ACQUIRE VALUED SHOPPERS

Supervisor : Dr. Dinesh Gopalani

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

Problem Statement

- Consumer brands often offer discounts to attract new shoppers to buy their products. The most valuable customers are those who return after this initial incentive purchase.
- With enough purchase history, it is possible to predict which shoppers, when presented an offer, will buy a new item.
- However, identifying the shopper who will become a loyal buyer, prior to initial purchase, is more challenging task.
- So our task is to provide a prediction model to predict which shoppers are most likely to repeat purchase.

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

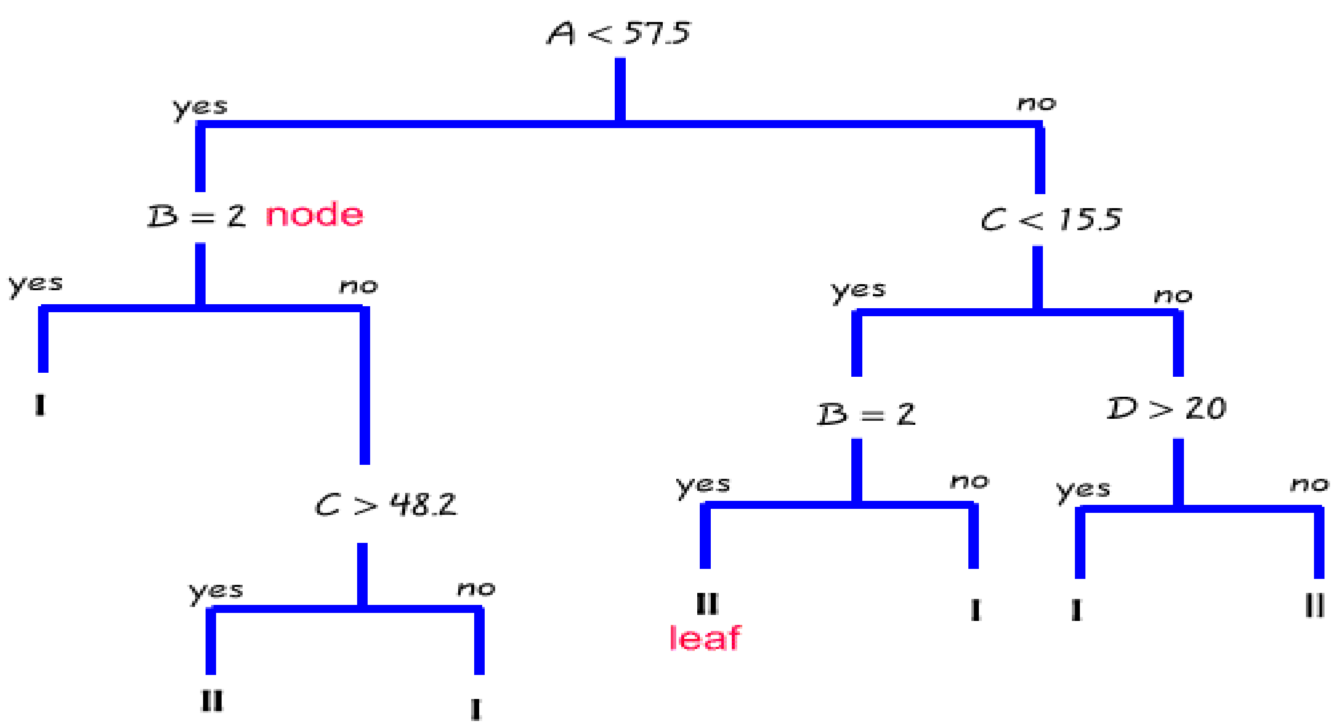
Data Files	Properties
Past Transactions	Customer ID, store, product department, product company, product category, product brand, date of purchase, product size, product size, product measure, purchase quantity, purchaseAmount
Training History	Customer ID, store, offer ID, geographical region, number 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 Trees
  - Bagging Trees
  - Random Forests

