3rd Seminar on Data Science and Analytics
Bangalore
July 27th, 2019

Use cases of Financial Data Science Techniques in Retail Analytics

SUDIPTO PAL
Staff Data Scientist, Walmart Labs India

Survival Models

Institute of Actuaries of India

Two kinds of Survival Models

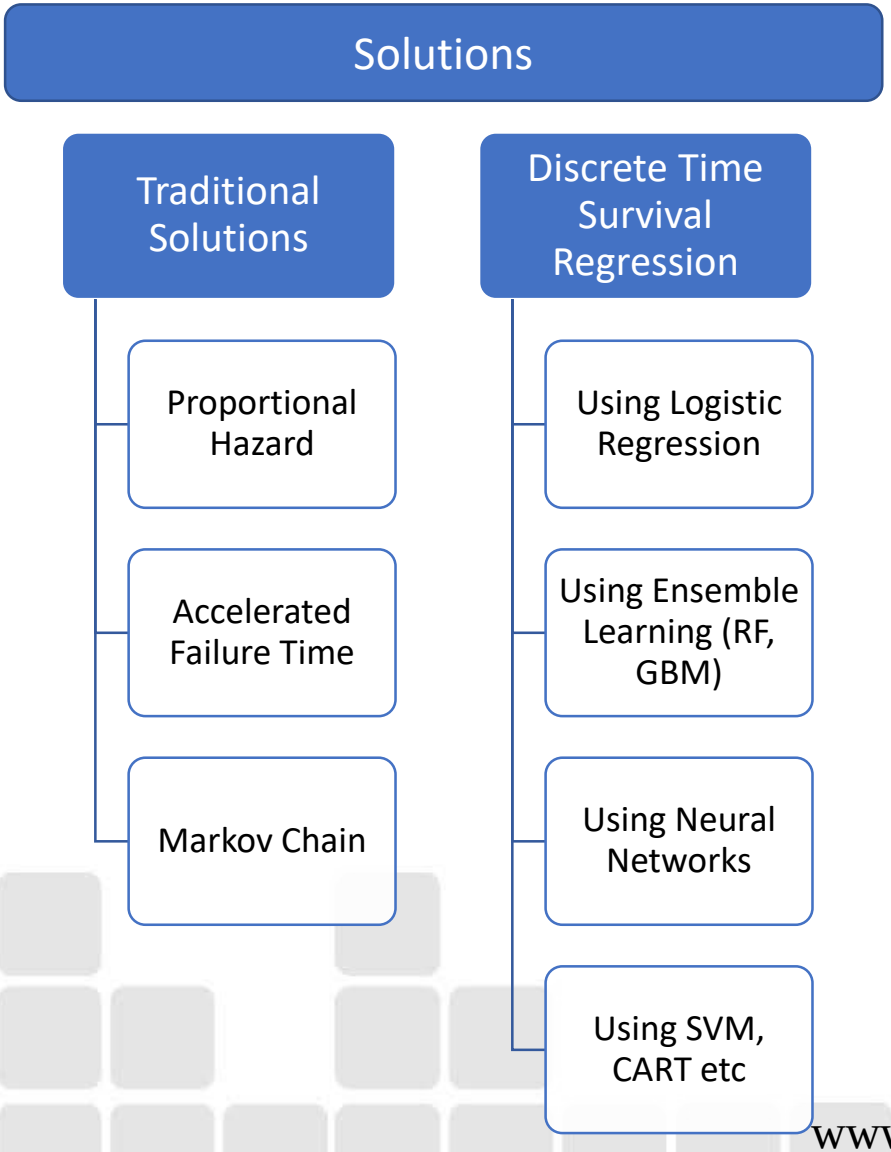
Event Model:

- modeling probability of surviving beyond a time threshold
- Example: Attrition Probability
- Popular technique: Cox Regression

Waiting time model:

- modeling waiting time till an event
- Example: Attrition Timing
- Popular Technique: AFT

