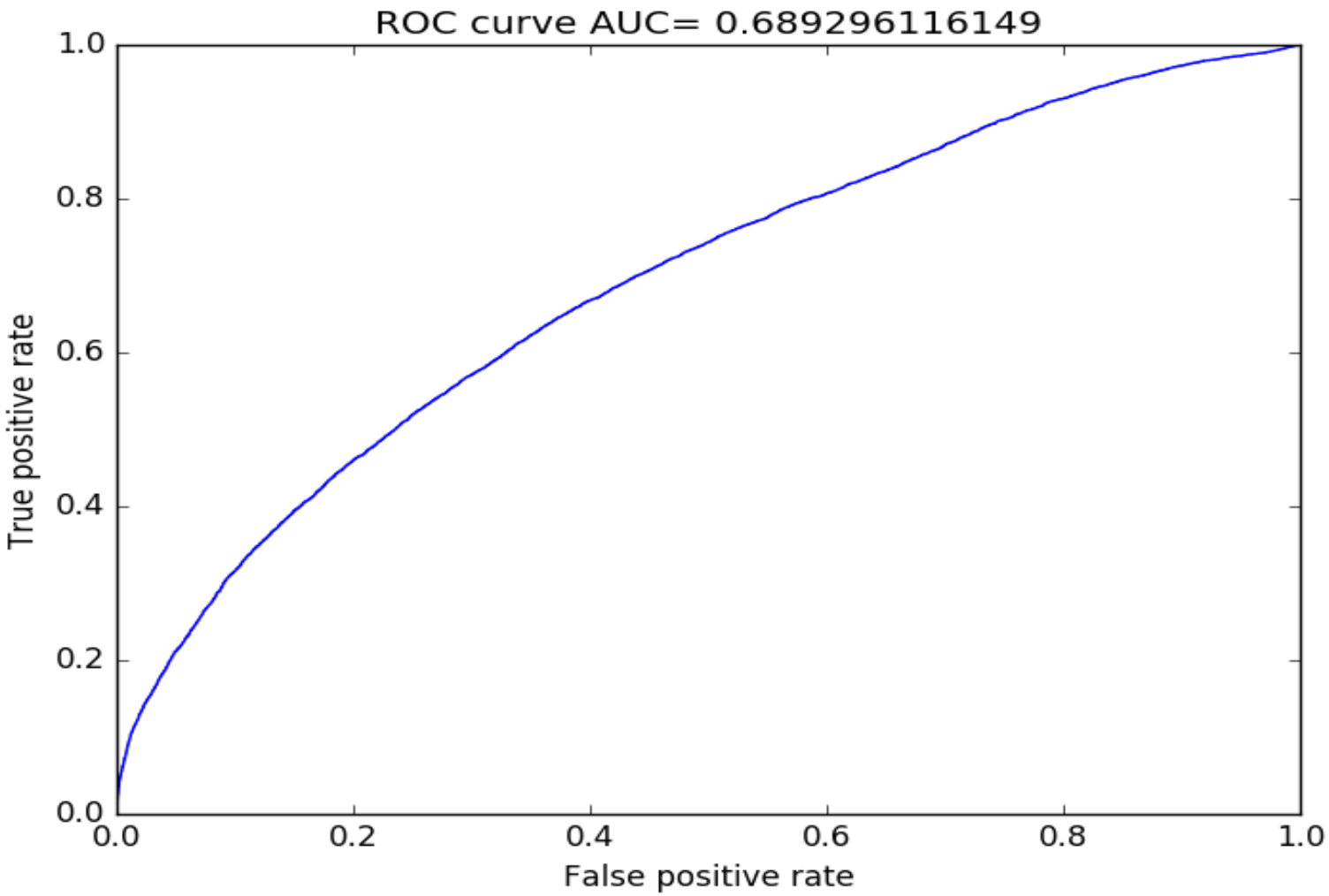
$$p(c|X) = \frac{1}{T}\sum_{i=1}^{T} p_i(c|X)$$

- This reduces the total variance.