Decision Tree

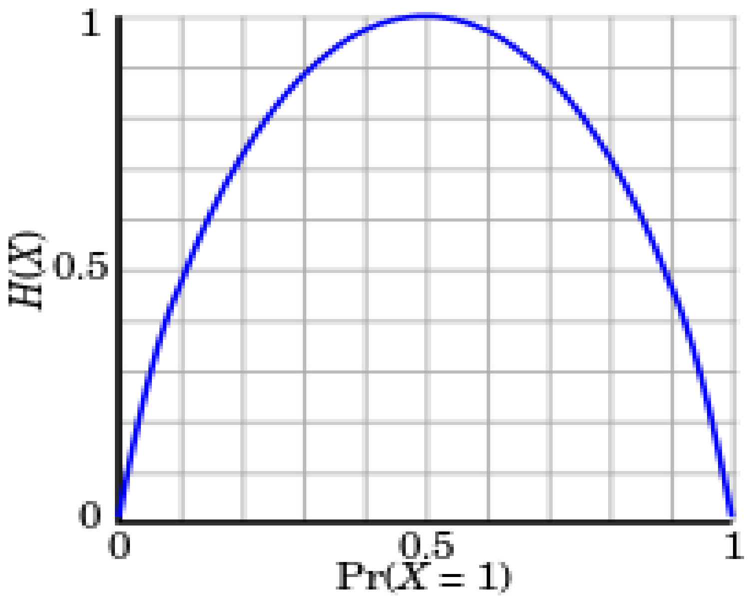
- Commonly used machine learning approach for classification and regression.
- It creates a hypothesis tree providing classification probabilities in the leaf nodes.



Splitting Criteria: Entropy

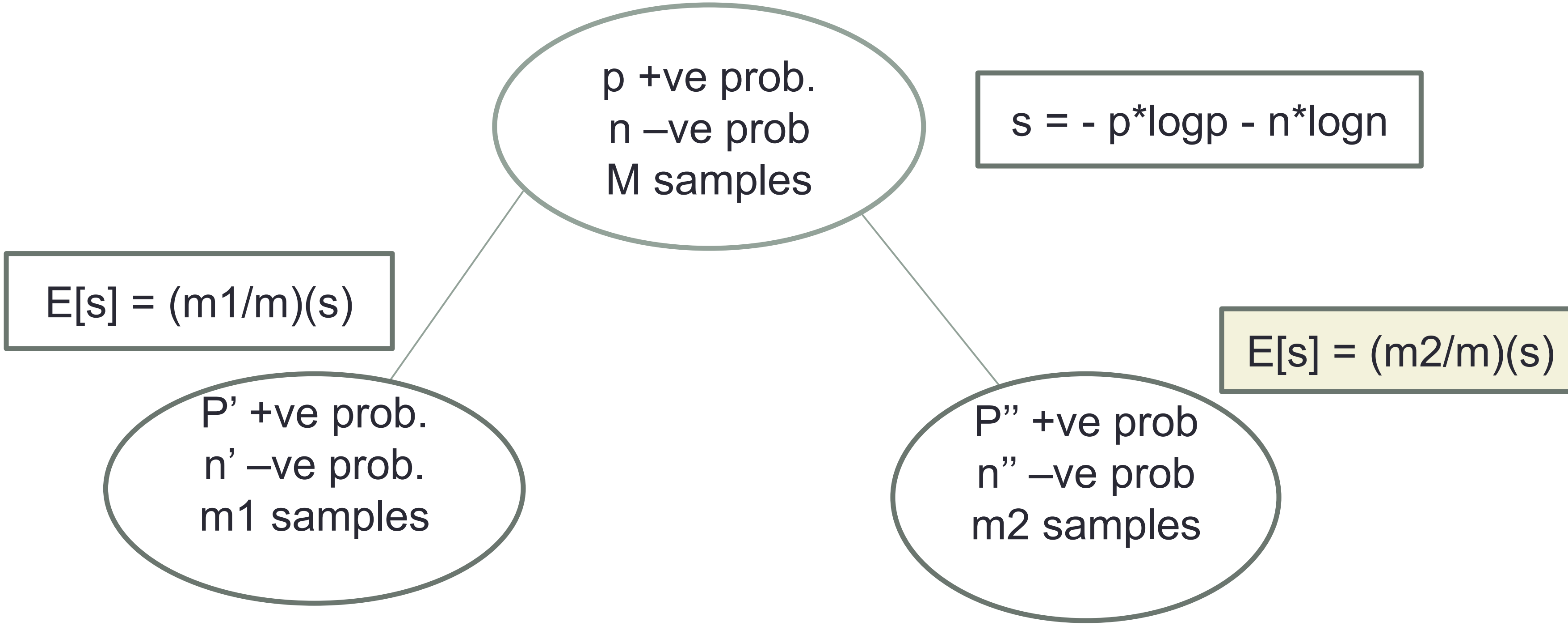
- Entropy is a measure of *unpredictability* of *information content*.