Survival Models



Survival Models: beyond mortality in Insurance

- Lapse
- Survival of reserves in non-life

Customer Lifetime Value & Attrition

In insurance:

- $CLTV = \sum_{t=1}^{\infty} e^{-rt} \times P(A \geq t) \times [P(S \geq t) \times I_t \ + P(S = t) \times D_t]$

S= Remaining lifetime after time 0
A = Time to attrition
I_t = Net income at time t if survives time t (usually positive)
D_t= Net income due to death at time t (usually negative)

Insurance:
Mortality and Lapse both important
Long tenure
Traceable customer
Death is usually reported

Retail:
Lapse contains mortality
Short Tenure (not really a lifetime)
Not 100% traceable customers
Death is untraceable

Customer Value in Retail

- Structural CLTV approach:
  - Survival model and a Money model
- Distribution based CLTV: Buy Till You Die (BTYD)
  - Two distributions calibrated from customer's own data
    - Probability of Being Alive
    - Probability of a purchase
  - Pareto/NBD, Beta-Geometric(BG)/NBD
- RFM or Recency Frequency Money
  - Customer ~~Lifetime~~ Value

Survival Models: beyond Cox & AFT

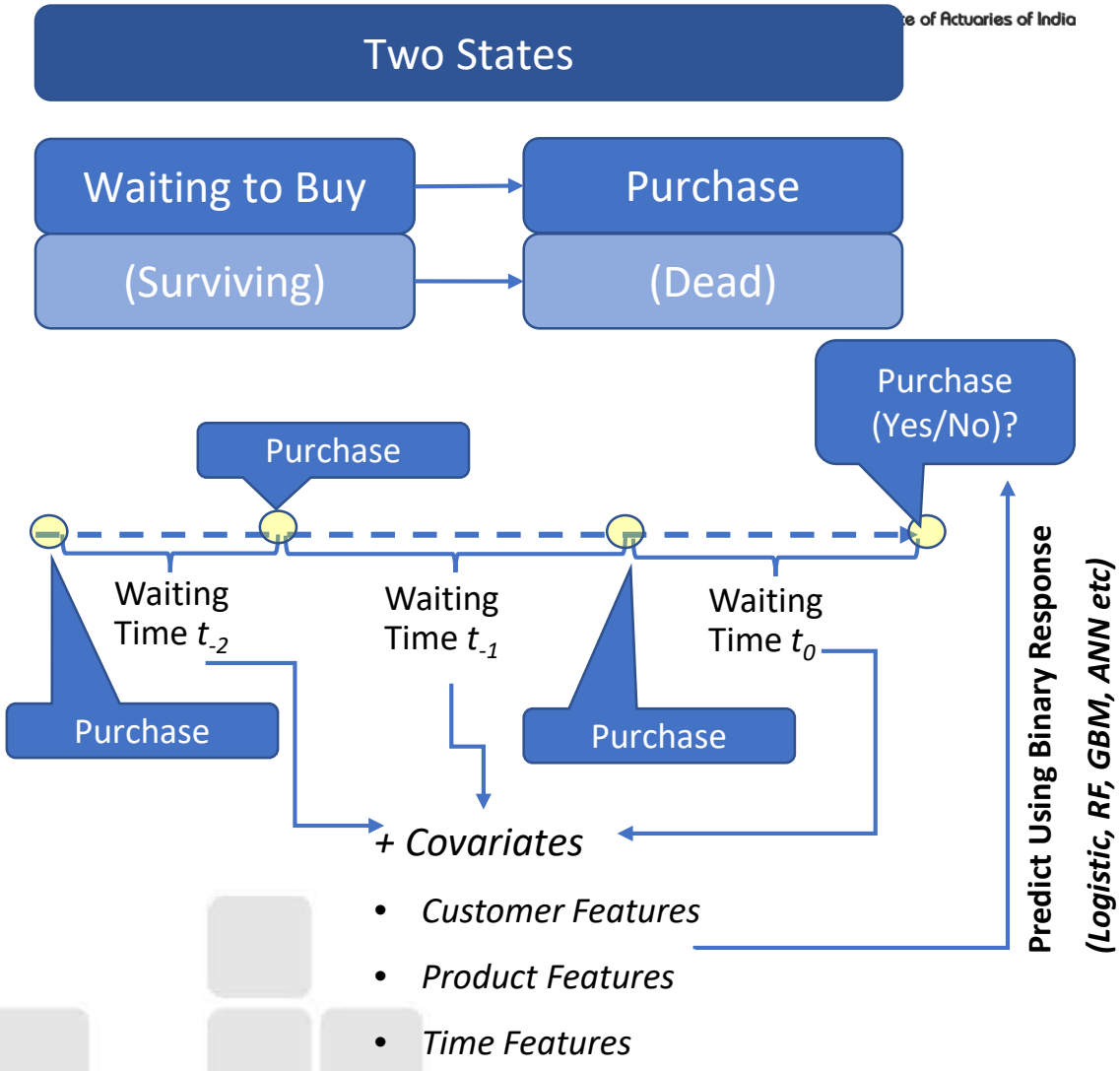
- Discrete time survival regression:
  - GLM based (Logistic/Poisson)
  - Ensemble Learning (Random Forest/ GBM)
  - Others (SVM/ANN)