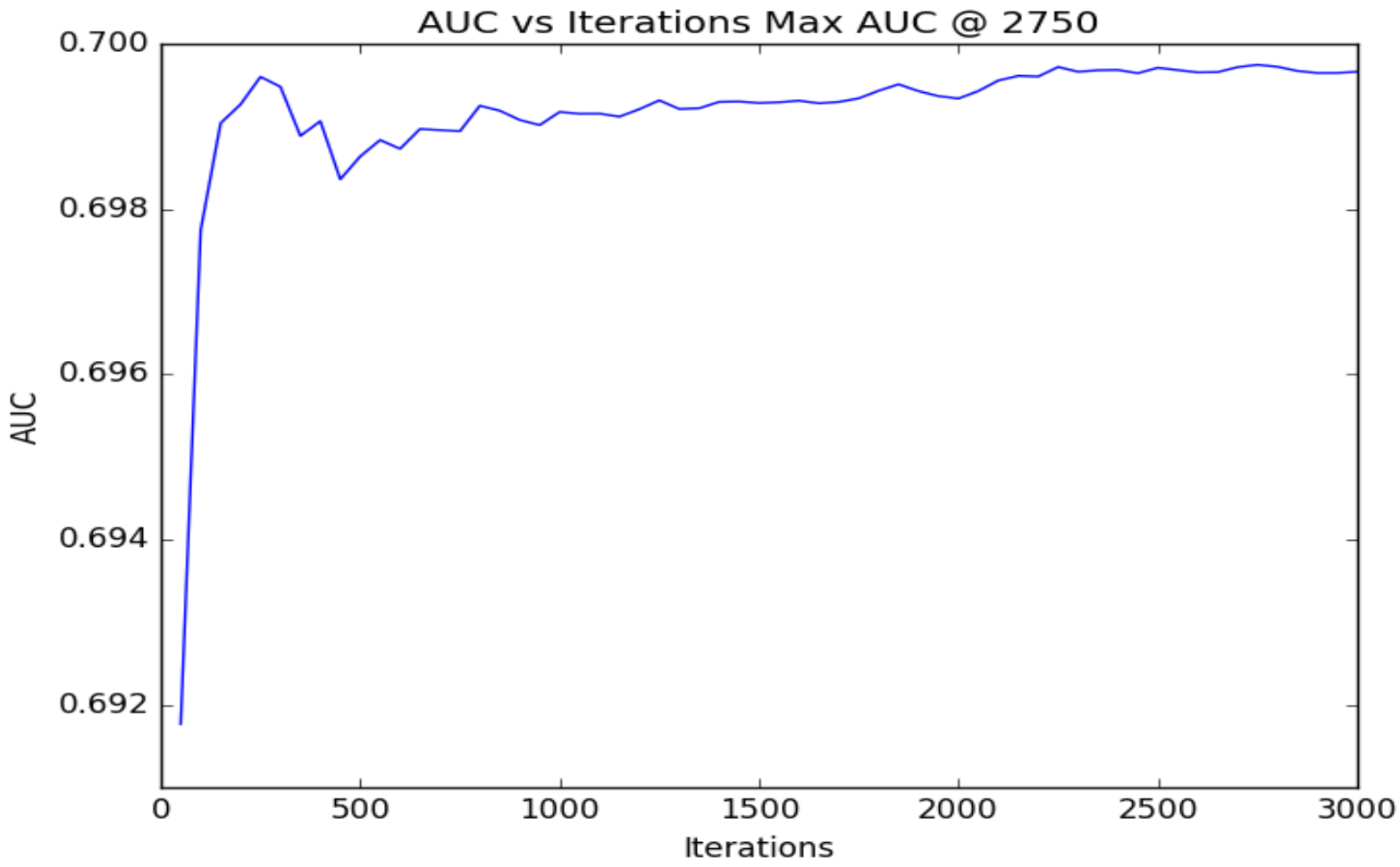
Random Forest: More randomness

- A number of decision trees created similar to bagging trees by bootstrapping samples.
- While selecting the best split variable the algorithm is given with comparatively less number of variables to choose from.
- Further reduces the variance by introducing more randomness while variable selection.