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



- 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

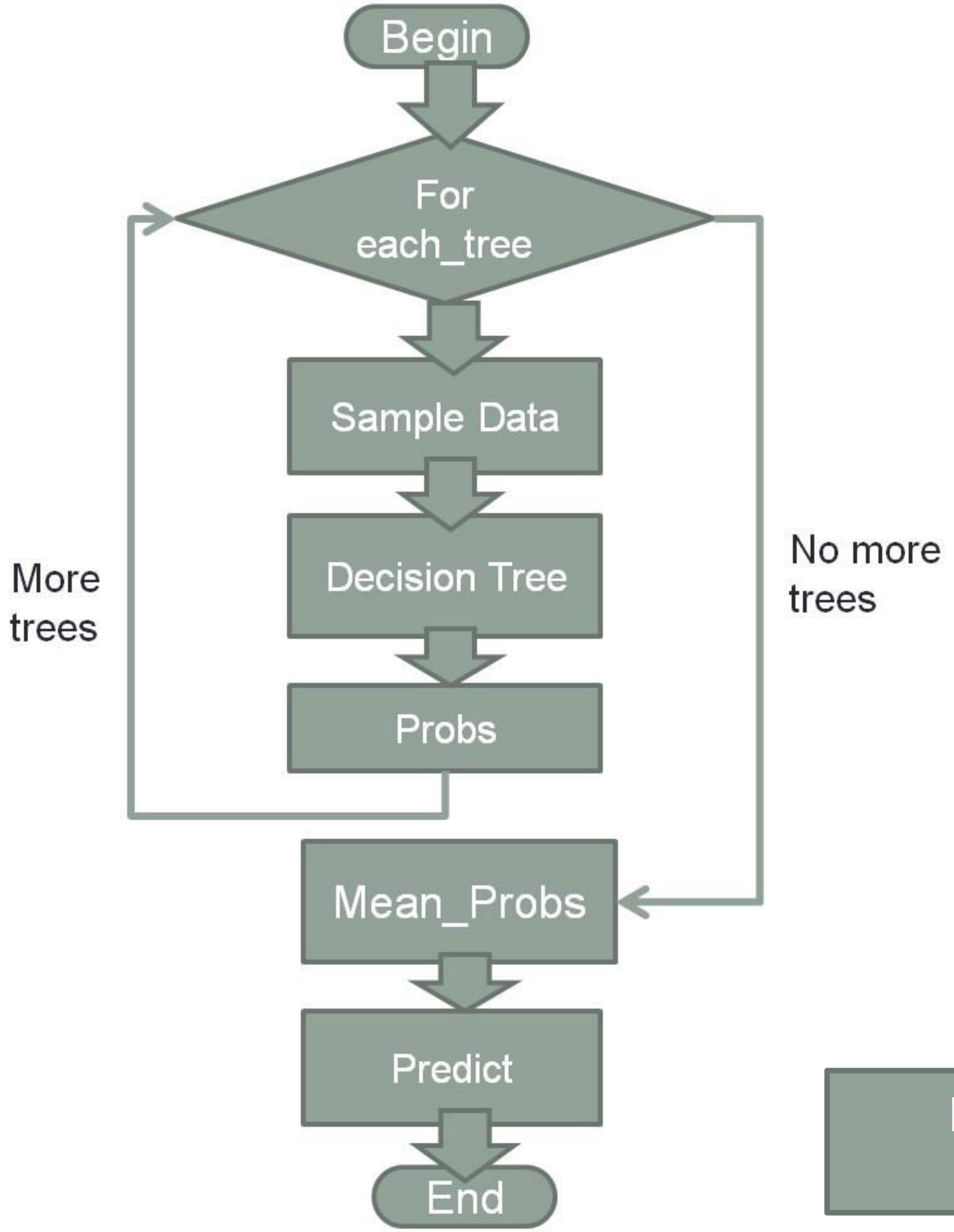
- Ensemble classifier that consist of many decision trees
- outputs the class that is the mode of the class's output by individual trees..
- Probabilities calculated by taking averages.