- Advantages:
  - No nonparametric estimation for baseline hazard
  - Handling of high dimensionality (Regularization/Ensemble)

Survival Models for Product Recommendation

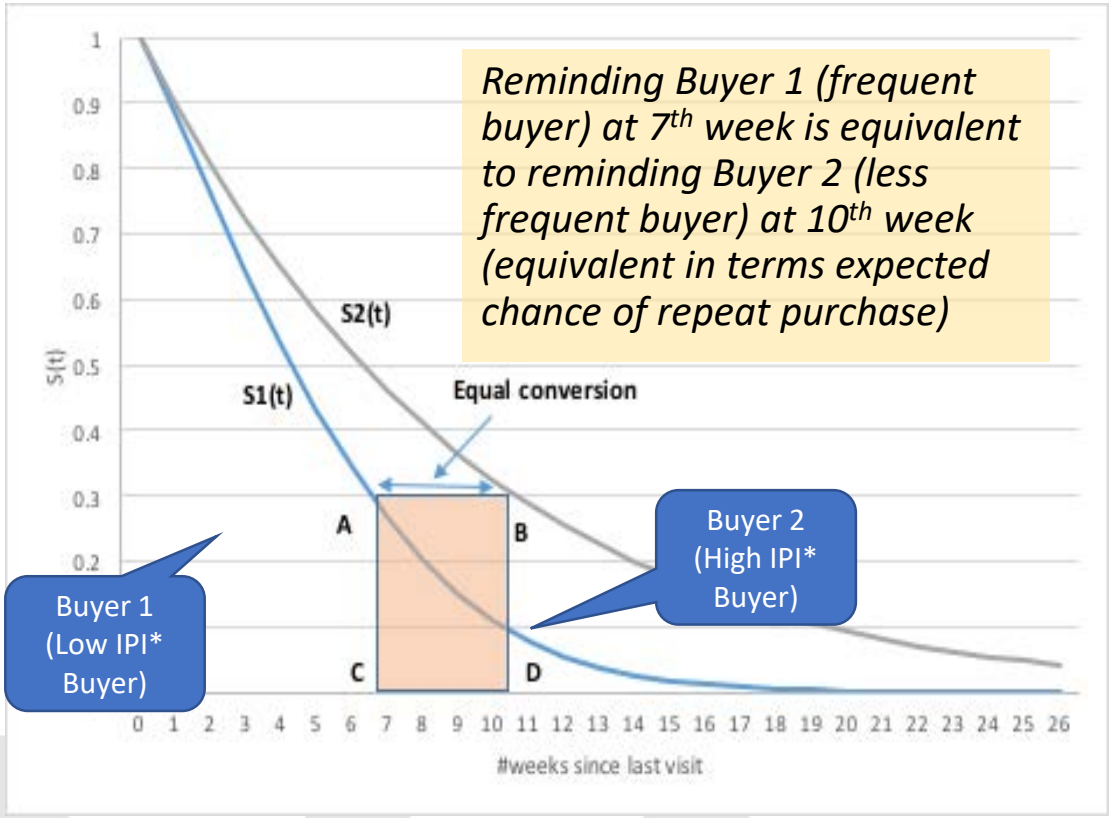
Product recommendation:

- Waiting time: Time between two purchase for repeating buyers
- A purchase is an event which is a function of how long the customer is waiting in a non-purchase state and other covariates (product features, customer features and time features)



Survival Models for Product Recommendation

Product recommendation as a waiting process until the customer's natural comeback time overshoots



*IPI= Inter Purchase Interval; S(t)= P[waiting time > t]

S(t)= Survival Probability i.e. Pr(Wait for customers come back continues beyond t) ::: Strictly non-increasing function of t

Survival Models: other use cases

- Inventory management:
  - Next demand generation as a function of time since previous demand generation
  - Next supply as a function of time since previous supply
- Queuing theory

The way we model: differences

Institute of Actuaries of India

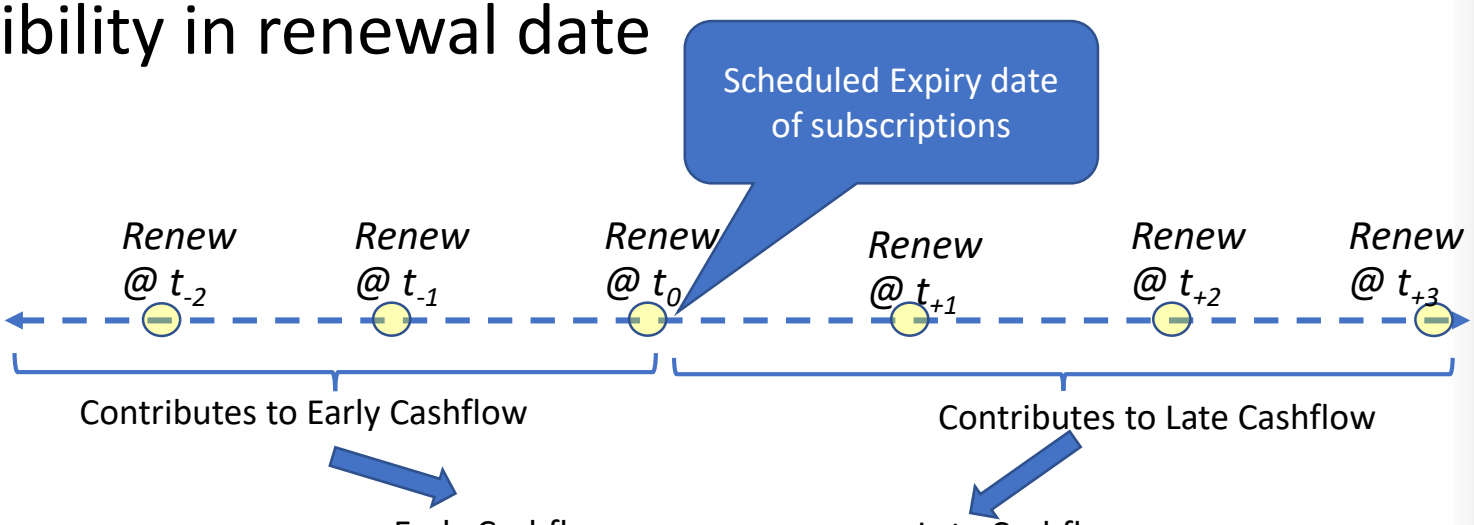
Insurance (Pricing and Reserving)	Insurance (Marketing)	Retail Analytics
Transparency for regulators	No role of regulators	No role of regulators
Explainability for decision makers	Explainability subjective	Explainability subjective
Whitebox models	Blackbox/ Whitebox both accepted	Blackbox/ Whitebox both accepted

