Credit Risk Model Building

Venkat Reddy

Note

- This presentation is just the lecture notes for the talk on Credit Risk Model building conducted for MBA students
- The best way to treat this is as a high-level summary; the actual session went more in depth and contained other information.
- Most of this material was written as informal notes, not intended for publication
- Please send your questions/comments/corrections to venkat@trenwiseanalytics.com or 21.venkat@gmail.com
- Please check my website for latest version of this document

-Venkat Reddy

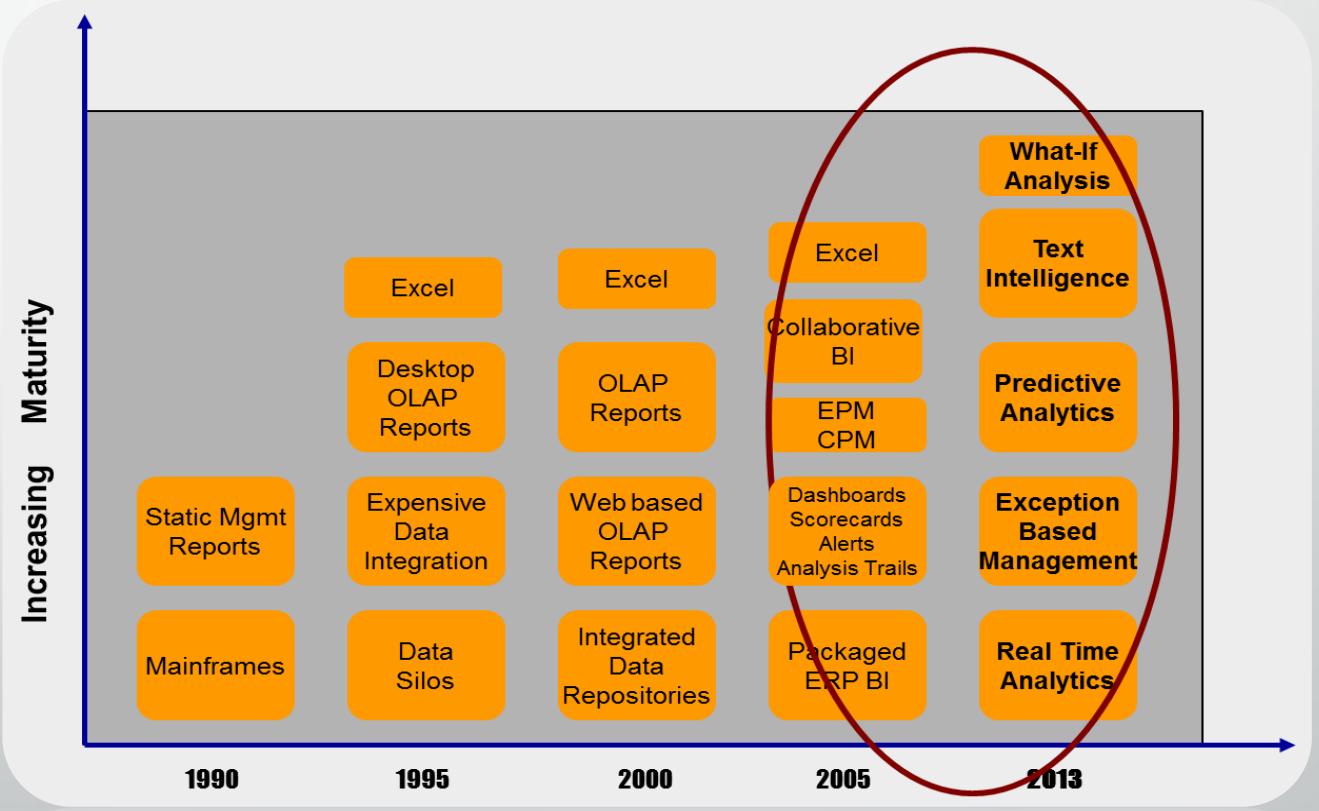
Contents

- Applications of Statistics in Business
- The Model Building Problem
- Credit Risk Model Building
- Other Applications of Model Building

Applications of Statistics in Business

Increasingly, business rely on intelligent tools and techniques to analyze data systematically to improve decision-making.