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

- Tries to reduce overall variance.

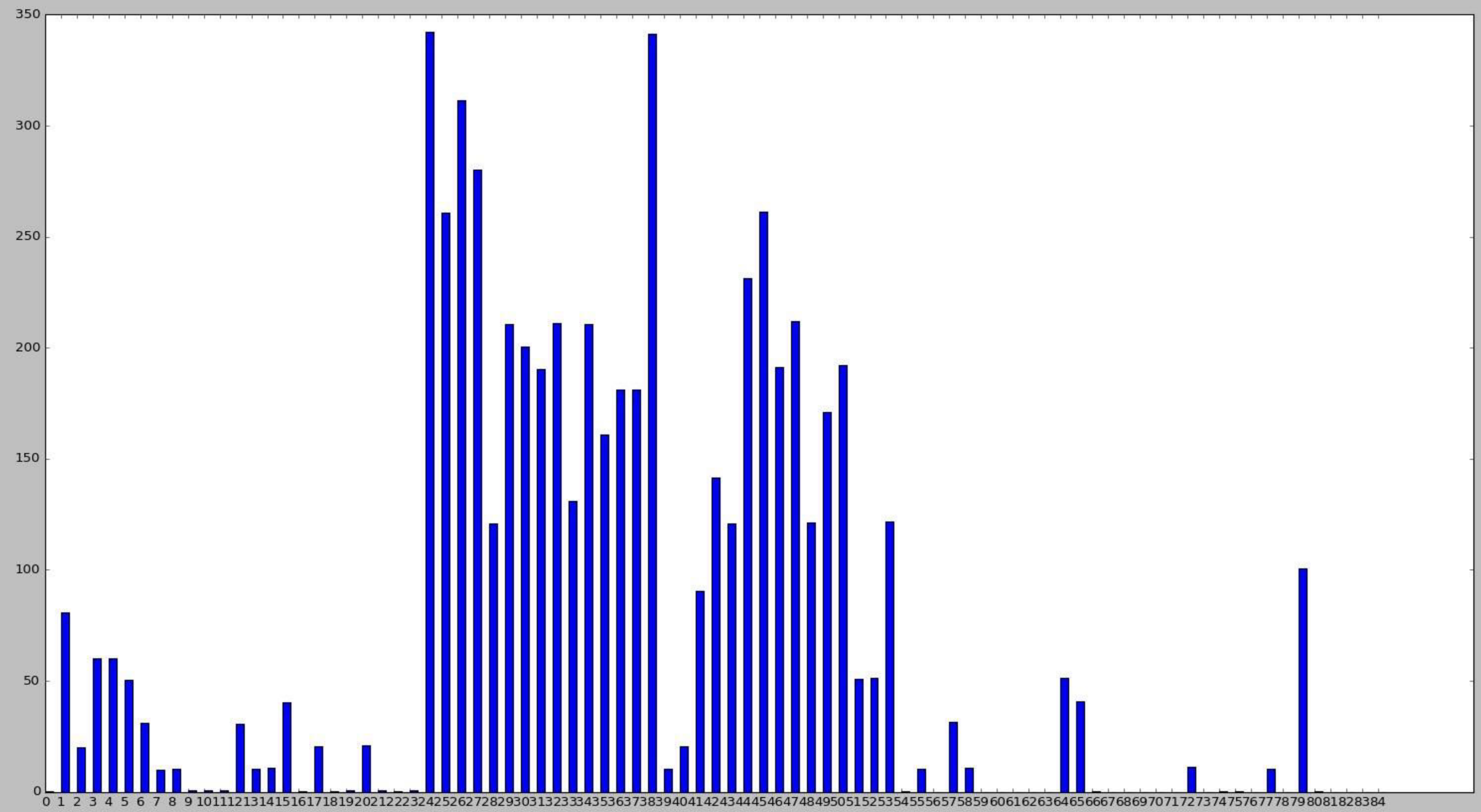
Random Forest: More randomness

- An ensemble of decision trees is created similar to bagging trees by boot-strapping samples.
- But, while selecting the best split variable the algorithm is given with comparatively less number of variables to choose from.
- Further tries to reduce the variance by introducing more randomness in variable selection.



Random Forest Algorithm

Variable Importance



Random Forest- num_trees: 600, height: 15

Naive Bayes Classification

- Naive Bayes is a simple technique for constructing classifiers: models that assign class labels to problem instances

- Naive Bayes Classification
  - Algorithm: Discrete-Valued Features
  - Algorithm: Continuous-valued Features

- Algorithm: Discrete-Valued Features

$$P(C \mid \mathbf{X}) \propto P(\mathbf{X} \mid C)P(C) = P(X_1, \cdots, X_n \mid C)P(C)$$

- Learning Phase: Given a training set **S**

$$\begin{array}{l}
\text{For each target value of } c_i \ (c_i = c_1, \cdots, c_L) \\
\quad \hat{P}(C = c_i) \leftarrow \text{estimate } P(C = c_i) \text{ with examples in } \mathbf{S}; \\
\quad \text{For every feature value } x_{jk} \text{ of each feature } X_j \ (j = 1, \cdots, n; \, k = 1, \cdots, N_j) \\
\qquad \hat{P}(X_j = x_{jk} \mid C = c_i) \leftarrow \text{estimate } P(X_j = x_{jk} \mid C = c_i) \text{ with examples in } \mathbf{S};