Cashflow models (Examples)

- An annual subscription based business offering flexibility in renewal date

Renewal Probabilities at different time thresholds modeled using Survival Models

Multiplied with Subscription Amount and Aggregated



Realized Renewals

Predicted Renewals

Original Expiry Month of Subscriotion	Maturity -2	-1	On Expiry Month	+1	+2	+3
Jan-2019	1.30%	9.50%	36.00%	7.70%	3.20%	1.90%
Feb-2019	0.60%	8.10%	33.40%	7.30%	2.80%	1.90%
Mar-2019	0.80%	9.10%	33.20%	6.40%	2.90%	1.60%
Apr-2019	1.00%	8.60%	35.00%	7.50%	3.10%	1.80%
May-2019	1.00%	9.40%	34.90%	7.30%	2.80%	1.80%
Jun-2019	0.70%	8.50%	35.20%	7.10%	3.00%	1.90%

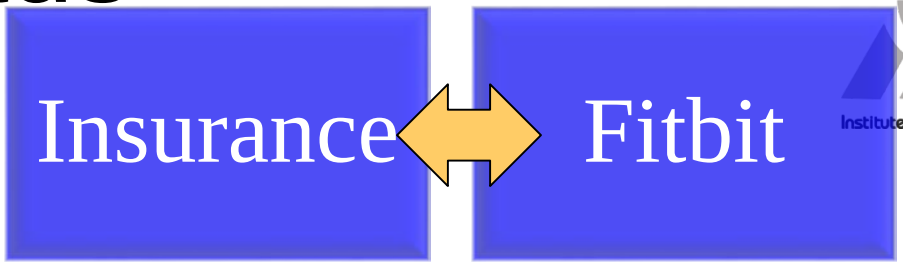
Original Expiry Month

Cashflow Models: main differences

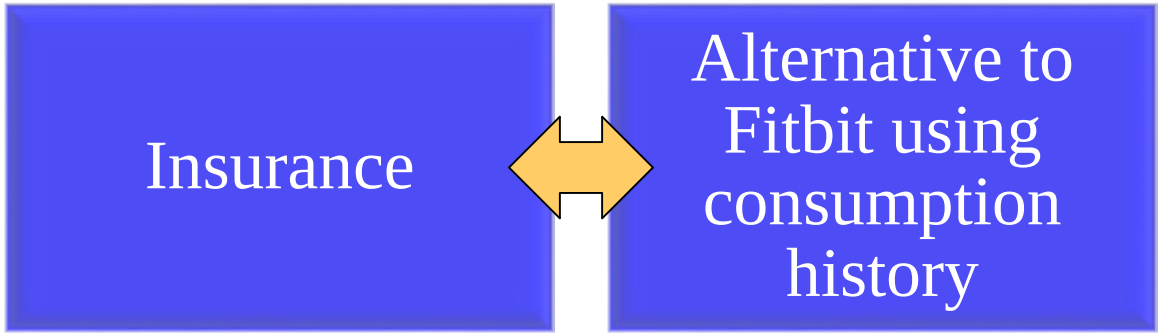
- Life Insurance (usual scenario):
  - Positive Cashflow on survival
  - Negative cashflow on death/maturity
  - Traceable death
- Pension:
  - Negative Cashflow on survival
  - No cashflow on death
  - Traceable death
- Retail Subscription:
  - Positive cashflow on survival
  - No cashflow on death
  - Untraceable death

Crossroads

Now:



Prospect:



Challenges:

- Personal Data
- Adverse selection and bias on both side through voluntary participation

Opportunities:

- Profiling at geographic level