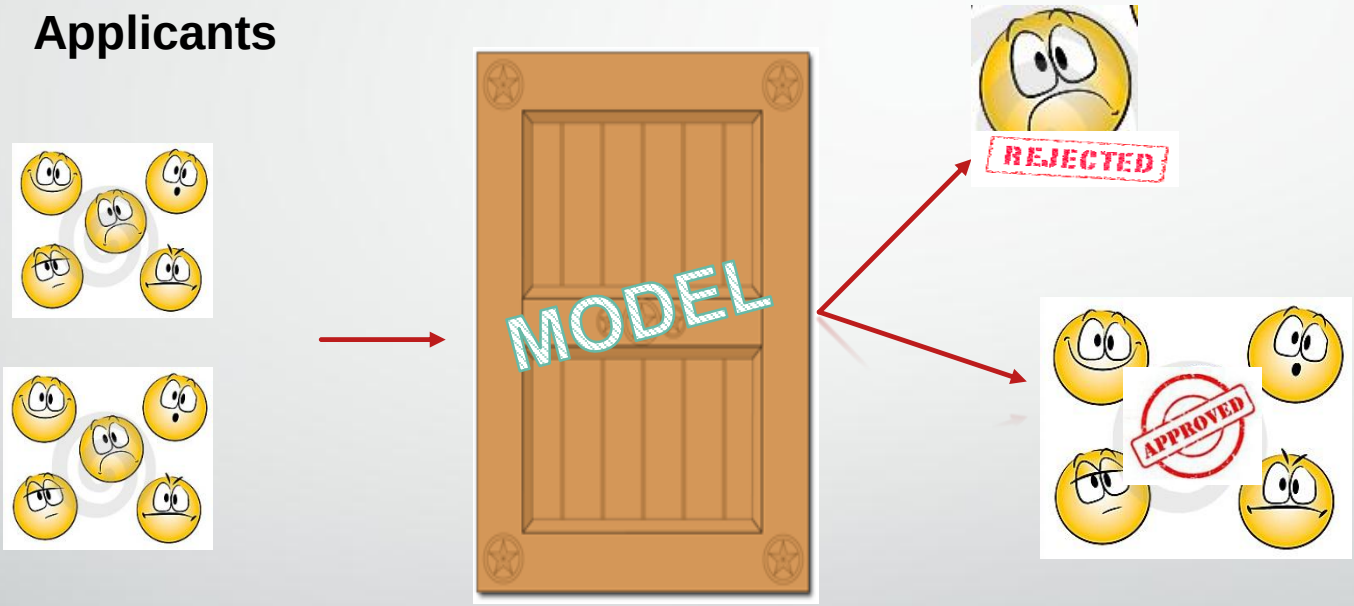
- Retail sales analytics
- Financial services analytics
- Telecommunications
- Supply Chain analytics
- Transportation analytics
- Risk & Credit analytics
- Talent analytics
- Marketing analytics
- Behavioral analytics
- Collections analytics
- Fraud analytics
- Pricing analytics



The Problem
Who will run away with my money?

- **Citi Bank :** Present in more than 90 countries.
- More than **100,000** customers apply for credit cards/loans every month
- All of them have different **characteristics**
- Out of 100,000 customers, who all have the higher probability of default/ Charge off?
- Basically, who will run away with my money?
- We need to predict the probability of "Running away"
- Who all have 'Gupta Bank' credit card applications in this room?

Bank builds a model that gives a score to each customer



"Developing set of equations or mathematical formulation to forecast future behaviors based on current or historical data."

Predictive Modeling
Lets try to understand predictive modeling

Predictive Modeling – Fitting a model to the data to predict the future.

- Predicting the future –and it is so easy some times
- Who is going to score more runs in IPL-2013?
- That's it you predicted the future..
- BTW how did you predict?
- Predicting the future based on historical data is nothing but Predictive modeling

Predictive Modeling
Lets try to understand predictive modeling

Predictive Modeling – Fitting a model to the data to predict the future.