\end{array}$$

- Output: conditional probability tables

- Test Phase: Given an unknown instance ,

  Look up tables to assign the label c^* to $\mathbf{X}'$ if

$$[\hat{P}(a_1' \mid c^*) \cdots \hat{P}(a_n' \mid c^*)]\hat{P}(c^*) > [\hat{P}(a_1' \mid c) \cdots \hat{P}(a_n' \mid c)]\hat{P}(c), \quad c \neq c^*, c = c_1, \cdots, c_L$$

- Label = max(P(c1),P(c2)……P(cn))

- Algorithm: Continuous-valued Features
- Conditional probability often modeled with the normal distribution

$$\hat{P}(X_j \mid C = c_i) = \frac{1}{\sqrt{2\pi}\sigma_{ji}} \exp\left(-\frac{(X_j - \mu_{ji})^2}{2\sigma_{ji}^2} \right)$$

μ_{ji} : mean(avearage) of feature values X_j of examples for which $C = c_i$

σ_{ji} : standard deviation of feature values X_j of examples for which $C = c_i$

- Learning Phase:

  Output: normal distributions and

- Test Phase: Given an unknown instance
  - Instead of looking-up tables, calculate conditional probabilities with all the normal distributions achieved in the learning phrase

Results

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	500	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

[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, April, 2009 pp. 5445-5449

[3] J. Li, J. Lim “Using Decision Tree to Predict Repeat Customer”

[4] 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

[5] Acquire Valued Shoppers Challenge “www.kaggle.com/c/acquire-valued-shoppers-challenge”


Thank you!