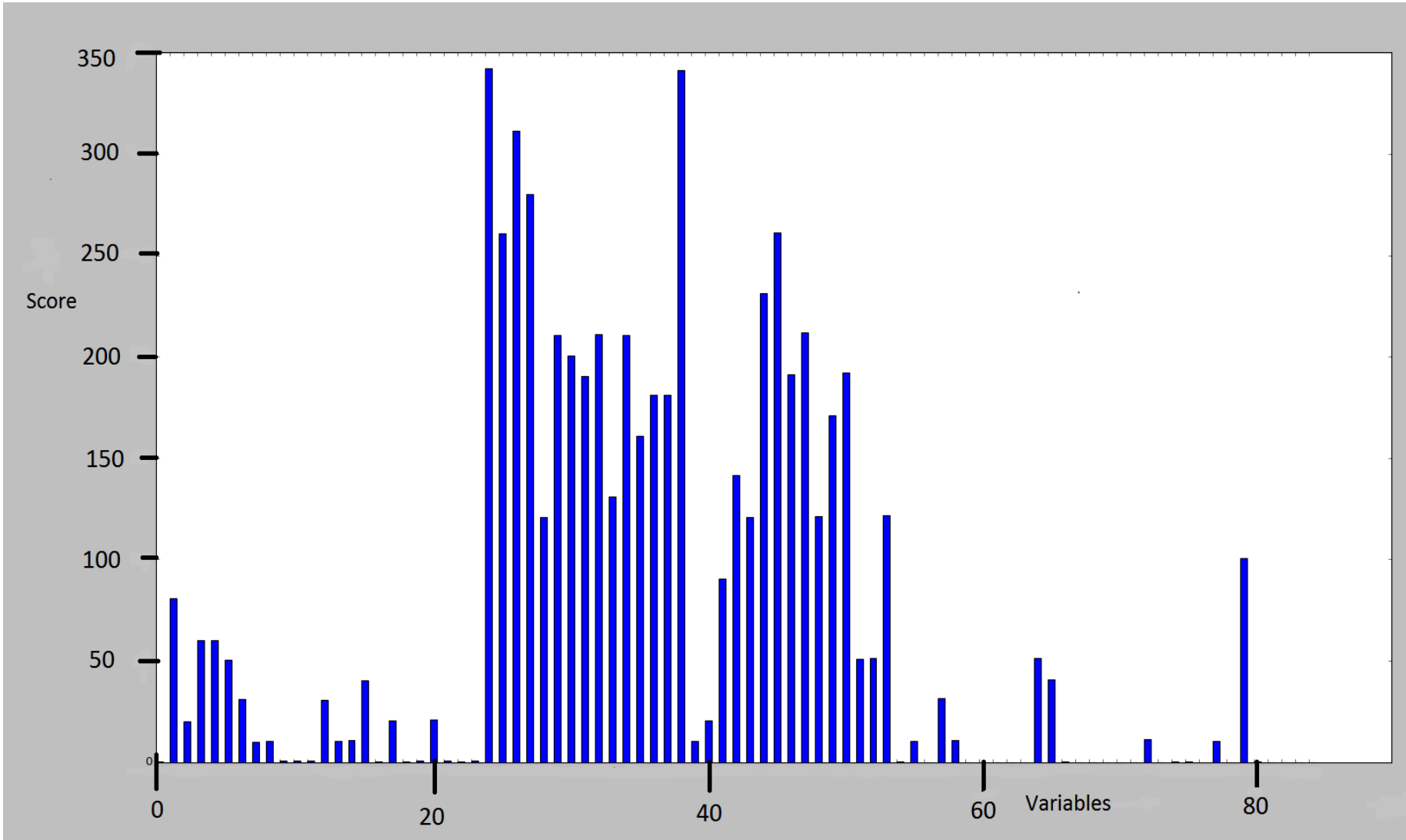
ROC Curve for 600 random trees



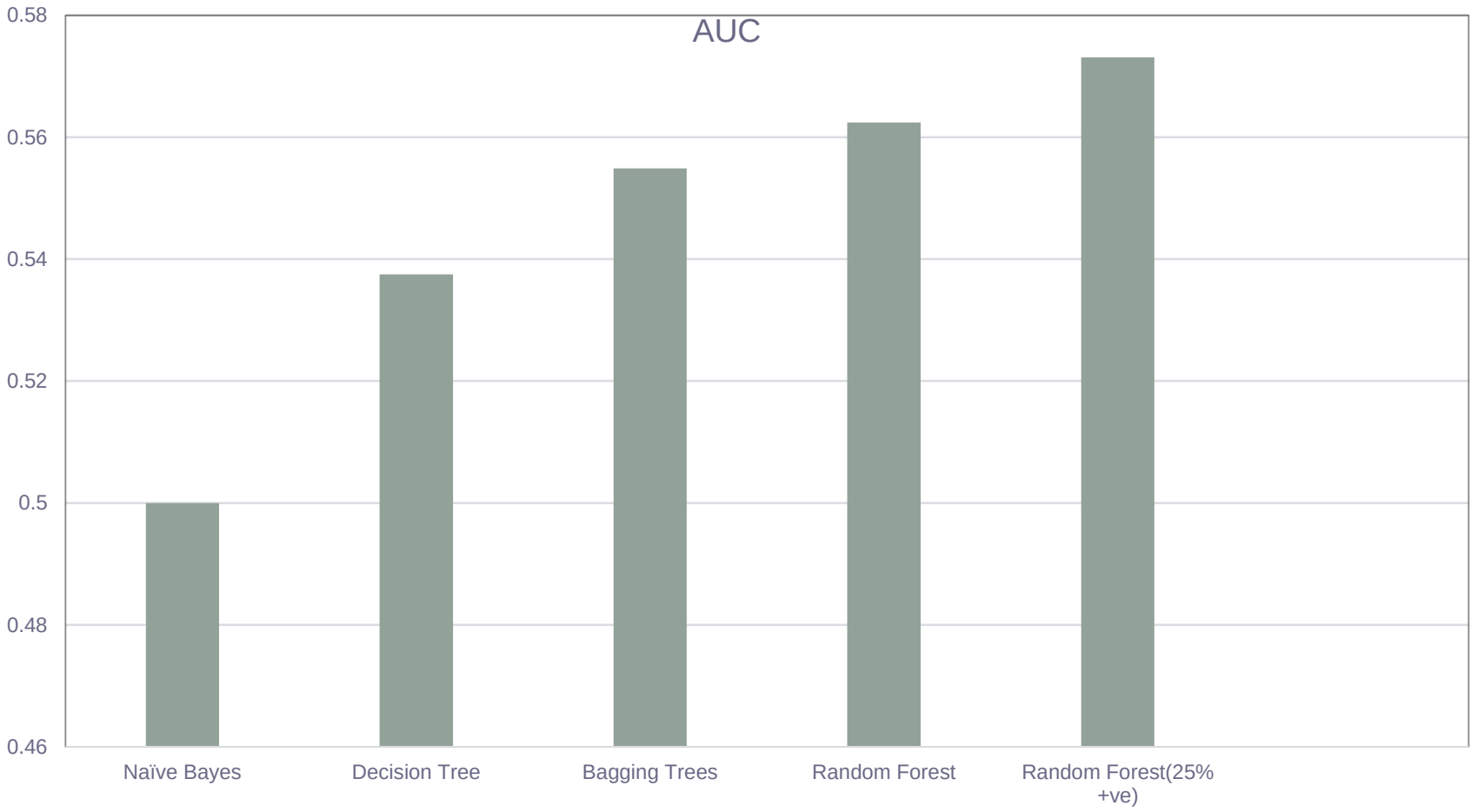
AUC vs Number of trees plot for random forest



Variable Importance Graph



Results



AUC Achieved for different algorithms

Algorithm	sample_size	num_trees	height	AUC
Naïve Bayes	-	-	-	0.50000
Decision Tree	-	1	15	0.53758
Bagging Trees	complete-datasize	600	15	0.55493
Random Forests	complete-datasize	600	15	0.56246
25% +ves Random Forest	complete-datasize	600	15	0.57313

References

[1] V. Nikulin, “On the Method for Data Streams Aggregation to Predict Shoppers Loyalty” in 2015 International Joint Conference on Neural Networks (IJCNN) pp.1-8, 12-17 July 2015, Killarney.

[2] Y. Xie, X. Li, E.W.T. Ngai, W. Ying, “Customer Churn Prediction using Improved Balanced Random Forests” in Expert Systems with Applications: An International Journal Volume 36 Issue 3, pp. 5445-5449, April 2009.

[3] M.Mehta, R. Agrawal, J. Rissanen, “SLIQ: A fast scalable classifier for data mining” in Advances in Database Technology — EDBT '96 Volume 1057 of the series Lecture Notes in Computer Science pp.18-32, June 10, 2005.

[4] L.Breiman, “Bagging Predictors” in Machine Learning, Volume 24, Issue 2, pp 123-140, August 1996.

[5] Leo Breiman, “Random forest” in machine Learning, Volume 45, Issue 1, pp5-32. Oct. 2001.

[6] J. Li, J. Lim “Using Decision Tree to Predict Repeat Customer”.[Online]. Available: http://cs229.stanford.edu/proj2015/035_report.pdf. [Accessed: Jan. 15, 2016].

[7] “Acquire Valued Shoppers Challenge” kaggle.com, April 10,2014 .[Online]. Available: www.kaggle.com/c/acquire-valued-shoppers-challenge. [Accessed: Jan. 20, 2016].


Thank you!