- Who is going to score more runs in IPL 2013?
- Predicting the future …well it is not that easy …

Predictive Modeling
Horse Race Analogy

How to bet on best horse in a horse race

The Historical Data
Win vs. Loss record in past 2 years

- **Long legs:** 75% (Horses with long legs won 75% of the times)
- **Breed A**: 55%, **Breed B**: 15 % **Others** : 30%
- **T/L (Tummy to length) ratio** <1/2 :75 %
- **Gender**: Male -68%
- **Head size:** Small 10%, Medium 15% Large 75%
- **Country**: Africa -65%

Given the historical data
Which one of these two horses would you bet on?

	Kalyan	Chethak
Length of legs	**150 cm**	110 cm
Breed	**A**	F
T/L ratio	**0.3**	0.6
Gender	**Male**	Female
Head size	**Large**	Small
Country	India	India

Given the historical data
Which one would you bet on….now?

	Kalyan	Chethak
Length of legs	110 cm	**150 cm**
Breed	C	**A**
T/L ratio	**0.45**	0.60
Gender	**Male**	Female
Head size	Small	**Large**
Country	**Africa**	India

Given the historical data
What about best one in this lot?

	Horse-1	Horse-2	Horse-3	Horse-4	Horse-5	Horse-6	Horse-7	Horse-8	Horse-9	Horse-10
Length of legs	109	114	134	130	149	120	104	117	115	135
Breed	C	A	B	A	F	K	L	B	C	A
T/L ratio	0.1	0.8	0.5	1.0	0.3	0.3	0.3	0.6	0.7	0.9
Gender	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
Head size	L	S	M	M	L	L	S	M	L	M
Country	Africa	India	Aus	NZ	Africa	Africa	India	India	Aus	Africa

Citi has a similar problem?
Who is going to run away with my money?

- Given Historical of the customers we want to predict the probability of bad
- We have the data of each customer on
  - Customer previous loans, customer previous payments, length of account credit history, other credit cards and loans, job type, income band etc.,
- We want to predict the probability of default

Credit Risk Model Building

Four main steps

1. Study historical data
   - What are the causes(Customer Characteristics)
   - What are the effects(Charge off)
2. Identify the most impacting factors
3. Find the exact impact of each factor(Quantify it)
4. Use these coefficients for future

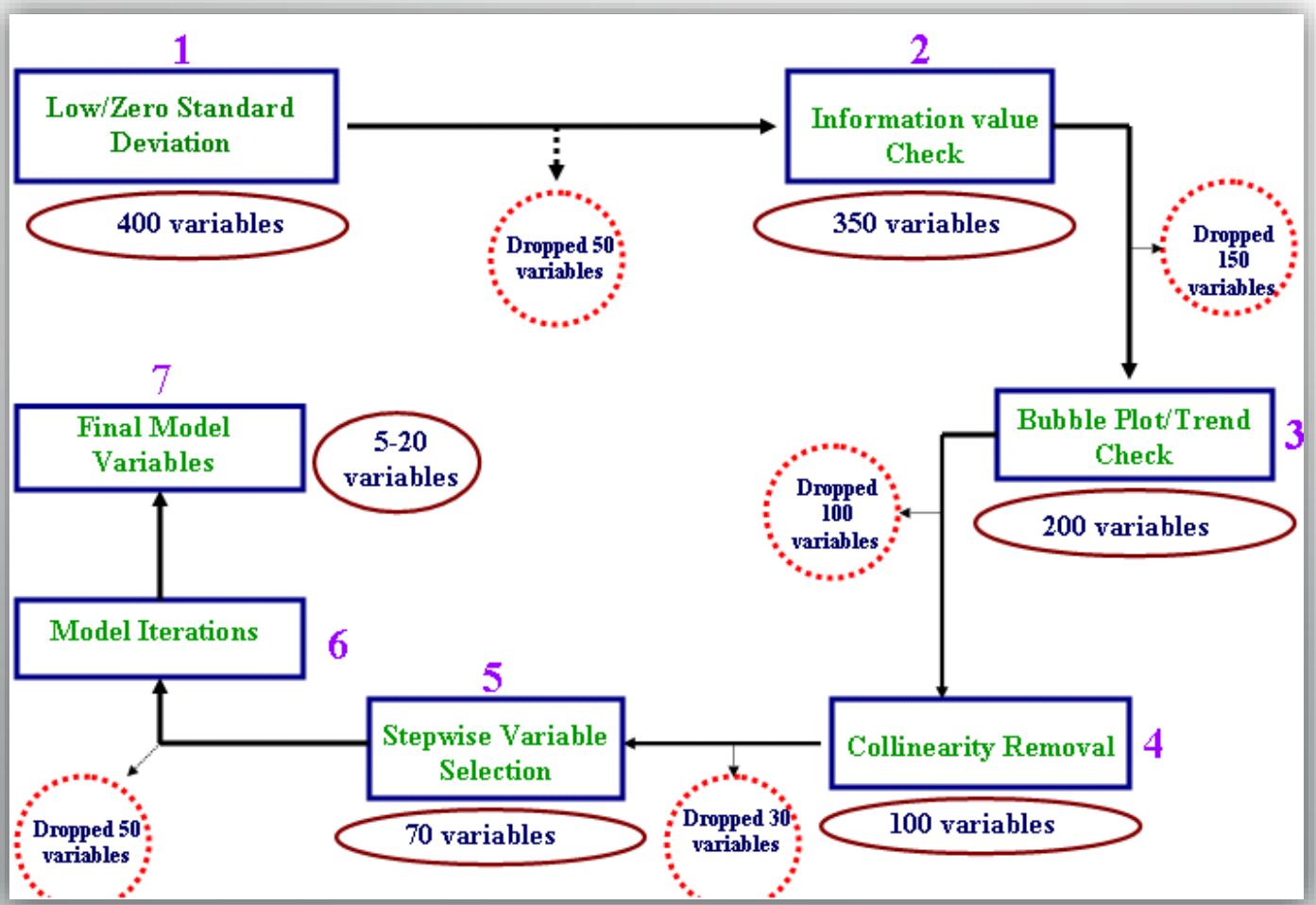
The Historical Data of Customers

- Contains all the information about customers
- Contains information across more than 500 variables
- Portion of data is present in the application **form**
- Portion of the data is available with **bank**
- Lot of data is maintained in **bureau**
  - Social Security number –in US
  - PAN Number in India

Attribute	Value
SSN	*111259005*
Age	*27*
Number of dependents	*2*
Number of current loans	*1*
Number of credit cards	*1*
Number Installments 30days late in last 2 years	*4*
Average utilization % in last 2 years	*30%*
Time since accounts opened	*60 months*
Number of previous applications for credit card	*2*
Bankrupt	***NO***

Identifying most impacting factors

Out of 500 what are those 20 important attributes



Model Building

logistic regression model to predict the probability of default

- Probability of bad = w1(Var1)+w2(Var2) +w3(Var1)……+w20(var20)
- Logistic regression gives us those weights
- Predicting the probability
  - **Probability of bad**=0.13(number of cards)+ 0.21 (utilization)+…….+0.06(number of loan applications)
- That's it ….we are done

Credit Risk Model Building Real-time Example

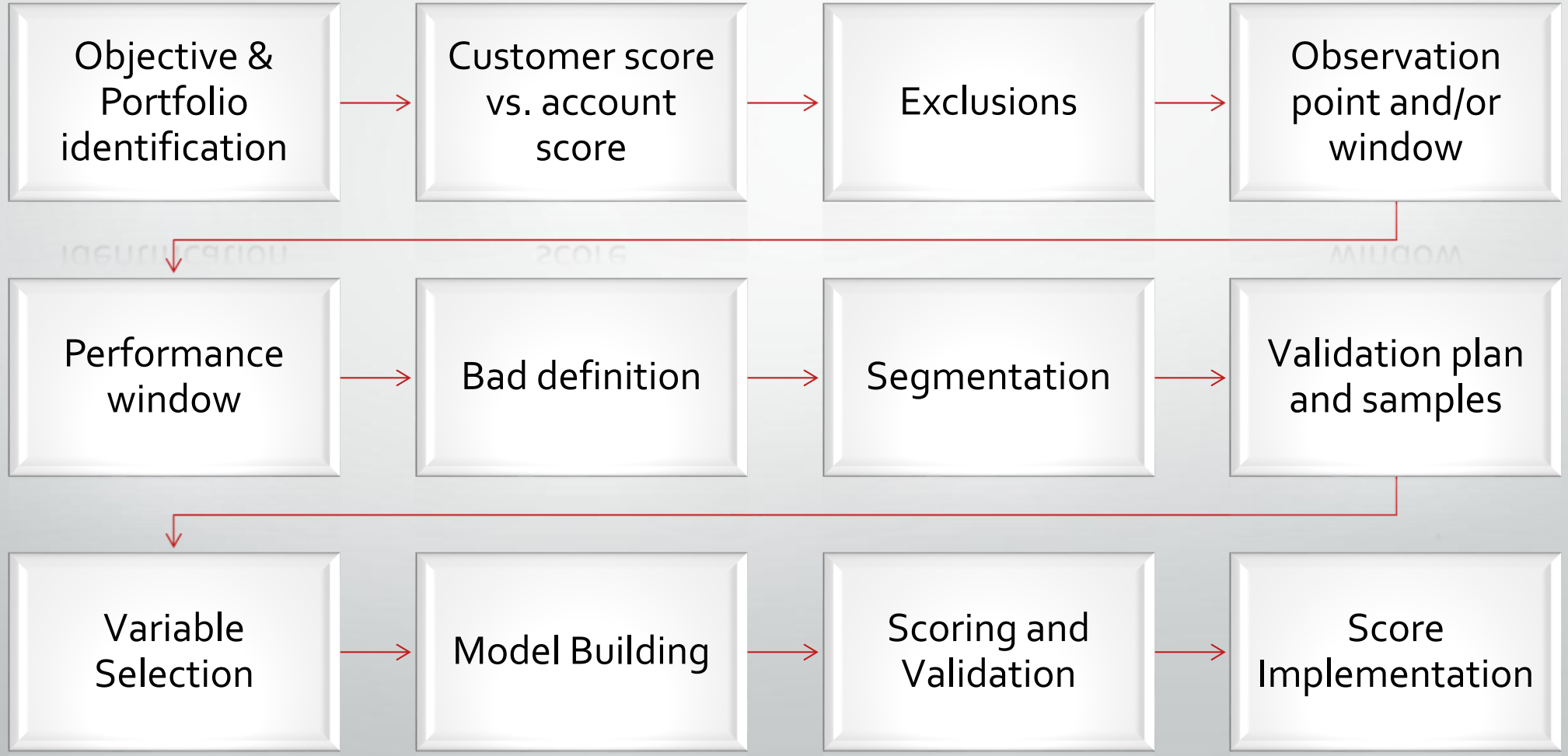
Attributes used on the model

1. MonthlyIncome
2. Number of loans
3. Number of times 30days late in last 2 years
4. Utilization in last 2 years
5. Age
6. DebtRatio – Monthly Debt / Monthly Income

‘Gupta Bank’ Credit Cards Approval

- Lets use above model for gupta bank
- Who are the applicants here?
- Lets get the bureau information for the applicants

Actual model Building Steps



Marketing Example

Predicting the response probability to a marketing Campaign

- Selling Mobile phones – Marketing campaign
- Who should we target?
  - Consider historical data of mobile phone buyers
  - See their characteristics
  - Find top impacting characteristics
  - Find weight of each characteristic
  - Score new population
  - Decide on the cut off
  - Try to sell people who score more than cut off

Other Applications of Model Building

- **Fraud transactions scorecard** – Fraud identification based on attributes like transaction amount, place, time, frequency of transactions etc.,
- **Attrition modeling** – Predicting employee attrition based on their characteristics

Thank You

Venkat Reddy

21.venkat@gmail.com